TECHNICKÁ UNIVERZITA V LIBERCI

Fakulta mechatroniky a mezioborových
inženýrských studií

METODY DETEKCE ZMĚNY MLUVČÍHO V AKUSTICKÉM SIGNÁLU

DISERTAČNÍ PRÁCE

2005

JINDŘICH ŽĎÁNSKÝ

METODY DETEKCE ZMĚNY MLUVČÍHO V AKUSTICKÉM SIGNÁLU

DISERTAČNÍ PRÁCE

Disertant:	Jindřich Žďánský
Studijní program:	2612V Elektrotechnika a informatika
Studijní obor:	2612V045 Technická kybernetika
Pracoviště:	Katedra elektroniky a zpracování signálů Fakulta mechatroniky a mezioborových inženýrských studií Technická univerzita v Liberci
Školitel:	Prof. Ing. Jan Nouza, CSc.

ROZSAH PRÁCE:

Počet stran:	103
Počet obrázků:	21
Počet tabulek:	25
Počet vzorců:	145
Počet příloh:	2

Prohlášení

Tuto práci jsem vypracoval samostatně s využitím uvedené literatury a na základě konzultací se svým školitelem.

V Liberci dne 10. října 2005

Jindřich Žďánský

Poděkování

Za všemožnou podporu během doktorského studia, jež vyústilo v tuto disertační práci, patří můj dík zejména školiteli Prof. Ing. Janu Nouzovi, CSc., ale přirozeně také rodičům, všem kolegům, kolegyním, přátelům, Míše Z. a Lucii F..

Anotace

Disertační práce je zaměřena na metody detekce změny mluvčího v akustickém signálu. Autor se v práci zabývá teoretickými východisky a formuluje úlohu detekce jednoho bodu změny jako testování hypotéz změny parametrů gaussovského procesu. Z rozboru problematiky vyplývá, že běžně používaný přístup k testování jednoho bodu změny na základě Bayesovského informačního kritéria není zcela v souladu s principy testování hypotéz a ukazuje, že teoreticky více ospravedlnitelné postupy vedou k lepším výsledkům. Při aplikaci teorie jednoho bodu změny na problém detekce více bodů změny se vymezuje vůči nejběžnějšímu přístupu - metodě fixních oken - a navrhuje algoritmus metody binárního dělení, dobře známý z jiných oblastí change-point analýzy, jako základní algoritmus detekce změny mluvčího. Jako alternativu nabízí vylepšení on-line metody s adaptivním oknem a zcela původní algoritmus přímé multiple change-point analýzy - metodu globální maximalizace BIC.

Práce je strukturována následujícím způsobem:

Kapitola 1 je věnována úvodnímu slovu.

Kapitola 2 krátce popisuje hlavní motivaci pro tuto práci, jíž je media mining systém vyvíjený na Technické univerzitě v Liberci.

Kapitola 3 se věnuje teoretickému rozboru úlohy detekce změny řečníka a popisu stávajících i nově navržených metod vhodných k jejímu řešení.

Kapitola 4 shrnuje základní údaje o databázích užitých v této práci a popisuje metodiky vyhodnocení obdržených výsledků.

Kapitola 5 se zabývá návrhem implementace, možnostmi trénování a vyhodnocením metody binárního dělení.

Kapitola 6 je zainteresována implementaci a vyhodnocení metody globální maximalizace Bayesovského informačního kritéria.

Kapitola 7 popisuje návrh modifikace metody s adaptivním oknem a její výsledky.

Kapitola 8 hodnotí výsledky dosažené v této práci a porovnává dílčí metody z hlediska jejich praktického využití.

Annotation

The dissertation thesis is focused on the issue of speaker change detection in acoustic signals. The author interprets the single change-point problem in terms of the hypothesis testing theory. After the theoretical analysis he shows that the common approach used to solve the single change-point detection problem via the Bayesian information criterion (BIC) is not always in accord with principles of hypothesis testing. It is shown that the methods based on proper theoretical assumptions provide better results. To solve the multiple change-point problem, the author analyses all known methods and proposes several own ones. First of all he discusses the most popular method that is based on a fixed window length scenario and demonstrates its weak points (namely many free parameters). In order to overcome them, he proposes a binary segmentation technique (well-known from other branches of change-point studies) as a fundamental approach to the speaker change detection problem. Furthermore, he suggests a modification of the method that uses an adaptive window length scenario in order to provide an on-line solution. Finally he proposes a novel approach named *Global BIC Maximization* which can be characterized as an attempt to solve the multiple change-point problem globally, i.e. not via a sequential application of single change-point detection.

The thesis consists of the following parts:

Chapter 2 briefly describes the main motivation for this work, which is the developement of a media mining system providing automatic transcription of broadcast news.

Chapter 3 deals with the theoretical analysis of the speaker change detection task and outlines principles of existing and newly proposed methods.

Chapter 4 describes databases that were created and utilized for evaluation tests.

Chapters 5-7 are devoted to the theory, implementation, training and evaluation of the proposed methods: *binary segmentation*, *global BIC maximization* and *a modified version of the adaptive window length method*.

Chapter 8 concludes the thesis and compares the individual methods from the practical point of view.

Obsah

Seznam obrázků

Seznam tabulek

Seznam zkratek

AIC ..... Akaike Information Criterion
AVD ..... Average Delay
AWIN ..... Adaptive WINdow length method
BIC ..... Bayesian Information Criterion
COST ..... CO-operation in Sciece and Technology
DCT ..... Discrete Cosine Transform
ETSI ..... European Telecommunications Standards Institute
FFT ..... Fast Fourier Transform
FPWSIC ..... Fixed Penalty Weight SIC
GMBIC ..... Global BIC Maximization
GMM ..... Gaussian Mixture Model
HMM ..... Hidden Markov Model
FTSIC ..... Fixed Threshold SIC
MAWIN ..... Modified AWIN
MAP ..... Maximum A Posteriori
MFCC ..... Mel-Frequency Cepstral Coeficients
ML ..... Maximum Likelihood
MLE ..... Maximum Likelihood Estimation
MLLR ..... Maximum Log-Likelihood Ratio
MMS ..... Media Mining System
MXD ..... Maximum Delay
NNO ..... Number of Numerical Operations
SIC ..... Schwarz Information Criterion
STFT ..... Short-Time Fourier Transform

KAPITOLA 1

ÚVOD

Na přelomu 20. a 21. století žijeme v období tzv. *informační společnosti*. Jejími hlavními rysy jsou převaha práce s informacemi, interaktivita, integrace a globalizační tendence. Z technologického hlediska lze informační společnost považovat za společnost s *vysokou mírou* využívání informačních a komunikačních technologií založených na prostředcích výpočetní techniky. V souvislosti s obrovským množstvím informací šířených elektronickými médii, z nichž jen nepatrná část je pro konkrétního jedince relevantní, vznikají prostředky pro jejich rychlé a přesné zpracování. V současné době existuje značné množství nástrojů umožňujících efektivní třídění a vyhledávání potřebných informací, takřka bezvýhradně ovšem vyžadují zdroje v *textové podobě*.

Dychtivost současné společnosti po rychlém a snadném přístupu k informačnímu obsahu aktuálních zpráv poněkud komplikuje fakt, že jedním z nejvýznamnějších prostředků vzájemné komunikace a výměny informací je *lidský hlas*. Převod lidské řeči do textové podoby se jeví být nejpřirozenějším postupem jak využít rozvinutých technologií zpracování textových informací k automatickému vytěžování informačního obsahu hlasových nahrávek. Prvopočátky řešení tohoto problému sahají do 50. let 20. století, kdy byl vyvinut první praktický systém pro rozpoznávání řeči v Bellových Laboratořích v USA, který umožňoval rozpoznávání číslovek nula až devět po telefonní lince. Během následujících 30 let se objevila celá řada různých přístupů k rozpoznávání řeči, nejperspektivnějším se stala metoda založená na teorii *skrytých Markovových modelů*. Principiálně se tuto metodu od 80. let nepodařilo překonat a v dnešní ji lze už považovat za standardní technologii převodu lidské řeči do textu.

Jedním z nejožehavějších témat současného výzkumu v oblasti rozpoznávání řeči je *automatický přepis* zvukových nahrávek různých televizních či rádiových pořadů, jako jsou politické debaty či zpravodajství. Tento přepis slouží jako základní materiál pro automatickou *indexaci* multimediálních archivů. Komplexní systémy umožňující vytěžování informací z těchto multimediálních archivů se nazývají *media mining systémy*. Pro nejrozšířenější světové jazyky (angličtina, japonština, francouzština a němčina) jsou intenzivně vyvíjeny již několik let. Obdobný media mining systém pro češtinu je vyvíjen i na Technické univerzitě v Liberci.

Každý vyspělý systém pro automatický přepis audio nahrávek je složen z několika základních komponent. Konkrétně se jedná o část *zpracování signálu*, *detektor řeč-neřeč*, *detektor změny řečníka*, modul *identifikace a verifikace řečníka* a *roz-*

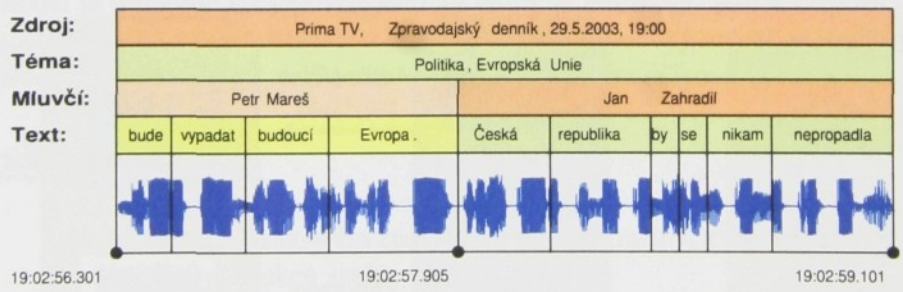
poznávač řeči. Tato práce se zabývá pouze třetím jmenovaným, tj. problematikou detekce změny řečníka. Detektor změny řečníka slouží k *segmentaci* akustického signálu do úseků náležejících jednotlivým mluvčím, což je nezbytným krokem, chceme-li identifikovat hovořící osobu. Totožnost mluvčího pak slouží také ke zvýšení robustnosti přepisu.

Disertační práce je strukturována následujícím způsobem. V kapitole 2 je krátce popsána hlavní motivace pro tuto práci, jíž je media mining systém vyvíjený na Technické univerzitě v Liberci. Náplní kapitoly 3 je teoretický rozbor úlohy detekce změny řečníka a popis stávajících i nově navržených metod vhodných k jejímu řešení. Základními údaji o databázích užitých v této práci a popisem metodiky vyhodnocení obdržených výsledků se zaobírá kapitola 4. Následující tři kapitoly jsou věnovány filosofii přístupu, popisu implementace a vyhodnocení úspěšnosti autorem navržených detektorů. Kapitola 5 se zabývá *metodou binárního dělení*, kapitola 6 *metodou globální maximalizace BIC* a kapitola 7 *on-line metodou s adaptivním oknem*. Pro vyšší přehlednost je porovnávání výsledků dílčích metod uskutečňováno inkrementálně, tj. každá následující metoda je srovnávána se všemi předchozími.

KAPITOLA 2

DETEKCE ZMĚNY ŘEČNÍKA V MEDIA MINING SYSTÉMECH

Vzhledem k obrovskému množství informací šířených masmédii, existuje velké množství společností, které se zabývají monitorováním a archivací moderních elektronických médií. Jejich úkolem je každodenní sběr co největšího množství např. zpravodajských pořadů, ať už v podobě textů nebo ve formě audio, audiovizuálních či obecně multimediálních záznamů, a jejich zařazení do databázových systémů. Získat tyto multimediální záznamy je z technického hlediska velmi snadné, ovšem jakmile jsou jednou zařazeny do archivu, je velmi obtížné se v nich orientovat a vyhledávat, což je hlavní příčinou, proč je nutné u každého pořadu provést důkladnou *transkripci*. Pod pojmem transkripce rozumíme víceúrovňový přepis multimediálního záznamu (s ohledem na typ zpracovávaných dat). Následné vytvoření rejstříku jednotlivých popisků se nazývá *indexace*. Na obrázku 2.1 je znázorněn typický příklad transkripce zpravodajského pořadu, kde předmětem zájmu je především text, mluvčí, téma, zdroj, datum a čas záznamu.



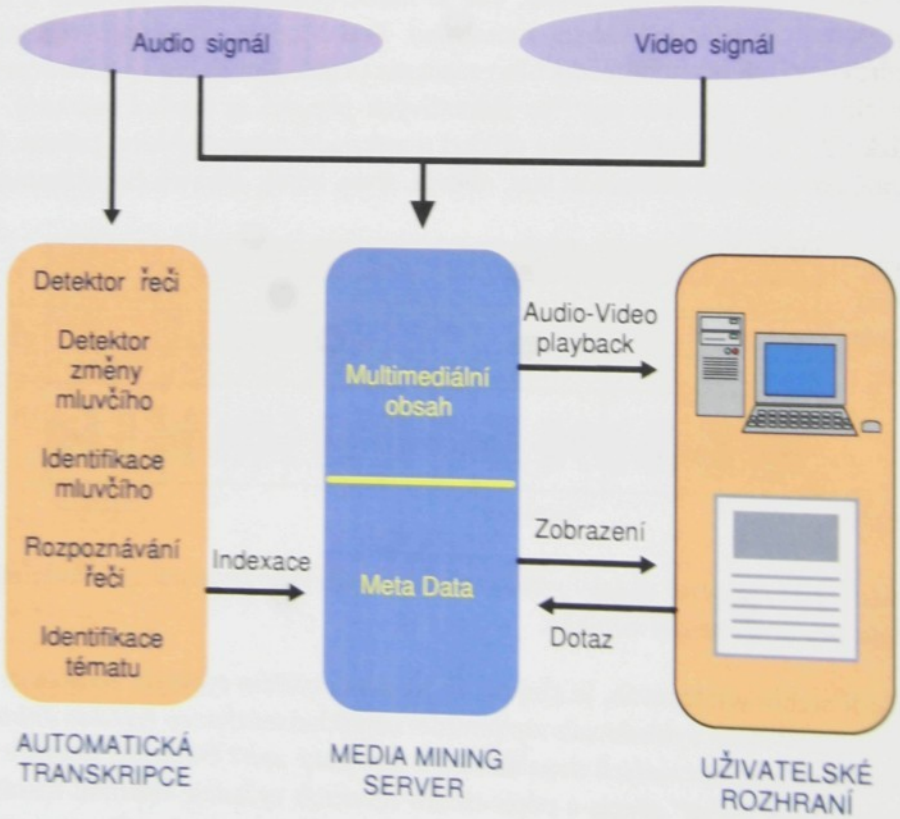
Obrázek 2.1: *Názorná ukázka víceúrovňové transkripce záznamu zpravodajského pořadu pro účely media miningu.*

Je-li archiv oindexován, je zřejmé, že již není problém rychlým způsobem pomocí *full-textového vyhledávače* zodpovědět například otázku: *co řekl Jan Zahradil v médiích během uplynulých dvou měsíců o Evropské unii*? Na druhé straně je pochopitelné, že „ruční" přepis a popis těchto záznamů vyžaduje enormní množství lidského úsilí, což je hlavní důvod, proč se laboratoře počítačového zpracování řeči na celém světě snaží o tvorbu co nejdokonalejších automatizovaných transkripčních systémů. Pojmem *media mining systém* (MMS) je pak označován sou-

bor softwarových a hardwarových nástrojů, jež umožňují automatizovanou tvorbu, správu, aktualizaci a prohledávání multimediálních archivů.

2.1 Architektura automatického media mining systému

Typická architektura automatického media mining systému pro zpracování multimediálních dat je znázorněna na obrázku 2.2. Přijímaný audiovizuální signál je digitalizován a komprimován do některého ze standardizovaných formátů. Jeho audio složka pak vstupuje do *modulu automatické transkripce*, jehož úkolem je poskytnout víceúrovňový časový popisu audio záznamu co možná nejvíce se blížícího ukázce na obrázku obr. 2.1. Následnou indexací je vytvořen rejstřík, jenž obsahuje odkazy na *meta data*, což jsou v podstatě dílčí multimediální záznamy spolu s korespondujícími přepisy. Meta data jsou uchovávána a organizována prostřednictvím *media mining serveru*. Pro vyhledávání v archivu, tj. komunikaci se serverem, pak slouží klientská aplikace s implementovaným *uživatelským rozhraním*.



Obrázek 2.2: *Typická architektura automatizovaného media mining systému.*

2.2 Modul automatické transkripce

Proces automatické transkripce začíná buď aplikací *detektoru řeči* nebo *detektoru změny mluvčího*, záleží na konkrétní implementaci systému. V prvním případě jsou nejprve odstraněny neřečové části záznamu, jimiž není pouze ticho, ale také hudba či jiné ruchy, a následně se hledají místa změny řečníka. Nalezením těchto bodů změny vznikají segmenty, jež obsahují řeč právě jednoho mluvčího. Tyto segmenty jsou poté podrobeny procesu *identifikace řečníka*, čímž vybereme nejpravděpodobnějšího mluvčího z existující databáze osob. Následná *verifikace* potvrdí či zamítne hypotézu, zda se opravdu jedná o odpovídající osobu.

V druhém případě se nejprve využije detektoru změny řečníka k nalezení akusticky homogenních segmentů. Místo označení detektor změny řečníka se pak používá pojmu *detektor akustických změn*, využívá však naprosto stejných principů. Detekované segmenty jsou následně odesílány do jednotky *identifikace segmentu*, která pomocí principů podobných identifikaci mluvčího umožňuje rozlišit i jiné třídy dat, jako jsou hudba, ticho...atd.

V obou případech je výsledkem audio signál segmentovaný dle jednotlivých mluvčích spolu s informací o identitě mluvčího. V ideálním případě, jedná-li se o osobu frekventovanou v médiích, známe přímo její jméno. V ostatních případech je užitečnou pomůckou alespoň pohlaví dané osoby. Toto je důležité nejen z hlediska transkripce jako takové, ale především pro robustnost převodu řeči do textové podoby.

Rozpoznávače řeči pro tyto účely jsou v současné době založeny výhradně na technologii *skrytých Markovových modelů* (HMM). Mimo technologie samé jsou dalšími důležitými součástmi rozpoznávače *akustický a jazykový model*. Akustický model je primárně koncipován jako *na mluvčím nezávislý*. Je-li k dispozici totožnost mluvčího, lze použít tzv. *na mluvčím závislý* akustický model, což vede k vyšší přesnosti přepisu. Obdobně je tomu i s jazykovým modelem a *identifikací tématu*.

2.3 Cíle práce

- Prozkoumat a jednotným způsobem popsat principy, z nichž lze vycházet při řešení úlohy detekce změny řečníka;
- modifikovat či nově vytvořit metody vhodné pro zadanou problematiku;
- vypracovat jejich efektivní algoritmizaci;
- navrhnout metodu pracující v on-line režimu;
- porovnat jednotlivé metody na uměle vytvořených datech simulujících změny mluvčích a ověřit nejslibnější přístupy na reálných nahrávkách;
- realizovat modul pro segmentaci skutečných zpravodajských pořadů využitelný jako součást vyvíjeného systému pro přepis televizních zpráv.

KAPITOLA 3

PRINCIPY DETEKCE ZMĚNY ŘEČNÍKA

Detekce změny řečníka je úloha spadající do oblasti statistické analýzy zabývající se hledáním *bodu změny* ve stochastickém procesu. Motivací, proč se věnovat této problematice, je poptávka po tzv. *change-point analýze* v mnoha vědních oborech - od ekonomie, přes medicínu a geologii, až po vědu literární. První pokusy o řešení *change-point problému* se objevily již na počátku 50. let. Od té doby byla publikována celá řada studií a článků, dosud však neexistuje ucelený teoretický pohled na danou problematiku.

Change-point analýza se zabývá dvěma kardinálními problémy. První, tzv. *single change-point problém*, je definován jako otázka detekce jednoho bodu změny ve stochastickém procesu. Při jeho řešení je nejčastěji uplatňován přístup pomocí *testování hypotéz*, jenž má také nejhlubší teoretické základy. Druhým předmětem zájmu je tzv. *multiple change-point analýza*, jejímž cílem je nalezení více bodů změny v náhodném procesu. Z teoretického pohledu je multiple change-point problém vzhledem ke komplikovanému matematickému aparátu velmi obtížně řešitelný. Úloha detekce více bodů změny je proto často převáděna na single change-point problém pomocí některé z více či méně známých *metod*.

Z pohledu statistického je bodem změny takové místo nebo čas stochastického procesu, o kterém lze tvrdit, že od počátku pozorování až do bodu změny byl zaznamenán odlišný proces, než od bodu změny do konce pozorování. Úkolem change-point analýzy je zodpovědět zdali:

- nastala v daném pozorování procesu nějaká změna;
- pakliže ano, kolik změn nastalo;
- a kde jsou místa změn.

Je zřejmé, že obtížné je vůbec definovat, co již za změnu považujeme a co ještě ne. Naštěstí v případě řešení úlohy segmentace audio signálu dle jednotlivých mluvčích je možné definovat změnu vcelku jasně - alespoň slovně: je to čas, kdy přestane mluvit jeden řečník a začne hovořit jiný[1].

[1] V praxi však i takto zřejmá definice naráží na problémy spojené např. s tím, že mezi promluvou jedné a druhé osoby je kratší či delší pauza. Otázkou je kam pak umístit bod změny: doprostřed či označit pauzu jako zvláštní segment?

3.1 Problematika detekce bodu změny stochastického procesu

V této části si ve zkratce nastíníme princip testování hypotéz a převedeme problém detekce změny řečníka na úlohu detekce změny parametrů náhodného procesu.

3.1.1 Testování hypotéz

Mějme plně specifikovanou třídu parametrických modelů $\{f(x|\theta), \theta \in \Theta\}$, kde Θ je rozděleno do dvou vzájemně se neprolínajících částí Θ_0 a Θ_1 a úkolem je rozhodnout, zda neznámé parametry θ náleží do Θ_0 či Θ_1. Jestliže označíme H_0 jako hypotézu, že $\theta \in \Theta_0$, a H_1 jako hypotézu, že $\theta \in \Theta_1$, pak existují dvě možná rozhodnutí: $a_0 \equiv$ přijmout H_0 nebo $a_1 \equiv$ přijmout H_1. Jelikož jsou obě hypotézy navzájem komplementární, přijetí H_1 se rovná zamítnutí H_0. Hypotéza H_0 se nazývá *nulová*, hypotéza H_1 se označuje jako *alternativní*.

V bayesovské teorii rozhodování se zavádí tzv. *pokutová funkce*, která definuje pokutu, již musíme zaplatit při přijetí rozhodnutí a_i. Nadefinujeme-li *pokutovou funkci* ve tvaru

$$L(a_i, \theta) = \begin{cases} 0 & \theta \in \Theta_i, \quad i = 0, 1 \\ l_{ij} & \theta \in \Theta_j, \quad j \neq i, \end{cases} \tag{3.1}$$

tvrdíme, že přijetím správného rozhodnutí neutrpíme žádnou ztrátu, kdežto přijetím chybného rozhodnutí zaplatíme pokutu l_{ij}. *Bayesovské riziko* přijetí hypotézy H_i, máme-li k dispozici data $\mathbf{x}$, je

$$R(a_i|\mathbf{x}) = l_{i0}P(H_0|\mathbf{x}) + l_{i1}P(H_1|\mathbf{x}) \tag{3.2}$$

a značí matematické očekávání ztráty dané přijetím rozhodnutí a_i. Hypotézu H_0 zamítneme tehdy a jen tehdy, když riziko jejího přijetí bude větší než než riziko přijetí H_1, tj. když

$$R(a_0|\mathbf{x}) > R(a_1|\mathbf{x}). \tag{3.3}$$

Po dosazení vztahu (3.2) do (3.3) bude mít podmínka zamítnutí H_0 tvar

$$\frac{l_{01}P(H_1|\mathbf{x})}{l_{10}P(H_0|\mathbf{x})} > 1. \tag{3.4}$$

S využitím *Bayesova vzorce*

$$P(H_i|\mathbf{x}) = \frac{f(\mathbf{x}|H_i)P(H_i)}{f(\mathbf{x})} \tag{3.5}$$

a dosazením do nerovnosti (3.4) získáme podmínku zamítnutí H_0 ve tvaru

$$\frac{l_{01}P(H_1)}{l_{10}P(H_0)} > \frac{f(\mathbf{x}|H_0)}{f(\mathbf{x}|H_1)} = B_{01}(\mathbf{x}), \tag{3.6}$$

kde poměr $B_{01}(\mathbf{x})$ se nazývá *Bayesovský faktor*.

Z hlediska klasické teorie rozhodování lze problém testování hypotéz chápat jako hledání rozhodovací *strategie* δ s odpovídajícím *kritickým regionem* R_δ, jenž je definován jako množina hodnot $\mathbf{x}$ takových, že H_0 zamítneme kdykoliv $\mathbf{x} \in R_\delta$. Kvalitu strategie - *testovací procedury* - δ lze popsat pomocí *power funkce*

$$pow(\theta|\delta) = P(\mathbf{x} \in R_\delta|\theta), \tag{3.7}$$

jež specifikuje „dlouhodobou" pravděpodobnost nesprávného zamítnutí pravdivé nulové hypotézy H_0 jako funkci parametrů θ. Ideální power funkce by měla nabývat tvaru:

$$pow(\theta|\delta) = \begin{cases} 0, & \theta \in \Theta_0 \\ 1, & \theta \in \Theta_1. \end{cases} \tag{3.8}$$

Nakolik se reálná power funkce blíží ideální, specifikuje tzv. *velikost* testu θ, jež je definována jako

$$\alpha = \sup_{\theta \in \Theta_0} pow(\theta|\delta). \tag{3.9}$$

Apriorním určením určité *úrovně signifikance* α_0 se omezujeme na testy, jejichž velikost nepřekročí α_0.

Pro jakoukoliv testovací proceduru δ můžeme explicitně očekávat dva druhy chyb: zamítnutí pravdivé nulové hypotézy, tj. *chyba I. druhu*, a přijetí nepravdivé nulté hypotézy, což je tzv. *chyba II. druhu*. Označíme-li pravděpodobnosti těchto chyb jako $\alpha(\delta|\theta)$ a $\beta(\delta|\theta)$, pak

$$\alpha(\delta|\theta) = \begin{cases} P(\mathbf{x} \in R_\delta|\theta) & \text{když } \theta \in \Theta_0, \\ 0 & \text{jinak} \end{cases} \tag{3.10}$$

$$\beta(\delta|\theta) = \begin{cases} P(\mathbf{x} \notin R_\delta|\theta) & \text{když } \theta \in \Theta_1, \\ 0 & \text{jinak.} \end{cases} \tag{3.11}$$

Při tvorbě testů je snahou navrhnout takový test, jenž bude mít oba typy chyb pokud možno co nejmenší. Typicky však změnou kritického regionu R_δ za účelem zmenšení pravděpodobnosti jednoho typu chyby, pravděpodobnost druhého typu chyby roste. Nejběžnějším řešením bývá například minimalizace jejich lineární kombinace $a\alpha(\delta|\theta) + b\beta(\delta|\theta)$.

Pokud mohou prostory parametrů Θ_0, Θ_1 obsahovat pouze jednu hodnotu θ, nazývají se odpovídající hypotézy *jednoduché*, v opačném případě *kompozitní*. V případě jednoduchých hypotéz nabývá Bayesův faktor tvaru

$$B_{01}(\mathbf{x}) = \frac{f(\mathbf{x}|\theta_0)}{f(\mathbf{x}|\theta_1)} \tag{3.12}$$

a při neznámých parametrech θ_i směřuje na *metodu maximální věrohodnosti* popsané v části 3.3.1. Uvažujeme-li kompozitní hypotézy, lze Bayesův faktor vyjádřit jako

$$B_{01}(\mathbf{x}) = \frac{\int_{\Theta_0} f(\mathbf{x}|\theta) f(\theta) d\theta}{\int_{\Theta_1} f(\mathbf{x}|\theta) f(\theta) d\theta}. \tag{3.13}$$

Aproximace integrálů v této rovnici vede na metodu založenou např. na *Schwarzově informačním kritériu* (SIC) a je jí věnována část 3.3.2.

3.1.2 Formulace problému

Zcela obecně uvažujme, že pozorujeme nějaký stochastický proces. Jeho záznamem získáme posloupnost náhodných příznakových vektorů $x_1, x_2, \ldots, x_T$ délky T. Dále předpokládejme, že se jedná o realizaci nezávislých náhodných veličin $X_1, X_2, \ldots, X_T$ popsaných distribučními funkcemi $F_1, F_2, \ldots, F_T$. Potom lze úlohu detekce bodů změny považovat za problém testování nulové hypotézy:

$$H_0 : F_1 = F_2 = \ldots F_T \tag{3.14}$$

versus alternativní

$$\begin{aligned} H_1 \quad &: \quad F_1 = \ldots = F_{t_1} \neq F_{t_1+1} = \ldots = F_{t_2} \\ &\neq \quad F_{t_2+1} = \ldots = F_{t_s} \neq F_{t_s+1} = \ldots = F_T, \end{aligned} \tag{3.15}$$

kde $1 < t_1 < t_2 < \ldots < t_s < T$, s je neznámý počet bodů změny a $t_1, t_2, \ldots, t_s$ jsou pozice hledaných bodů změn. Předpokládejme, že distribuční funkce $F_1, F_2, \ldots, F_T$ patří do stejné parametrické třídy $F(\theta)$, kde $\theta \in R^p$ jsou parametry dané třídy rozložení náhodných veličin. Pak lze problém detekce bodů změny převést z testování distribučních funkcí na testování jejich parametrů θ_i, $i = 1, \ldots, T$. Testujeme tedy nulovou hypotézu:

$$H_0 : \theta_1 = \theta_2 = \ldots \theta_T = \theta \quad \text{(neznámé)} \tag{3.16}$$

oproti alternativní

$$\begin{aligned} H_1 \quad &: \quad \theta_1 = \ldots = \theta_{t_1} \neq \theta_{t_1+1} = \ldots = \theta_{t_2} \\ &\neq \quad \theta_{t_2+1} = \ldots = \theta_{t_s} \neq \theta_{t_s+1} = \ldots = \theta_T, \end{aligned} \tag{3.17}$$

kde s a $t_1, t_2, \ldots, t_s$ potřebujeme odhadnout.

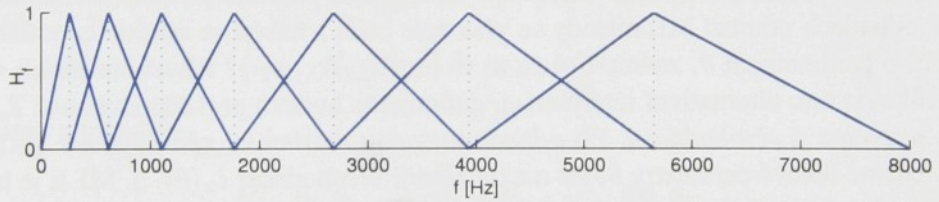
3.2 Specifika úlohy detekce změny řečníka

Lidský hlas se šíří formou stlačování a zřeďování vzduchu. Měřením těchto výchylek tlaku pomocí mikrofonu vzniká elektrický signál, jenž je následně digitalizován, a to s ohledem na předpokládané použití a vzorkovací teorém. Takto diskretizovaný signál však není příliš vhodný ani k rozpoznávání řečníka, natož pak k detekci jeho změny. Proto je nutné transformovat původní jednorozměrný signál do vhodného příznakového prostoru. Zřejmě nepopulárnější reprezentací řečového signálu v úloze změny řečníka jsou tzv. *kepstrální příznaky*, mezi nejoblíbenější patří *mel-frekvenční kepstrální koeficienty* (MFCC), neboť vykazují robustnost vůči šumu a chybám v odhadu spektra. Stručným popisem jejich výpočtu se zabývá část 3.2.1.

V oblastech rozpoznávání řeči a mluvčího je zavedeným statistickým modelem kepstrálních příznaků *multimodální vícerozmněrné normální (Gaussovo) rozložení* (GMM), neboť dokáže při dostatečném počtu modů aproximovat jakékoliv jiné statistické rozložení. Čím více má však distribuční funkce modelu parametrů[2], tím více potřebuje dat pro jejich spolehlivý odhad, což v případě úlohy detekce změny řečníka je značný problém. V této práci je unimodální vícerozměrný gaussovský model považován za rozumný model řečníka a je popsán v části 3.2.2.

3.2.1 Mel-frekvenční kepstrální koeficienty

Mel-frekvenční kepstrální koeficienty patří do rodiny tzv. perceptuálně motivovaných příznaků a využívají poznatků psychoakustiky o nelineárním subjektivním vnímání výšky tónu lidským sluchem. Pro nízké kmitočty je tato subjektivní stupnice shodná s objektivní frekvenční stupnicí, nad 1 kHz pak stoupá pomaleji [Jiříček, 2002]. Mel-kepstrální reprezentace řečového signálu je definována jako *reálné kepstrum* krátkodobých stacionárních úseků řeči[3] odvozené z *krátkodobé Fourierovy transformace* (STFT). Na rozdíl od klasického postupu při výpočtu reálného kepstra, je ještě na spektrum signálu aplikována banka trojúhelníkových filtrů se středy logaritmicky rozmístěnými po frekvenční ose - *melovská banka filtrů* (viz obr. 3.1).



Obrázek 3.1: *Ukázka melovské banky trojúhelníkových filtrů užívané při výpočtu MFCC příznaků.*

Parametrizace řečového signálu začíná *preemfází*, tj. filtrováním přes jednoduchou horní propust prvního řádu. Následně je signál segmentován a váhován *hammingovým okénkem*, typicky je aplikováno každých 10 ms na segment délky 25 ms. Tyto krátké části signálu jsou transformovány pomocí *rychlé Fourierovy transformace* (FFT) do frekvenční oblasti a vstupují do melovské banky filtrů, klasicky o rozsahu 24-26 filtrů. Poté je spočítán logaritmus energie v každém pásmu a výsledné hodnoty jsou pomocí *kosinové transformace* (DCT) převedeny na kepstrální koeficienty. Podrobnější popis této parametrizace lze nalézt např. v telekomunikačním standardu [ETSI, 2003]. Zatímco pro rozpoznávání řeči se využívá prvních 13 koeficientů ($c_0 - c_{12}$) doplněných o jejich první a druhé diference, pro rozpoznávání a detekci změny řečníka jsou vhodné pouze koeficienty $c_1 - c_{12}$.

[2]Počet parametrů se zhruba zdvojnásobuje s každým modem.

[3]Lidská řeč je považována za tzv. kvazi-stacionární signál, tj. signál stacionární v krátkodobých úsecích, typicky 10–30 ms.

3.2.2 Vícerozměrné normální rozložení

Náhodná d-rozměrná veličina X s normálním rozložením je popsána hustotou pravděpodobnosti

$$f_X(x) = \frac{1}{\sqrt{(2\pi)^d |\Sigma|}} \exp\left[-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right], \tag{3.18}$$

kde μ značí d-rozměrný *vektor středních hodnot*, Σ je $d \times d$ rozměrná *kovarianční matice* a $|\Sigma|$ její determinant. Parametry μ a Σ plně specifikují normální náhodnou veličinu a lze je vyjádřit jako:

$$\mu = E(x), \tag{3.19}$$

$$\Sigma = E\left[(x-\mu)(x-\mu)'\right], \tag{3.20}$$

kde $E(\cdot)$ značí operátor matematického očekávání.

3.2.2.1 Odhad parametrů metodou maximální věrohodnosti

Metoda maximální věrohodnosti (ML) je jednou z nejrozšířenějších procedur v oblasti estimace parametrů a testování hypotéz. Myšlenka metody maximální věrohodnosti je vcelku jednoduchá. Běžně považujeme hustotu pravděpodobnosti $f_X(x|\theta)$ za funkci hodnoty náhodné veličiny x a parametry modelu θ za fixní. V odhadech pomocí ML metody se však role otáčí a nabízí se otázka: co můžeme říct o parametrech θ, známe-li data $\mathbf{x} = \{x_1, x_2, \ldots, x_T\}$? Abychom mohli specifikovat tuto alternativní interpretaci, definujeme hustotu pravděpodobnosti $\ell_x(\theta)$ a nazveme ji *věrohodností*. Při *odhadu metodou maximální věrohodnosti* (MLE) hledáme takové parametry $\hat{\theta}$, jež maximalizují věrohodnost $\ell_x(\theta)$, tj. MLE je hodnota θ, jež by při daném modelu nejvěrohodněji vygenerovala data $\mathbf{x}$.

Pakliže lze předpokládat nezávislost proměnných $\mathbf{X} = \{X_1, X_2, \ldots, X_T\}$, můžeme zapsat věrohodnost následujícím způsobem:

$$\ell_x(\theta) = \prod_{i=1}^{T} f_X(x_i|\theta). \tag{3.21}$$

Je-li $\theta = (\theta_1, \theta_2, \ldots, \theta_k)'$ k-složkový vektor parametrů modelu a ∇_θ je operátor gradientu, pak řešíme následující soustavu k-rovnic:

$$\nabla_\theta \ell_x(\theta) = 0, \tag{3.22}$$

pro vícerozměrný Gaussův model je řešení uvedeno v příloze A.1 a nabývá notoricky známé podoby:

$$\hat{\mu} = \frac{1}{T}\sum_{i=1}^{T} x_i, \tag{3.23}$$

$$\hat{\Sigma} = \frac{1}{T}\sum_{i=1}^{T}(x_i-\mu)(x_i-\mu)'. \tag{3.24}$$

Hodnota věrohodnosti v bodě $\ell_x(\hat{\theta})$ je pak (viz A.1):

$$\ell_x(\hat{\mu}, \hat{\Sigma}) = (2\pi)^{-Td/2} |\hat{\Sigma}|^{-T/2} e^{-Td/2}, \tag{3.25}$$

kde d je rozměr náhodné veličiny.

Odhad parametrů pomocí metody maximální věrohodnosti je označován jako asymptoticky *konzistentní* a *nejlepší*, což znamená, že je-li učiněn z nekonečného množství trénovacích vzorků, konverguje ke skutečným parametrům procesu, jenž tyto vzorky vygeneroval, tj.

$$\lim_{T \to \infty} \hat{\theta} = \tilde{\theta} \tag{3.26}$$

a zároveň má nejmenší rozptyl. Na druhou stranu však ML odhad nemusí být nestranný.

3.3 Detekce jednoho bodu změny

V této části si uvedeme dva základní přístupy k detekci jednoho bodu změny v gaussovském procesu - na základě testování jednoduchých a kompozitních hypotéz. Předpokládejme, že $\mathbf{X} = \{X_1, X_2, \ldots, X_T\}$ jsou nezávislé náhodné normálně rozložené veličiny s parametry $\theta_i = (\mu_i, \Sigma_i), i = 1 \ldots T$. Řešení problému detekce bodu změny v realizaci $\mathbf{x} = \{x_1, x_2, \ldots, x_T\}$ pak spočívá v testování hypotéz:

$$H_0 : \theta_1 = \theta_2 = \ldots = \theta_T \tag{3.27}$$

versus alternativa

$$H_1 : \theta_1 = \ldots = \theta_t \neq \theta_{t+1} = \ldots = \theta_T, \tag{3.28}$$

kde t je testovaný bod změny. Principiálně, uvažujeme-li gaussovský proces, můžeme testovat:

1. změnu ve vektoru středních hodnot;
2. změnu kovarianční matice;
3. obecnou změnu parametrů.

Výběr vhodného testu závisí na typu řešené úlohy. Je zřejmé, že bude-li cílem úlohy detekovat např. změnu parametrů přenosového kanálu při telefonickém spojení, bude záhodno testovat střední hodnoty, neboť „konvoluční šum" je v kepstrální oblasti aditivní. Změna řečníka je však úloha z teoretického pohledu velmi nespecifická a nelze přesně říci, zdali dochází ke změně střední hodnoty či kovariance. Z tohoto důvodu se v této práci autor zabývá výhradně obecnou změnou parametrů, tj. testováním následujících hypotéz:

$$H_0 : \mu_1 = \mu_2 = \ldots = \mu_T \quad \text{a} \quad \Sigma_1 = \Sigma_2 = \ldots = \Sigma_T \quad (\mu, \Sigma \text{ neznámé}) \tag{3.29}$$

oproti hypotéze

$$\begin{aligned} H_1 : \mu_1 &= \ldots = \mu_t \neq \mu_{t+1} = \ldots = \mu_T \quad \text{a} \\ \Sigma_1 &= \ldots = \Sigma_t \neq \Sigma_{t+1} = \ldots = \Sigma_T, \end{aligned} \tag{3.30}$$

kde $d < t < T - d$ a d je rozměr náhodného vektoru.

3.3.1 Metoda maximální věrohodnosti

V části 3.1.1 věnované teoretickým základům procedury testování hypotéz bylo uvedeno, že postup kdy považujeme obě hypotézy za jednoduché vede na metodu maximální věrohodnosti. Vyjdeme z nerovnosti (3.6) a využijeme vztah (3.12), kde nahradíme parametry θ_i jejich ML odhadem $\hat{\theta}_i$. Hypotézu H_0 zamítneme, tzn. nastala změna, když bude poměr věrohodností splňovat podmínku (viz např. [Lehmann, 1986]):

$$r(x) = \frac{\sup_{\theta \in \Theta_0} \ell_x(\theta)}{\sup_{\theta \in \Theta_1} \ell_x(\theta)} = \frac{\ell_x(\hat{\theta}_0)}{\ell_x(\hat{\theta}_1)} < C, \tag{3.31}$$

kde C je bod determinující kritický region. Nerovnost (3.31) se nazývá *test* nebo *procedura poměru věrohodností* a poměr $r(x)$ definuje odpovídající *statistiku*.

V případě hypotézy H_0 definované v (3.29) bude maximum věrohodnostní funkce odpovídat přímo vztahu (3.25), tj.

$$\ell_x(\hat{\theta}_0) = (2\pi)^{-Td/2} |\hat{\Sigma}|^{-T/2} e^{-Td/2}, \tag{3.32}$$

kde

$$\hat{\mu} = \frac{1}{T} \sum_{i=1}^{T} x_i, \tag{3.33}$$

$$\hat{\Sigma} = \frac{1}{T} \sum_{i=1}^{T} (x_i - \hat{\mu})(x_i - \hat{\mu})'. \tag{3.34}$$

Za předpokladu hypotézy H_1 lze vyjádřit maximum věrohodnostní funkce v závislosti na volbě místa bodu změny t jako

$$\ell_x(\hat{\theta}_1) = (2\pi)^{-Td/2} |\hat{\Sigma}_1|^{-t/2} |\hat{\Sigma}_T|^{-\frac{T-t}{2}} e^{-Td/2}, \tag{3.35}$$

kde

$$\hat{\mu}_1 = \frac{1}{t} \sum_{i=1}^{t} x_i, \tag{3.36}$$

$$\hat{\Sigma}_1 = \frac{1}{t} \sum_{i=1}^{t} (x_i - \hat{\mu}_1)(x_i - \hat{\mu}_1)', \tag{3.37}$$

$$\hat{\mu}_T = \frac{1}{T-t}\sum_{i=t+1}^{T} x_i, \tag{3.38}$$

$$\hat{\Sigma}_T = \frac{1}{T-t}\sum_{i=t+1}^{T} (x_i - \hat{\mu}_T)(x_i - \hat{\mu}_T)'. \tag{3.39}$$

Uvážíme-li vztahy (3.32) a (3.35), pak statistika procedury poměru věrohodností bude rovna

$$r(x|t) = \frac{|\hat{\Sigma}_1|^{t/2}|\hat{\Sigma}_T|^{(T-t)/2}}{|\hat{\Sigma}|^{T/2}}. \tag{3.40}$$

Není-li znám bod změny t, zavádí se tzv. *statistika maximálního typu* [Lauro & kol, 2002], jež má podobu

$$\Lambda_{\hat{t}} = \max_{d<t<T-d} r(x|t)^{-2} = \max_{d<t<T-d} \frac{|\hat{\Sigma}|^T}{|\hat{\Sigma}_1|^t|\hat{\Sigma}_T|^{T-t}} \tag{3.41}$$

a díky operaci záporné mocniny[4], se mění kritérium pro zamítnutí H_0 na

$$\Lambda_{\hat{t}} > \frac{1}{C^2} = K. \tag{3.42}$$

V literatuře [Chen J., 2000] [Lauro & kol, 2002] lze nalézt, že řešení rovnice

$$\hat{t} = \arg \max_{d<t<T-d} r(x|t)^{-2} \tag{3.43}$$

poskytuje konzistentní odhad bodu změny t.

Protože $\Lambda_{\hat{t}}$ je náhodná veličina, není nezajímavé znát její distribuční funkci. Ačkoliv její přesná podoba není dosud známa, Horváth v roce 1993 [Horváth, 1993] odvodil alespoň asymptotickou podobu distribuční funkce náhodné veličiny

$$Y = a(\log T)(\log \Lambda_{\hat{t}})^{\frac{1}{2}} - b_{2d}(\log T), \tag{3.44}$$

kde

$$a(\log T) = (2\log\log T)^{\frac{1}{2}} \text{ a} \tag{3.45}$$

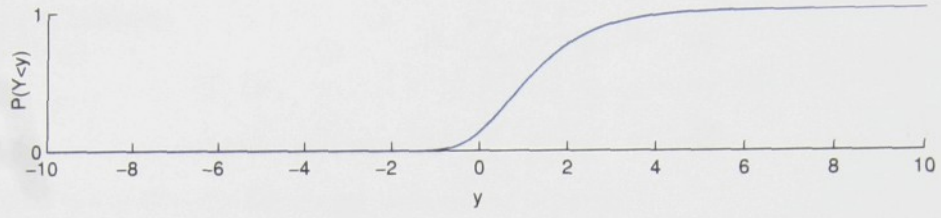
$$b_{2d}(\log T) = 2\log\log T + d\log\log\log T - \log\Gamma(d). \tag{3.46}$$

Za předpokladu pravdivé nulové hypotézy H_0, pro $T \to \infty$ má Y distribuční funkci

$$\lim_{T\to\infty} P(Y < y) = \exp(-2e^{-y}), \tag{3.47}$$

kde $y \in \Re$ a její průběh je znázorněn na obrázku 3.2.

[4]Tato operace z faktického hlediska není nutná. Druhá mocnina je zavedena kvůli zjednodušení vztahu. Záporné znaménko u mocniny je použito pouze z toho důvodu, aby autor udržel přiměřenou konzistenci terminologie, neboť ta se v literatuře značně různí.



Obrázek 3.2: *Asymptotická distribuční funkce náhodné veličiny Y definované vzorcem (3.44).*

Závěrem této části je testovací procedura vycházející z transformace (3.44). Dosadíme-li do ní vztah (3.41), obdržíme po několika úpravách finální řešení testování obecné změny parametrů gaussovského modelu metodou maximální věrohodnosti ve formě

$$y = \alpha \sqrt{\max_{d<t<T-d} \left[T \log |\hat{\Sigma}| - t \log |\hat{\Sigma}_1| - (T-t) \log |\hat{\Sigma}_T| \right]} - \beta, \qquad (3.48)$$

kde $\alpha = a(\log T)$, $\beta = b_{2d}(\log T)$ jsou definovány v (3.45), (3.46) a $\hat{\Sigma}$, $\hat{\Sigma}_1$, $\hat{\Sigma}_T$ se dají vypočítat dle vztahů (3.34), (3.37), (3.39). Změna parametrů nastane, když hodnota y bude větší než nějaká kritická mez K, jejíž hodnotu odhadneme na trénovacích datech.

3.3.2 Přístup pomocí informačních kritérií

V části 3.1.1 bylo uvedeno, že test kompozitních hypotéz směřuje na přístup přes tzv. Schwarzovo informační kritérium. Zcela obecně jsou informační kritéria využívána pro výběr správného řádu modelu. Předpokládejme, že $X_1, X_2, \ldots, X_T$ je sekvence nezávislých identicky distribuovaných náhodných veličin s hustotou pravděpodobnosti $f(\cdot|)$, kde f je model s M parametry (řád modelu), tj.,

$$\mathcal{M}(M) : \{f(\cdot|\theta) : \theta = (\theta_1, \theta_2, \ldots, \theta_M), \theta \in \Theta_M\}. \qquad (3.49)$$

Zavedeme-li prostor s omezeným počtem parametrů

$$\Theta_m = \{\theta \in \Theta_M | \theta_{m+1} = \theta_{m+2} = \ldots = \theta_M = 0\} \qquad (3.50)$$

a označíme-li odpovídající model jako $\mathcal{M}(m)$, pak dle principu informačních kritérií vyberme model, jehož hodnota informačního kritéria bude nejnižší.

Základ teorii informačních kritérií položil v roce 1973 Hirogutu Akaike (viz např. novější komentovaný přetisk [Kotz & kol., 1992]), když navrhl tzv. *Akaikovo informační kritérium* ve tvaru:

$$AIC(m) = -2 \log \ell_x(\hat{\theta}) + 2m, \; m = 1, 2, \ldots, M, \qquad (3.51)$$

kde $\ell_x(\hat{\theta})$ je funkce maximální věrohodnosti modelu $\mathcal{M}(m)$. Od té doby byla navržena celá řada jiných informačních kritérií založených na různých principech, ale

takřka vždy se skládají ze dvou částí. První je hodnota ML odhadu a druhou nějaká penalizační funkce. Jedno z nejzajímavějších je *Schwarzovo informační kritérium*

$$SIC(m) = -2\log \ell_x(\hat{\theta}) + m \log T,\ m = 1, 2, \ldots, M. \tag{3.52}$$

Ačkoliv lze změnu pozorovat pouze v penalizačním členu, kdy místo $2m$ je $m \log T$, Schwarz v roce 1978 prokázal [Schwarz, 1978], že SIC na rozdíl od AIC poskytuje alespoň asymptoticky konzistentní odhad řádu modelu. Ve stejném článku taktéž odvodil vztah týkající se testování hypotéz: asymptoticky platí

$$-2\log B_{01} = -2\log r(x) - D\log T, \tag{3.53}$$

kde B_{01} je Bayesův faktor definovaný ve vztahu (3.13), $r(x)$ je poměr věrohodností definovaný v (3.31) a D je rozdíl v dimenzi modelu hypotézy H_1 a H_0.

3.3.2.1 Fixní hranice kritického regionu

Dosazením vztahu (3.53) do nerovnosti (3.6), tj. $B_{01} < C$, získáme podmínku zamítnutí hypotézy H_0

$$\Lambda_t = -2\log r(x) - D\log T > K. \tag{3.54}$$

Dosazením rovnic (3.32) a (3.35) do výše uvedeného vztahu (3.54) po několika úpravách obdržíme testovací statistiku maximálního typu ve formě

$$\Lambda_{\hat{t}} = \max_{d<t<T-d} \left(T\log|\hat{\Sigma}| - t\log|\hat{\Sigma}_1| - (T-t)\log|\hat{\Sigma}_T| - D\log T \right), \tag{3.55}$$

kde rozdíl v dimenzi hypotézy H_1 a H_0, je-li d rozměr příznakového vektoru, je roven

$$D = d + \frac{1}{2}d(d+1). \tag{3.56}$$

Nejlepším bodem změny pak je bod $\hat{t}$, jenž maximalizuje testovací statistiku Λ_t (viz [Chen J., 2000]), což předpokládá řešení rovnice

$$\hat{t} = \arg \max_{d<t<T-d} \Lambda_t. \tag{3.57}$$

Asymptotická distribuční funkce statistiky maximálního typu $\Lambda_{\hat{t}}$ je odvozena např. v monografii [Chen J., 2000].

3.3.2.2 Fixní váha penalizační funkce

Poněkud odlišný přístup než v předchozím případě je volen v metodě s fixní váhou penalizační funkce. Jelikož se jedná o nejčastěji využívaný přístup a autorovi není známo jeho opodstatnění, neb se jedná spíše o přístup ryze heuristický, pokusí se ho podat tak, aby byl alespoň částečně teoreticky ospravedlnitelný.

SIC je někdy nazýváno *Bayesovské informační kritérium*[5] a uváděno ve formě

$$BIC(m) = \log \ell_x(\hat{\theta}) - \frac{m}{2} \log T, \; m = 1, 2, \ldots, M \tag{3.58}$$

s tím, že nejlepší model z pohledu kritéria je ten, jenž maximalizuje BIC. Dvojí definice významově úplně stejného kritéria však není zcela zbytečná, neboť lze ukázat [Chickering & Heckerman, 1996], že BIC je asymptotickou aproximací marginální distribuční funkce modelu $\mathcal{M}(m)$

$$BIC(m) \approx \log f(\mathbf{x}|m) = \log \int f(\mathbf{x}|\theta, m) f(\theta|m) d\theta \tag{3.59}$$

získané na základě *Laplaceovy metody Taylorovým rozvojem* druhého řádu v okolí bodu maxima aposteriorní pravděpodobnosti $\tilde{\theta}$ - viz příloha A.2. Jelikož se jedná o aproximaci, předpokládáme, že pro danou úlohu se lze správné hodnotě $f(\mathbf{x}|m)$ přiblížit vhodným váhováním penalizační funkce koeficientem λ. Jinými slovy: koeficientem λ nastavíme „pracovní bod" aproximace.

BIC za podmínky hypotézy H_0 lze vyjádřit ve formě

$$BIC(H_0) = -\frac{T}{2} \log |\hat{\Sigma}| - \lambda \frac{D}{2} \log T, \tag{3.60}$$

za podmínky hypotézy H_1 pak

$$BIC(H_1) = \max_{d<t<T-t} \left(-\frac{t}{2} \log |\hat{\Sigma}_1| - \frac{T-t}{2} \log |\hat{\Sigma}_T| \right) - \lambda D \log T. \tag{3.61}$$

Hypotézu H_0 zamítneme, když $BIC(H_1) > BIC(H_0)$, což znamená, že testovací statistika maximálního typu bude mít tvar

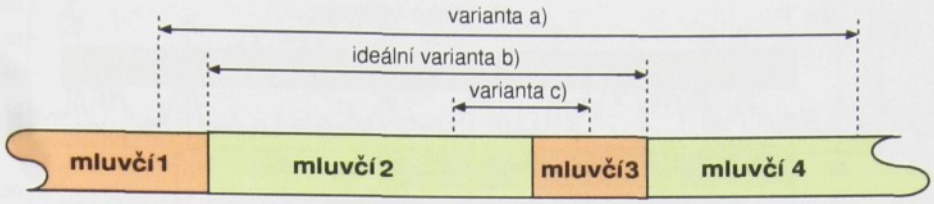
$$\Lambda_{\hat{t}} = \max_{d<t<T-t} \frac{T \log |\hat{\Sigma}| - t \log |\hat{\Sigma}_1| - T - t \log |\hat{\Sigma}_T|}{D \log T} \tag{3.62}$$

a změna nastane když $\Lambda_{\hat{t}}$ bude větší než kritická hranice K.

3.4 Detekce více bodů změny

Chceme-li řešit problematiku detekce více bodů změny pomocí nástrojů single change-point analýzy, je nutné se pozastavit nad některými aspekty, jež mohou ovlivňovat úspěšnost jednotlivých metod. Jelikož máme k dispozici signál s neznámým počtem změn, musíme umět vhodně vybírat části signálu, ve kterých bude detekce jedné změny prováděna. Konkrétní postup, jak automaticky nastavovat počátek a konec analyzujícího okna, je pak označován jako *metoda* detekce více bodů změny.

[5] Někdy také Schwarz-Bayesovo kritérium.



Obrázek 3.3: *Tři základní možnosti nastavení analyzujícího okna.*

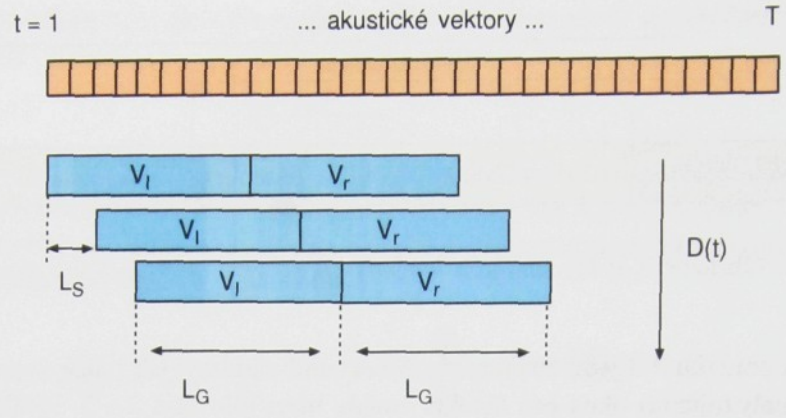
Na obrázku 3.3 jsou znázorněny 3 základní varianty, jak může vypadat nastavení analyzujícího okna pro detekci změny mezi mluvčím 2 a 3. Ideální je možnost b), neboť dané okno obsahuje pouze jednu změnu a pro rozhodnutí je využito maximální možné množství dat obou mluvčích. Tím totiž získáme nejvěrnější možný ML odhad parametrů a rozhodnutí o změně pak budeme činit s nejvyšší jistotou. Existuje několik možností, jak se k této variantě přiblížit, když na počátku s jistotou známe pouze počátek a konec signálu. Konkrétně se jedná o *metodu fixních oken*, *metodu binárního dělení* a *metodu s adaptivním oknem*. Zcela mimo tento rámec stojí *metoda globální maximalizace BIC*, u které problematiku počátku a konce analyzujícího okna není nutné řešit, neboť díky svému pojetí žádné takové okno nepotřebuje.

3.4.1 Metoda fixních oken

V odborných článcích věnovaných praktickým implementacím detektorů změny řečníka je nejčastěji zmiňována on-line *metoda fixních oken*. Vysoká míra oblíbenosti této metody je zapříčiněna pravděpodobně přímočarostí, s jakou lze obdržet vcelku rozumné výsledky bez větší znalosti dané oblasti. Problematickou se tato metoda stává při jejím trénování, neboť má příliš mnoho volných parametrů. Nicméně je hojně využívaná a v literatuře lze nalézt desítky variací na toto téma.

Metoda je založena na principu měření statistické vzdálenosti mezi dvěma sousedícími sekvencemi akustických vektorů - viz obr. 3.4. Po signálu se posunuje dvojice oken o délce L_G s krokem L_S. Toto umožňuje měřit v každém čase t vzájemnou vzdálenost $D(t)$ odpovídajících sad příznakových vektorů $V_l = x(t - L_G + 1), \ldots, x(t)$ a $V_r = x(t + 1), \ldots, x(t + L_G)$. Za body změny mluvčího jsou považována lokální maxima křivky $D(t)$. Na obrázku 3.5 jsou uvedeny její průběhy pro případ, kdy je jako míra statistické vzdálenosti použita značně populární *Kullback-Leiblerova vzdálenost*. Její definice je uvedena například v knize [Huang & kol., 2001]. Velmi oblíbeným tématem je testování různých statistických měr - např. článek [Pietquin & kol., 2001], případně testování jejich různých modifikací - viz např. [Kwon & Narayanan, 2002].

Protože se jedná o první metodu, kterou se autor zabýval [Zdansky & kol., ICSLP 2004], dovoluje si níže shrnout problémy, s nimiž se setkal při její implementaci.



Obrázek 3.4: *Princip metody fixních oken, respektive její první úrovně.*

1. Lokální maxima křivky $D(t)$ neposkytují příliš přesný odhad bodu změny.

2. Z obrázku 3.5 je patrné, že jednoduchá není ani volba délky okna L_G. Příliš krátké okno produkuje příliš mnoho falešných maxim, dlouhé okno naopak zase způsobuje nedetekovatelnost krátkých segmentů.

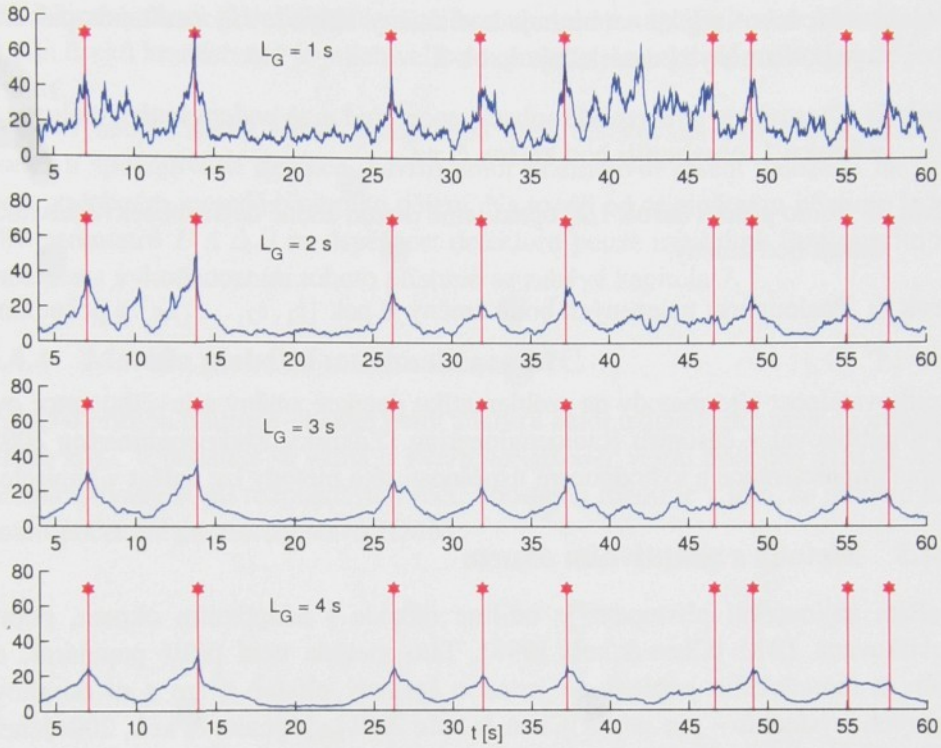
3. Metoda vyžaduje detektor lokálních maxim, jejichž detekce na nehladké křivce $D(t)$ není zrovna triviální záležitost.

4. Zakomponováním detektoru lokálních maxim do procesu detekce bodů změny přibývají volné parametry systému, jež je nutné natrénovat.

5. Neexistuje rozumný způsob odhadu těchto parametrů, tj. je nutné zkusit všechna možná nastavení parametrů a vybrat to, jež na trénovacích datech poskytne nejlepší výsledky. Čím více volných parametrů, tím déle trvá jejich odhad.

Výše uvedené problémy se často řeší tím, že detekovaná lokální maxima nejsou považována přímo za body změny, ale pouze za tzv. *kandidáty* na bod změny. Pak hovoříme o *dvojúrovňovém* systému, kdy v první úrovni jsou nalezeni kandidáti na bod změny, v druhé jsou pak tyto kandidáti buď potvrzeni nebo zamítnuti, většinou na principu testování hypotéz pomocí BIC s fixní váhou penalizační funkce - viz část 3.3.2.2. Pro takovýto dvojúrovňový systém se vžilo označení DISTBIC [Delacourt & Wellekens, 2000] a v literatuře se setkáme s celou řadou jeho rozličných implementací - viz např. [Lu & Zhang, 2002] či [Vandecatseye & Martins, 2003]:

Odladění této metody je velmi komplikované, neboť ke své činnosti potřebuje detektor lokálních maxim. Kandidáti na bod změny nalezené tímto detektorem slouží pouze pro nastavení hranic analyzující okna. S tímto nastavením je pak v druhé úrovni uskutečněna klasická detekce jednoho bodu změny. Přitom *metoda*



Obrázek 3.5: *Průběh hodnoty Kullback-Leiblerovy vzdálenosti v závislosti na čase pro různé délky oken L_G. Červeně jsou označeny skutečné body změny .*

s adaptivním oknem dělá za cenu nepatrně vyšších výpočetních nároků totéž v jednom průchodu a bez nutnosti implementace detektoru maxim. Z autorova pohledu je metoda fixních oken nepraktická a místo ní preferuje on-line metodu s adaptivním oknem - viz kapitola 7.

3.4.2 Metoda binárního dělení

Metodu binárního dělení navrhla a její konzistenci ověřila již v roce 1981 Ruska Vostrikova [Vostrikova, 1981]. Zde je nutné podotknout, že tato metoda zůstala v oblasti detekce změny řečníka zcela bez povšimnutí, ačkoliv je v ostatních oblastech brána jako základní způsob aplikace single change-point analýzy na řešení problému vícenásobné změny - viz např. [Chen J., 2000], [Lauro & kol, 2002]. Velkou výhodou této metody je její snadná trénovatelnost, neboť je nutno odhadnout hodnotu pouze jediného volného parametru - kritické hranice zamítnutí nulové hypotézy. Mezi další výhody patří nízká výpočetní náročnost.

Princip binárního dělení lze shrnout do následujících kroků.

krok 1: Otestujeme signál zdali obsahuje či neobsahuje bod změny, tj. H_0 vs. H_1 dané vztahy (3.29) a (3.30). Pakliže je nulová hypotéza přijata, algoritmus

končí, neboť signál neobsahuje bod změny. Když je H_0 zamítnuta, pak existuje bod změny t_s a následuje **krok 2**.

krok 2: Otestujeme dvě vzniklé subsekvence před a za bodem změny nalezeným v **kroku 1**, obsahují-li bod změny či ne.

krok 3: Tento proces (**krok 1,2**) opakujeme dokud žádné další subsekvence neobsahují bod změny.

krok 4: Posloupnost nalezených bodů změny je pak $\{t_1, t_2, \ldots, t_S\}$ a počet změn je S.

Aplikovatelnost této metody na problematiku detekce změny mluvčího autor ověřil a publikoval v časopisu Radioengineering [Zdansky, Radioengineering 2005]. Popis implementace a vyhodnocení úspěšnosti této metody lze nalézt v kapitole 5.

3.4.3 Metoda s adaptivním oknem

Dalším zajímavým přístupem je on-line metoda s adaptivním oknem, poprvé publikovaná IBM [Chen & kol., 1998]. Tato metoda není příliš populární, ačkoliv z teoretického pohledu se jedná o správný přístup a lze s ní dosahovat dobrých výsledků - viz např. [Chen & kol., 2001], [Ajmera & kol., 2004] nebo [Zhou & Hansen, 2005]. Tato metoda je založena na hledání jednoho bodu změny v okénku posouvajícím se po signálu. Velikost a poloha okénka je přitom řízena na základě algoritmu sestávajícího z následujících 4 kroků.

krok 1: Inicializujeme interval $[a, b]$, kde $a = 1$ a b je nějaká minimální délka inicializačního okna, tj. $b = B$ příznakových vektorů.

krok 2: Testujeme hypotézu zda interval $[a, b]$ obsahuje či neobsahuje jeden bod změny, v případě originálního článku [Chen & kol., 1998] se jednalo o BIC s fixní váhou penalizační funkce.

krok 3: Pakliže není žádná změna v intervalu $[a, b]$ nalezena, zvětšíme rozsah testovaného okna na $b = b + \Delta B$. V případě, že byl nalezen bod změny t_s, pak počátek testovacího okna nastavíme na $a = t_s + 1$ a konec na $b = a + B$.

krok 4: Pokud není konec signálu, přejdeme ke **kroku 2**.

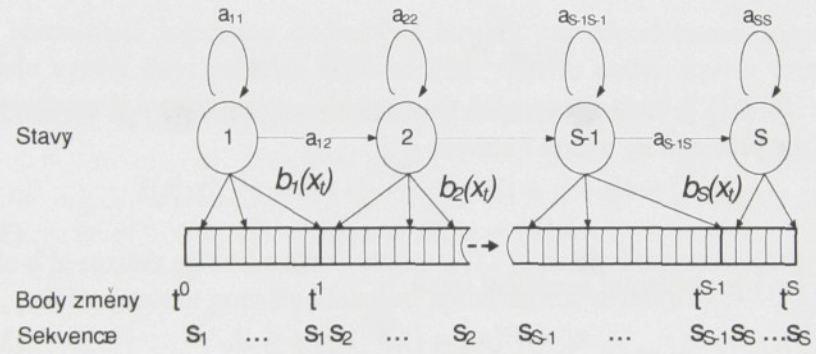
Problematickou částí této metody je možnost jejího natrénování. Opomeneme-li otázku odhadu hranice kritického regionu K, pak zbývají ještě dva volné parametry metody. Velikost inicializačního okna B a koeficient rozšíření ΔB. Čím nižší bude hodnota obou parametrů, tím větší bude výpočetní náročnost. Větší nepříjemností však je skutečnost, že se zmenšováním těchto parametrů bude klesat spolehlivost detekce, neboť hrozí riziko nevyužití maximálního množství dat pro spolehlivý odhad kovariančních matic. To znamená, že učiníme-li rozhodnutí o změně bez maximálního možného množství dat, bude učiněno s menší jistotou, než v opačném případě, z čehož jasně vyplývá vyšší očekávaná chybovost. Na druhou stranu,

příliš široké okno či velký krok rozšiřování mohou zapříčinit nedetekovatelnost hranic úseků kratších než je jejich velikost, protože je očekávána vždy pouze jedna změna.

Aby nebylo nutné hledat kompromis mezi detekovatelností hranic krátkých úseků a spolehlivostí detekce, navrhl autor alternativní řešení založené na principu podobném metodě binárního dělení. Na rozdíl od originálního přístupu je pak vliv parametrů B a ΔB na úspěšnost detektoru pouze minoritní. Popisem implementace a vyhodnocením tohoto přístupu se zabývá kapitola 7.

3.4.4 Metoda globální maximalizace BIC

Pokus o přímou multiple-change point analýzu autor označil jako metodu maximalizace BIC. V podstatě se jedná o *Viterbiho dekodér* velmi podobný HMM dekodérům užívaných při rozpoznávání řeči. Podstatný rozdíl je v tom, že předem není znám počet ani parametry stavů HMM.



Obrázek 3.6: *Vyhodnocení pravděpodobnosti segmentace jako pravděpodobnosti sekvence stavů HMM.*

Předpokládejme, že máme k dispozici záznam audio nahrávky skládající se z nepřekrývajících se částí, jež odpovídají jednotlivým mluvčím. Úkolem je identifikovat hranice těchto úseků, což znamená nalézt dílčí body změny, přičemž neznáme jejich polohu ani jejich počet.

3.4.4.1 Formální definice

Nechť máme data

$$\mathbf{x} = \{x_1, x_2, \ldots, x_T\}, \tag{3.63}$$

což je sekvence T příznakových vektorů neznámého stochastického procesu. Nadefinujeme rostoucí posloupnost přirozených čísel

$$\mathbf{t}_i^S = \{t_i^0, t_i^1, t_i^2, \ldots, t_i^S\} \tag{3.64}$$

a označíme ji jako i-tou *segmentaci* S-tého řádu, kde t_i^s je pozice konce s-tého segmentu. Zřejmě platí, že *segmentace* musí splňovat podmínku

$$t_i^0 < t_i^1 < t_i^2 < \cdots < t_i^S \tag{3.65}$$

kde $t_i^0 = 0$, $t_i^S = T$, pro libovolné i.

3.4.4.2 Model segmentace

Pojmeme-li úlohu segmentace formálněji, můžeme ji popsat následovně: majíce sekvenci dat, hledáme její rozdělení do takových částí (subsekvencí), aby bylo možné o každých dvou sousedících tvrdit, že pochází ze vzájemně nezávislých zdrojů. Předpokládejme tedy, že lze o jednotlivých částech prohlásit, že byly vygenerovány dílčími stavy S-stavového *skrytého Markovova modelu* prvního řádu (viz obr. 3.6), jenž je plně charakterizován sadou parametrů $\Theta = (\mathbf{A}, \mathbf{B}, \pi)$.

- $\mathbf{A} = \{a_{ij}\}$ je matice pravděpodobnosti přechodů, kde a_{ij} jsou pravděpodobnosti přechodů ze stavu i do stavu j, tj.

$$a_{ij} = P(s_t = j | s_{t-1} = i). \tag{3.66}$$

- $\mathbf{B} = \{b_i(k)\}$ je matice výstupních věrohodností, kde $b_i(k)$ je věrohodnost emitace příznaku x_k v čase t stavem i, tj.

$$b_i(k) = p(X_t = x_k | s_t = i). \tag{3.67}$$

- $\pi = \{\pi_i\}$ je počáteční rozdělení stavů, tj.

$$\pi_i = P(s_0 = i), \quad 1 \leq i \leq S. \tag{3.68}$$

Pro data $\mathbf{x}$ a segmentaci $\mathbf{t}_i^S$ předpokládejme, že pravděpodobnost segmentace odpovídá pravděpodobnosti sekvence stavů $\mathbf{S}_i^S = \{s_1, s_1, \ldots, s_s, s_s, \ldots, s_S\}$ modelu, tj. $P(\mathbf{t}_i^S|\mathbf{x}) = P(\mathbf{S}_i^S|\mathbf{x})$, což lze pomocí *Bayesova vzorce* rozepsat jako

$$P(\mathbf{S}_i^S|\mathbf{x}) = \frac{p(\mathbf{x}|\mathbf{S}_i^S)P(\mathbf{S}_i^S)}{p(\mathbf{x})}, \tag{3.69}$$

kde

$$p(\mathbf{x}|\mathbf{S}_i^S) = p(\mathbf{x}|\mathbf{t}_i^S, \Theta) = \prod_{s=1}^{S} p(\mathbf{x}_s|\theta_i^s), \tag{3.70}$$

a θ_i^s je neznámý vektor parametrů s-tého stavu a $\mathbf{x}_s = \{x_{t_i^{s-1}+1}, \ldots, x_{t_i^s}\}$ jsou data odpovídající s-tému segmentu. Budeme-li předpokládat *levo-pravý* HMM[6], pak lze apriorní pravděpodobnost segmentace vyjádřit jako

$$P(\mathbf{S}_i^S|\Theta) = \pi_0 \prod_{s=1}^{S} a_{ss}^{t_i^s - t_i^{s-1}} (1 - a_{ss}), \tag{3.71}$$

kde a_{ss} jsou pravděpodobnosti setrvání ve stavu s.

[6]Levo-pravý HMM znamená, že nejsou povoleny přechody zprava doleva, tj. $a_{ij} = 0 \; \forall i > j$.

3.4.4.3 Optimální segmentace

K nalezení optimální segmentace využijeme *Bayesovského klasifikátoru s minimální chybovostí*, což znamená, že hledáme řešení založené na maximalizaci aposteriorní pravděpodobnosti $P(\mathbf{S}_i^S|\mathbf{x})$.

$$t_{opt} \equiv S_{opt} = \arg\max_{\forall S \forall i} p(\mathbf{x}|\mathbf{S}_i^S) P(\mathbf{S}_i^S). \tag{3.72}$$

Maximalizací výše uvedeného vztahu přes všechny možné polohy bodů změn a všechny možné počty segmentů tedy získáme optimální segmentaci.

3.4.4.4 Aproximace sdružené pravděpodobnosti

Problém je, že parametry HMM Θ neznáme. Pokusíme se tedy nalézt alespoň aproximaci logaritmu[7] sdružené pravděpodobnosti

$$\log P(\mathbf{x}, \mathbf{S}_i^S) = \log p(\mathbf{x}|\mathbf{S}_i^S) + \log P(\mathbf{S}_i^S). \tag{3.73}$$

K aproximaci logaritmu podmíněné hustoty pravděpodobnosti $\log p(\mathbf{x}|\mathbf{S}_i^S)$ můžeme využít *Bayesovského informačního kritéria* definovaného vztahem (3.58). Zavedeme-li logaritmickou věrohodnost dat v pozici mezi t_i^{s-1} a t_i^s

$$\ell(t_i^s, t_i^{s-1}) = \frac{t_i^{s-1} - t_i^s}{2} \left[\log |\Sigma| + d + d \log (2\pi)\right], \tag{3.74}$$

kde d je rozměr příznakového vektoru a $|\Sigma|$ je determinant kovarianční matice dat $\mathbf{x}_s$, pak lze hustotu pravděpodobnosti aproximovat ve tvaru

$$\log p(\mathbf{x}|\mathbf{S}_i^S) \approx \mathcal{BIC}(\mathbf{x}|\mathbf{S}_i^S) = \sum_{s=1}^{S} \ell(t_i^s, t_i^{s-1}) - \lambda \frac{Sc}{2} \log T. \tag{3.75}$$

λ je váha penalizační funkce a proměnná c značí počet volných parametrů gaussovského modelu. c lze spočítat z rozměru d jako

$$c = d + \frac{1}{2} d(d+1). \tag{3.76}$$

Apriorní pravděpodobnost sekvence stavů je možné získat odhadem z trénovacích dat tím, že budeme předpokládat *degenerovaný* HMM s pravděpodobností setrvání ve stavu sdílenou pro všechny stavy. Pokud bychom uvažovali *semi-Markovův skrytý model* [Ferguson, 1980], bylo by možné pravděpodobnost setrvání ve stavu modelovat např. *logaritmicko-normálním* nebo *gamma* rozložením. V této práci se však autor omezil na klasický HMM, jenž má *exponenciální* model setrvání ve stavu s jediným parametrem a. Pro degenerovaný HMM tedy platí, že

$$a_{ss} = a \quad \forall s = 1, \ldots, S. \tag{3.77}$$

[7] Z výpočetního hlediska je výhodnější maximalizovat logaritmus sdružené pravděpodobnosti. Logaritmus je monotónní rostoucí funkce, takže na výsledku maximalizační úlohy se nic nezmění.

Dosazením do rovnice (3.71) a logaritmováním získáme vztah

$$\log P(\mathbf{S}_i^S|\Theta) = \log \pi + (T - S)\log a + S\log(1 - a). \tag{3.78}$$

Jelikož při odvození BIC jako aproximace marginální věrohodnosti byly zanedbávány všechny členy, jež nerostou s T - viz A.1, zanedbáme je i zde. Vzhledem k tomu, že a značí pravděpodobnost setrvání ve stavu, lze ji zhruba odhadnout jako dobu kdy nedochází ke změně k celkové délce signálu $a = 1 - S/T$. Poněvadž $S \ll T$, bude mít apriorní informace o segmentaci v této podobě zanedbatelný vliv na výsledek a celý proces segmentace se bude řídit rovnicí

$$t_{opt} = \arg\max_{\forall S \forall i} \left[\sum_{s=1}^{S} \ell(t_i^s, t_i^{s-1}) - \lambda \frac{Sc}{2} \log T \right]. \tag{3.79}$$

Aplikovatelnost této metody na problematiku detekce změny mluvčího autor ověřil a publikoval na mezinárodní konferenci Interspeech 2005 v Lisabonu [Zdansky & Nouza, Interspeech 2005]. Popisu efektivní implementace a vyhodnocení úspěšnosti této metody se věnuje kapitola 6.

KAPITOLA 4

DATABÁZE A METODY VYHODNOCENÍ ÚSPĚŠNOSTI SEGMENTACE

Trénování a testování navržených metod detekce změny řečníka bylo realizováno na pěti různých databázích. Zdrojem prvních čtyř „uměle namíchaných" databází se staly záznamy různých pořadů českých televizních a rozhlasových stanic. Tyto pořady byly pečlivě anotovány a údaje o počátku a konci promluvy jednotlivých mluvčích byly využity jako zdroj pro vznik umělé databáze ART. Jelikož tato databáze obsahovala „příliš reálná data", což znamená, že se v ní vyskytovaly i různé neřečové části a aditivní rušení na pozadí řečového signálu, byly vytvořeny navíc „ideální" databáze S-ART, kde byly tyto jevy v maximální míře potlačeny. Konkrétně se jedná o databázi FS-ART, složenou pouze z ženských hlasů, MS-ART, jež obsahuje pouze mužskou řeč a smíšenou databázi S-ART. Dalším dobrým důvodem tvorby umělých databází byla potřeba mít trénovací a testovací data, u nichž jsou zcela jednoznačně známy přesné pozice bodů změny. Pro testování navržených metod na reálných datech pak posloužila panevropská databáze televizních zpráv, která vznikla v rámci projektu Evropské Unie COST 278.

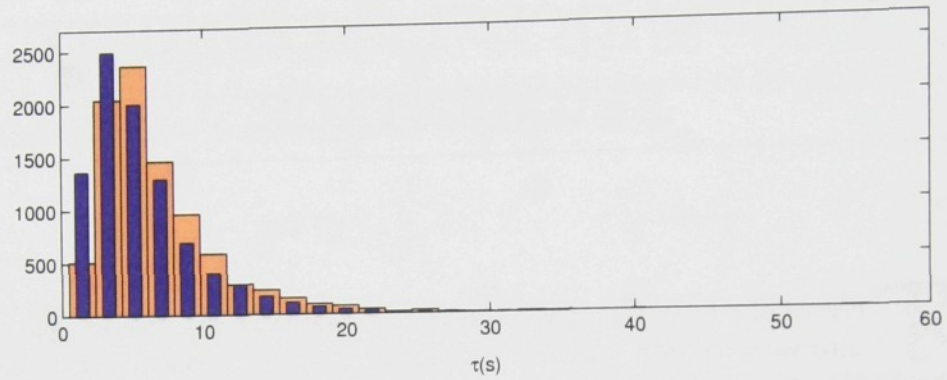
4.1 Databáze ART

	Počet změn					Délka segmentu (s)		
Část	$\sum$	M-F [%]	F-F [%]	M-M [%]	N	ϕ	max	min
Trén.	8788	48.38	18.57	33.05	100	6.81	55.99	0.51
Test.	8949	50.21	18.44	31.36	100	6.67	57.32	0.54

Tabulka 4.1: *Základní údaje o databázi ART.*

Jako zdroj databáze ART posloužilo 5346 segmentů pocházejících od 456 mluvčích uložených ve formátu WAV vzorkovaných na frekvenci 16 kHz. Jednotlivé segmenty byly očíslovány a pomocí generátoru náhodných čísel byly pospojovány do úseků o délce přibližně 10 minut. Tímto způsobem vznikla trénovací a

testovací část - každá o rozsahu $N = 100$ položek. Základní údaje o databázi ART jsou shrnuty v tabulce 4.1, histogram délek segmentů je na obrázku 4.1. Pro informaci jsou v tabulce uvedeny procentuální zastoupení změn typu muž–muž (M–M), žena–žena (F–F) a muž–žena (M-F). Symbolem sumace je označen celkový počet změn a ϕ je údaj o průměrné délce segmentu.



Obrázek 4.1: *Histogram délek segmentů trénovací (světle) a testovací (tmavě) části databáze ART.*

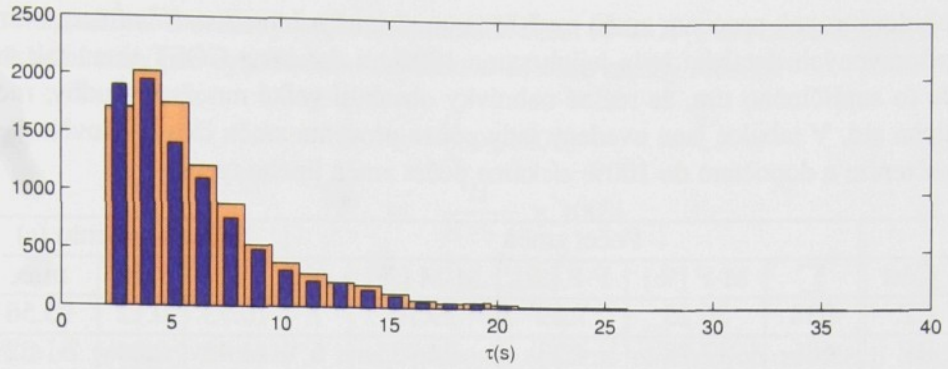
4.2 Databáze S-ART

Při vytváření databáze S-ART bylo využito stejných zdrojů jako u databáze ART, pouze z nich byly vyřazeny ty položky, jež obsahovaly jakékoliv aditivní rušení na pozadí, čímž se zredukovalo množství zdrojových dat na 3895 segmentů od 428 mluvčích. Jelikož měl autor k dispozici podrobný fonetický popis dílčích segmentů a odpovídající akustické modely, využil metody automatického *vynuceného zarovnání* časových hranic fonémů[1] k odstranění neřečových částí z jednotlivých zdrojových segmentů. Cílem této operace bylo umožnit vyhodnocení přesnosti detekce bodu změny z hlediska jeho časové polohy. Detailnější informace o databázi S-ART lze nalézt v tabulce 4.2 a histogram délek segmentů na obrázku 4.2.

	Počet změn					Délka segmentu (s)		
Část	$\sum$	M-F [%]	F-F [%]	M-M [%]	N	ϕ	max	min
Trén.	9455	49.20	18.30	32.50	100	6.32	39.94	2.01
Test.	9468	49.00	19.10	31.90	100	6.31	39.94	2.00

Tabulka 4.2: *Tabulka základních údajů o databázi S-ART.*

[1] Forced alignment.



Obrázek 4.2: *Histogram délek segmentů trénovací (světle) a testovací (tmavě) části datbáze S-ART.*

4.3 Databáze FS-ART a MS-ART

Databáze FS-ART a MS-ART vznikly za stejných podmínek a ze stejných zdrojů jako databáze S-ART. Jediný rozdíl je ten, že do databáze FS-ART byly vybírány pouze osoby ženského a do MS-ART pouze mužského pohlaví. Obě byly vytvořeny z důvodu ověření, nakolik úspěšnost detekce změny souvisí s pohlavím sousedících mluvčích. Základní údaje jsou shrnuty tabulce 4.3, histogram délek segmentů je obdobný jako u databáze S-ART a tudíž není uváděn.

Databáze	Část	Počet změn		Délka segmentu (s)		
		$\sum$	N	ϕ	max	min
FS-ART	Trénovací	9814	100	6.10	35.34	2.00
	Testovací	9917	100	6.03	35.34	2.00
MS-ART	Trénovací	9129	100	6.55	39.94	2.00
	Testovací	9146	100	6.54	39.94	2.00

Tabulka 4.3: *Základní údaje o databázi FS-ART a MS-ART.*

4.4 Databáze COST 278

Databáze reálných nahrávek vznikla v rámci aktivity 278[2] evropského programu COST, jehož se v době psaní této práce účastnilo 10 institucí z 9 různých zemí. Každá instituce dodala 3 hodiny víceúrovňově anotovaných dat, jimiž byly audiovizuální záznamy národních televizních zpráv. Popisem databáze se zabývá článek publikovaný na konferenci LREC 2004 [Vandecatseye & kol, 2004].

Parametry jednotlivých komponent zajímavých z hlediska detekce změny mluvčího jsou uvedeny v tabulce 4.4. Za povšimnutí stojí sloupec počtu změn,

[2]Spoken Language Interaction in Telecommunication

především pak procenta změn muž-žena, muž-muž a žena-žena. Zatímco u uměle připravených databází byla jejich suma 100%, u databáze COST tomu tak není. Je to zapříčiněno tím, že reálné nahrávky obsahují velké množství hudby, ruchů, ticha atd. V tabulce jsou uvedeny tedy pouze procenta změn člověk-člověk, jejich sečtením a dopočtem do 100% získáme počet změn jiného typu.

	Počet změn					Délka segmentu (s)		
Část	$\sum$	M-F [%]	F-F [%]	M-M [%]	N	ϕ	min	max
CZ	938	36.25	7.89	29.10	7	10.73	0.12	93.56
BE	686	17.78	4.23	24.93	6	11.49	0.34	81.52
GA	544	34.56	19.85	17.65	3	17.40	0.50	292.74
GR	752	19.95	3.72	18.09	3	12.73	0.27	141.14
HR	571	40.11	10.51	34.33	6	15.95	0.23	110.12
HU	916	19.43	3.38	15.61	11	10.80	0.47	113.70
PT	1286	11.90	1.40	10.19	6	9.12	0.03	398.95
SI	701	34.09	6.85	15.98	3	14.10	0.15	150.90
SI2	528	25.00	2.08	32.01	3	13.29	0.40	144.28
SK	1142	6.48	2.71	9.54	9	9.16	0.01	107.92
$\sum$	8073	22.36	5.43	19.03	57	11.79	0.01	398.95

Tabulka 4.4: *Základní údaje o databázi COST 278.*

Kladem této databáze je rozsáhlost a velká variabilita spojená s různými jazyky a národními zvláštnostmi v televizních programech. Určitým negativním rysem je fakt, že každá národní část byla zpracována jiným anotátorem a ne každý se při své práci držel zcela striktně pravidel. Anotace dílčích částí je tudíž občas poněkud nekonzistentní a někdy se tím pádem v databázi objevují chyby.

4.5 Metody vyhodnocení výsledků

Vyhodnocení výsledků je běžně praktikováno formou obousměrného hledání nejbližšího souseda mezi referenčními a vypočtenými body změny. *i-tý* vypočtený bod změny t_{ci} považujeme za správně nalezený (HIT) a odpovídající *j-tému* referenčnímu bodu změny t_{rj} tehdy a jen tehdy, když

1. t_{ci} je vypočtený bod změny nejblíže referenčnímu t_{rj},
2. t_{rj} je referenční bod změny nejblíže vypočtenému t_{ci},
3. vzdálenost mezi nimi je menší než určitá mez τ_{max}, typicky $|t_{ci} - t_{rj}| < 1\ s$.

O takovýchto dvojicích pak tvrdíme, že tvoří *pár* a jejich počet označíme H. Všechny vypočtené body, jež netvoří pár, označujeme jako INZERCE a jejich počet I. Obdobně pak pro referenční body změny, jež nebyly nalezeny (netvoří pár), se používá označení DELECE, pro jejich počet D.

Označíme-li celkový počet referenčních bodů $N = H + D$, lze nadefinovat tři základní míry pro vyhodnocení úspěšnosti detekce změny řečníka:

$$R = \frac{H}{N} \times 100\% \tag{4.1}$$

$$P = \frac{H}{H+I} \times 100\% \tag{4.2}$$

$$F = \frac{2 \times R \times P}{R+P}. \tag{4.3}$$

Míra R se nazývá *recall* a značí procento správně nalezených ze všech hledaných bodů změny. V mezním případě, umístí-li detektor bod změny do každého diskrétního časového okamžiku, se bude recall blížit ideálním 100%, což ovšem neznamená, že máme k dispozici kvalitní detektor. Z tohoto důvodu se navíc používá míra zvaná *precision* (P), která vyjadřuje procento správně nalezených ze všech nalezených bodů změny. Tyto dvě míry jsou protichůdné, tj. roste-li jedna, klesá druhá a naopak. Avšak ani jedna z těchto měr nemá lokální maximum, pročež nejsou vhodné jako kritéria pro trénování detektoru. Toto je důvodem zavedení míry zvané *F-rate* (F), která tuto podmínku splňuje.

Dalším hlediskem vyhodnocení detektoru změn je přesnost poloh správně nalezených bodů změny. Pro tento účel využijeme histogram chyb časového zarovnání

$$\Delta t_h = |t_{ci} - t_{rj}|, \tag{4.4}$$

kde $h = 1, \ldots, H$ a H je celkový počet správně detekovaných změn. Z histogramu odečteme tři hodnoty vypovídající o přesnosti detektoru:

1. $\Delta_{2/3}$ označíme maximální chybu časového zarovnání pro 2/3 všech správně detekovaných bodů změny;
2. $\Delta_{0.95}$ označíme maximální chybu časového zarovnání pro 95% všech správně detekovaných bodů změny;
3. δ_{10} označíme procento správně detekovaných bodů změny, jejichž chyba zarovnání je menší než 10 ms.

4.5.1 Test statistické signifikance

Předpokládejme, že máme k dispozici výsledky 2 systémů (metod) získaných na stejných datech. Pro oba jsme obdrželi velmi podobné výsledky a předmětem zájmu je rozhodnout, zdali je rozdíl mezi oběma systémy statisticky významný. Zkoumáme tedy, zdali je systém **A** lepší než systém **B**, k čemuž využijeme test statistické signifikance.

Nechť E_A^p, E_B^p značí *chybovost* systému **A**, **B** měřenou na p-té položce databáze. Chybovost E_X^p vypočteme jako poměr počtu chyb ku očekávanému počtu změn N

$$E_X^p = \frac{D_X^p + I_X^p}{N}, \tag{4.5}$$

kde D_X^p, I_X^p značí počet delecí respektive inzercí v odpovídající položce databáze. Zavedeme náhodnou rozdílovou veličinu

$$Z = E_B - E_A, \tag{4.6}$$

o níž budeme předpokládat, že má normální rozdělení. Nyní lze vyslovit hypotézu H_0: systém **A** je stejný nebo horší než systém **B**, oproti alternativní H_1: systém **A** je lepší než **B**. Neboť pojmem horší rozumíme vyšší chybovost, plyne z této formulace, že za podmínky H_0 bude $E(Z) \leq 0$. Pro normální rozdělení je matematické očekávání asymptoticky rovno ML odhadu střední hodnoty μ. Budeme tedy testovat následující hypotézy:

$$H_0 : \mu \leq 0 \tag{4.7}$$

$$H_1 : \mu > 0. \tag{4.8}$$

V monografii [Lehmann, 1986] je pak odvozena patřičná testovací statistika

$$r(z) = \frac{\sqrt{P}\bar{z}}{\sqrt{\frac{1}{P-1}\sum_{p=1}^{P}(z_p - \bar{z})^2}}, \tag{4.9}$$

kde P je počet testovaných položek databáze a $\bar{z}$

$$\bar{z} = \frac{1}{P}\sum_{p=1}^{P} z_p \tag{4.10}$$

je aritmetický průměr hodnot z_p. Ve stejné knize je také dokázáno, že $r(z)$ má normální rozdělení $R \propto N(0,1)$.

Systém **A** lze považovat za lepší oproti systému **B** na úrovni signifikance α, tj. zamítneme H_0, když

$$P\{R \geq r(z)\} \leq \alpha. \tag{4.11}$$

Čím nižší zvolíme α, s tím vyšší jistotou zamítáme H_0. Mnohem praktičtější je ekvivalentní formulace ve tvaru

$$r(z) \geq u_\alpha, \tag{4.12}$$

kde u_α se nazývá α–procentním kvantilem a lze ho vypočíst z rovnice

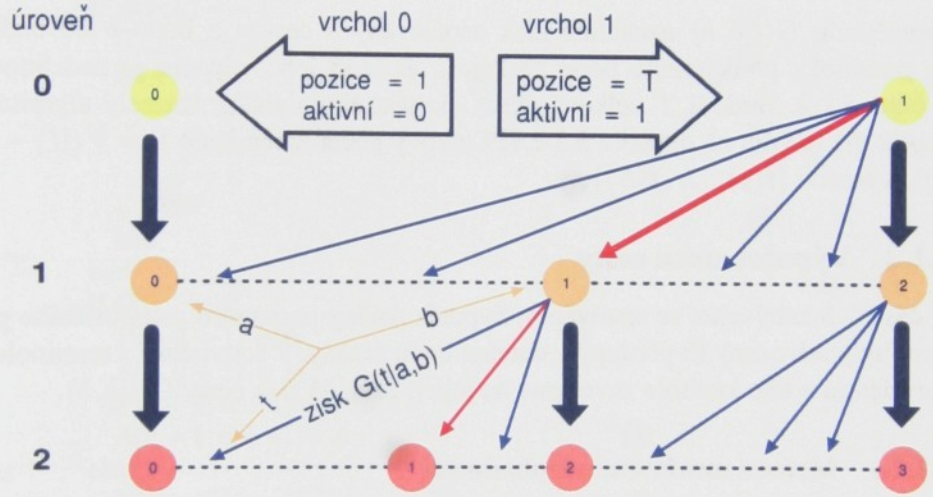
$$\alpha = 1 - P\{R < u_\alpha\}. \tag{4.13}$$

Pro normální rozložení je tato rovnice analyticky neřešitelná a hodnoty kvantilů jsou tabelovány.

KAPITOLA 5

METODA BINÁRNÍHO DĚLENÍ

Princip binárního dělení byl zjednodušeně popsán již v části 3.4.2. V této kapitole bude upřesněn popis metody s ohledem na její implementaci, možnosti trénování a následně bude provedeno vyhodnocení na databázích popsaných v kapitole 4.



Obrázek 5.1: *Grafické znázornění metody binárního dělení.*

5.1 Princip binárního dělení

Princip metody binárního dělení je znázorněn na obrázku 5.1. Nadefinujeme si nultou *úroveň* procesu detekce změny řečníka se dvěma *vrcholy* odpovídajícími počátku a konci akustického signálu. Každému vrcholu přiřadíme dvě *vlastnosti*. První je pozice vrcholu, druhou pak vlastnost, zdali je vrchol *aktivní*, tj. je-li možné přejít z daného vrcholu do následující úrovně. Tenké šipky na obrázku značí *cestu* z jedné úrovně do úrovně následující a *zisk* daný tímto přechodem odpovídá výpočtu rovnic uvedených v části 5.1.2. V souladu s teorií popsanou v části 3.3 vybereme tu cestu, jež přináší nejvyšší zisk. Když bude zisk větší než určitá kritická hranice K, ustanovíme nový vrchol v následující úrovni a nastavíme mu následující vlastnosti: *pozice* bude odpovídat místu, kam vedla nejvýhodnější cesta a

aktivní = 1. Poté zkopírujeme testovaný vrchol do následující úrovně. Pakliže zisk nejlepší cesty nepřekročí kritickou hranici K, nebude nový vrchol ustaven, dojde pouze ke zkopírování původního do nové úrovně a změní se nastavení vlastnosti *aktivní* = 0. Algoritmus ukončíme na té úrovni, kde již nebude žádný *aktivní* vrchol.

5.1.1 Návrh algoritmu

Označme číslo úrovně písmenem $u = 0, \ldots, U$ číslo vrcholu $v = 0, \ldots, V(u)$. Dále nadefinujeme dvě dvourozměrná pole vlastností:

1. $A(v, u) = 1$, když vrchol v na úrovni u bude *aktivní*, v opačném případě bude nabývat hodnoty 0;

2. $P(v, u)$ ponese informaci o *pozici* vrcholu v na úrovni u.

Proměnnou $G(t|a, b)$ označíme zisk asociovaný s cestou z bodu b do bodu t za podmínky předchozího bodu změny a. Způsob jeho výpočtu je nadefinován v části 5.1.2. Bude-li T celková délka akustického signálu, lze celý algoritmus zapsat dle návrhu na obrázku 5.1.2.1. Celkový počet změn bude $S = V(U) - 1$ a jejich pozice $\{\hat{t}(1), \ldots, \hat{t}(S)\}$.

5.1.2 Výpočet zisku cesty

V části 3.3 zabývající se testováním hypotéz změny parametrů gaussovského procesu byly odvozeny tři přístupy k detekci bodu změny. V souvislosti s terminologií zavedenou v této kapitole zavedeme na jejich základě zisk cesty $G(t|a, b)$.

5.1.2.1 Metoda maximální věrohodnosti

Z přístupu metodou maximální věrohodnosti vyplynula testovací statistika maximálního typu definovaná vzorcem (3.48). Zisk G cesty z bodu b do bodu t lze z tohoto vzorce odvodit ve tvaru

$$G_{MLLR}(t|a, b) = \alpha \sqrt{\left[(b - a + 1) \log |\hat{\Sigma}| - (t - a + 1) \log |\hat{\Sigma}_1| - (b - t) \log |\hat{\Sigma}_T|\right]} - \beta, \tag{5.1}$$

když $a - d > t > b + d$ a d je rozměr příznakového vektoru. Proměnné α a β vychází ze vzorců (3.45), (3.46) a lze je vyjádřit jako:

$$\alpha = (2 \log \log (b - a + 1))^{\frac{1}{2}} \text{ a} \tag{5.2}$$

$$\beta = 2 \log \log (b - a + 1) + d \log \log \log b - a + 1) - \log \Gamma(d). \tag{5.3}$$

Matice $\hat{\Sigma}$, $\hat{\Sigma}_1$, $\hat{\Sigma}_T$ jsou kovariance dat $\{x_a, \ldots, x_b\}$, $\{x_a, \ldots, x_t\}$, $\{x_{t+1}, \ldots, x_b\}$. Metoda využívající funkci G_{MLLR} bude v dalším textu označována zkratkou MLLR (Maximum Log-Likelihood Ratio).

Inicializace:

$A(0,0) = 0;$
$P(0,0) = 1;$
$A(1,0) = 1;$
$P(1,0) = T;$

Cyklus:

$u = 0;$
while $\sum_{\forall v} A(v,u) > 0$
  for $v = 1, \ldots, V(u) - 1$
    if $A(v,u) == 1$
      $G_{max} = -\infty;$
      for $t = P(v-1,u) + d, \ldots, P(v,u) - d$
        if $G(t|P(v-1,u), P(v,u)) > G_{max}$
          $G_{max} = G(t|P(v-1,u), P(v,u));$
          $\hat{t} = t;$
        end
      end
    end
    if $G_{max} > K$
      $P(v,u+1) = \hat{t};$
      $A(v,u+1) = 1;$
      $P(v+1,u+1) = P(v,u);$
      $A(v+1,u+1) = A(v,u);$
    else
      $P(v,u+1) = P(v,u);$
      $A(v,u+1) = 0;$
    end
  end
  $u = u + 1;$
end
$U = u;$

Ukončení:

for $v = 1, \ldots, V(U) - 1$
  $\hat{t}(v) = P(v,U);$
end

Obrázek 5.2: *Algoritmus metody binárního dělení.*

5.1.2.2 SIC s fixní hranicí kritického regionu

Budeme-li uvažovat přístup k testování hypotéz pomocí SIC s pevnou hranicí kritického regionu, zisk G cesty z bodu b do bodu t za podmínky omezení předchozím bodem změny a lze získat z rovnice (3.55).

$$G_{FTSIC}(t|a,b) = \\ (b-a+1)\log|\hat{\Sigma}| - (t-a+1)\log|\hat{\Sigma}_1| - (b-t)\log|\hat{\Sigma}_T| - D\log(b-a+1), \tag{5.4}$$

kde D je definováno vzorcem (3.56). Tato metoda bude v následujícím textu označována zkratkou FTSIC (Fixed Threshold SIC).

5.1.2.3 SIC s fixní váhou penalizační funkce

Posledním ověřovaným přístupem je z teoretického hlediska nejméně opodstatněná, leč nejvíce využívaná, metoda s pevnou váhou penalizační funkce. Ze vzorce (3.62) plyne, že odpovídající zisk bude

$$G_{FPWSIC}(t|a,b) = \\ \frac{(b-a+1)\log|\hat{\Sigma}| - (t-a+1)\log|\hat{\Sigma}_1| - (b-t)\log|\hat{\Sigma}_T|}{D\log(b-a+1)}, \tag{5.5}$$

kde D je definováno vzorcem (3.56). V dalším textu ponese označení FPWSIC (Fixed Penalty Weight SIC).

5.1.3 Efektivní implementace

Při implementaci jakéhokoliv algoritmu je vždy snahou minimalizovat jeho výpočetní náročnost. Z tohoto hlediska je nejkritičtějším místem výpočet zisku $G(t|a,b)$, konkrétně výpočet determinantů $|\hat{\Sigma}_i|$. Všechny tři funkce $G(t|a,b)$ maximalizujeme přes proměnnou t, zatímco a a b jsou fixní. Při pohledu na rovnice (5.1), (5.4) a (5.5) zjistíme, že pozici nejlepšího bodu změny lze získat minimalizací vztahu

$$\hat{t} = \arg\min_{\forall t}\left[(t-a+1)\log|\hat{\Sigma}_1| + (b-t)\log|\hat{\Sigma}_T|\right] \tag{5.6}$$

a hodnotu $G(\hat{t}|a,b)$ lze dopočítat až po jeho nalezení. Tímto se ušetří 1/3 výpočtů determinantů.

5.1.3.1 Výpočet kovariančních matic

Dalšího razantního snížení počtu numerických operací lze dosáhnout při výpočtu kovariančních matic. Nadefinujeme si d-rozměrné pole z_1 a $d \times d$ rozměrné

pole z_2[1]:

$$z_1(t) = z_1(t-1) + x_t, \tag{5.7}$$
$$z_2(t) = z_2(t-1) + x_t x_t', \tag{5.8}$$

kde x_t' značí transpozici vektoru x_t. Vyjdeme-li z definice výpočtu kovarianční matice a bude-li $E(\cdot)$ značit operátor matematického očekávání, pak lze psát

$$\bar{X} = E(X) \tag{5.9}$$
$$\Sigma = E\left[(X - \bar{X})(X - \bar{X})'\right], \tag{5.10}$$

což lze upravit do podoby

$$\Sigma = E(XX') - \bar{X}\bar{X}'. \tag{5.11}$$

Ze vztahu (5.11) pak lze odvodit maximálně věrohodné odhady parametrů gaussovského modelu počítané z dat $\{x_a, \ldots, x_b\}$ ve formě jednoduchých vzorců:

$$\hat{\mu} = \frac{z_1(b) - z_1(a-1)}{b - a + 1} \tag{5.12}$$
$$\hat{\Sigma} = \frac{z_2(b) - z_2(a-1)}{b - a + 1} - \hat{\mu}\hat{\mu}'. \tag{5.13}$$

5.1.3.2 Výpočet determinantu

Protože kovarianční matice Σ je čtvercová, *symetrická*, tj. $\sigma_{ij} = \sigma_{ji}$ pro všechna $i, j = 1, \ldots, d$, a *pozitivně definitní*, tj. $v \cdot \Sigma \cdot v > 0$ pro všechny nenulové vektory v, je z hlediska výpočetní náročnosti nejvýhodnějším postupem při výpočtu determinantu tzv. *Choleskiho dekompozice*.

Choleskiho dekompozice je speciální případ tzv. *LU* rozkladu čtvercové matice na dolní $\mathbf{L}$ a horní $\mathbf{U}$ trojúhelníkovou matici, kdy díky symetrii je možné nahradit $\mathbf{U} = \mathbf{L}^T$. Potom hledáme rozklad matice Σ

$$\mathbf{L} \cdot \mathbf{L}^T = \Sigma. \tag{5.14}$$

Řešení tohoto rozkladu lze nalézt např. v knize [Press & kol., 2002]

$$L_{ii} = \left(\sigma_{ii} - \sum_{k=1}^{i-1} L_{ik}^2\right)^{\frac{1}{2}} \text{ a} \tag{5.15}$$
$$L_{ij} = \frac{1}{L_{ii}}\left(\sigma_{ij} - \sum_{k=1}^{i-1} L_{ik}L_{jk}\right) \quad j = i+1, \ldots, d, \tag{5.16}$$

a jeho výpočet čítá $d^3/6$ operací.

[1] Ve skutečnosti bude postačovat $0.5d \times (d+1)$ rozměrné, neboť pole $z_2(t)$ je symetrické podle hlavní diagonály.

Jelikož je z elementární algebry známo, že determinant součinu horní a dolní trojúhelníkové matice je roven součinu dílčích determinantů a operace transpozice na jeho hodnotě nic nemění, lze determinant kovarianční matice zapsat

$$|\Sigma| = |\mathbf{L} \cdot \mathbf{L}^T| = |\mathbf{L}| \cdot |\mathbf{L}^T| = 2|\mathbf{L}|. \tag{5.17}$$

Další základní vlastností trojúhelníkové matice je, že její determinant je roven součinu prvků na diagonále. Logaritmus determinantu kovarianční matice tedy lze tedy vypočíst ze vzorce:

$$\log |\Sigma| = 2 \sum_{i=1}^{d} \log L_{ii}, \tag{5.18}$$

kde $\mathbf{L}$ je dolní trojúhelníková matice Choleskiho rozkladu Σ.

5.1.4 Efektivní odhad kritické hranice - trénování algoritmu

Důležitou vlastností každého algoritmu je jeho snadná trénovatelnost, čímž se rozumí, možnost rychle a efektivně odhadnout jeho volné parametry. Velkou výhodou metody binárního dělení je, že má pouze jeden volný parametr - kritickou hranici K. V implementaci popsané v části 5.1.1 existuje navíc velmi jednoduchá a efektivní možnost odhadu optimální velikosti prahu K.

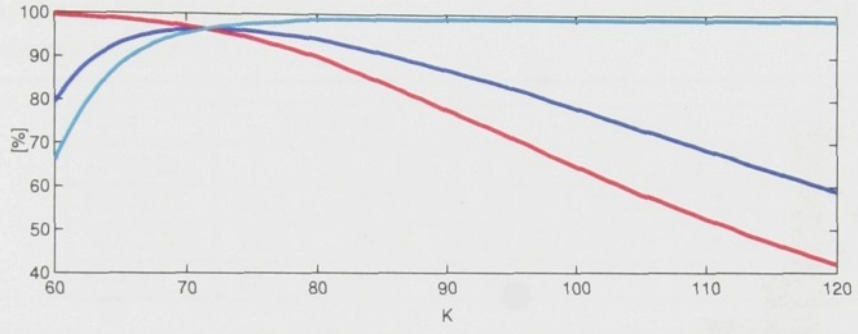
K polím vlastností $A(v,u)$ a $P(v,u)$ přidáme pole $K(v,u)$, jemuž při každém založení nového vrcholu v na úrovni u přiřadíme hodnotu odpovídající zisku G_{max}. Tuto hodnotu si daný vrchol s sebou ponese do každé následující úrovně. Pak již stačí zvolit hodnotu prahu K_{min} odpovídající maximálnímu povolenému počtu změn v trénovacích datech a kritickou mez K_{max}, jež ovlivňuje minimální počet detekovaných změn. Celý algoritmus, s výjimkou fáze **ukončení**, spustíme pouze jednou s nastavením $K = K_{min}$. Pak si stačí zvolit počet hodnot k mezi K_{min} a K_{max} a jim odpovídající místa změn získáme velmi rychle pozměněnou fází **ukončení**. Nechť $\Delta k = (K_{max} - K_{min})/k$, pak i-tou možnou segmentaci $\mathbf{t}_i^S = \{t_i^1, \ldots, t_i^S\}$ odpovídající hranici $K_{min} + i\Delta k$ získáme následovně:

$s = 1$;
for $v = 1, \ldots, V(U) - 1$
  if $K(v,U) > K_{min} + i\Delta k$
    $\hat{t}_i^s = P(v,U)$;
    $s = s + 1$;
  end
end
$S = s$;

Tento postup zopakujeme pro $i = 0, \ldots, k$, pro všechny položky trénovací databáze $p = 1, \ldots, P$ a provedeme globální vyhodnocení pro každé i přes všechny položky P.

5.2 Experimentální výsledky

V této části jsou shrnuty výsledky detekce změny řečníka na všech dostupných databázích. Základem vyhodnocení dílčích metod je idealizovaná databáze S-ART, neboť obsahuje pouze řeč od jednotlivých mluvčích, tj. neobsahuje žádné ticho ani rušení. Trochu realističtější umělá databáze ART pak umožňuje trénování dílčích metod pro účely jejich aplikace na reálná data, jež jsou reprezentována databází COST 278. Databáze FS-ART a MS-ART umožňují zodpovědět otázku, jak se liší detekovatelnost změn v závislosti na pohlaví mluvčích.



Obrázek 5.3: *Graf závisloti měr recall (červěně), precision (zeleně) a F-rate (modře) na hranici kritického regionu K pro metodu MLLR a trénovací část databáze S-ART.*

5.2.1 Databáze S-ART

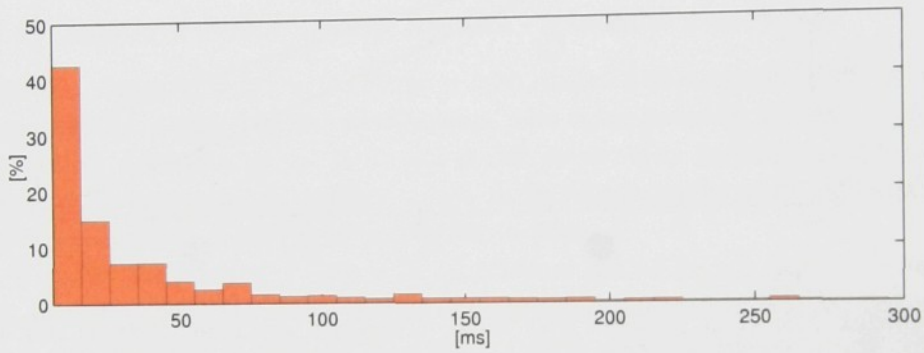
Typický graf měr *recall* (R), *precision* (P) a *F-rate* (F) v závislosti na hranici kritického regionu K je pro trénovací část databáze S-ART uveden na obrázku 5.3. Jelikož jsou tyto křivky velmi podobné pro všechny tři testované přístupy, je zobrazen průběh pouze pro metodu MLLR.

Databáze	Trénovací		Testovací		
Princip	K	F_{max} [%]	F [%]	R [%]	P [%]
MLLR	72.5	96.51	96.27	95.73	96.82
FTSIC	570	96.32	96.05	95.73	96.73
FPWSIC	2.05	94.39	94.19	92.92	95.49

Tabulka 5.1: *Vyhodnocení úspěšnosti detekce změny řečníka metod MLLR, FTSIC a FPWSIC na databázi S-ART.*

Jak již bylo uvedeno dříve, autor zvolil jako kritérium, které má být maximalizováno, míru *F-rate*. Kritická hranice K odpovídající maximu F_{max} dosaženému

na trénovací databázi byla použita k vyhodnocení metody na testovací části databáze. Výsledky všech tří přístupů jsou shrnuty v tabulce 5.1. Z této tabulky vyplývá, že za daných podmínek lze odhad parametrů u všech metod nazvat konzistentním, tj. na trénovací a testovací části dosahuje *F-rate* velmi podobných hodnot. Nejhůře ($F = 94.19\%$) dopadla metoda využívající váhování penalizační funkce Schwarz-Bayesova kritéria FPWSIC. Nicméně tento přístup je běžně využívaný, což je možná dáno tím, že absolutní rozdíl oproti zbylým přístupům $\Delta F \approx 2\ \%$ není nijak dramatický. Rozdíl mezi přístupy metodou maximální věrohodnosti a SIC je z praktického hlediska velmi malý $\Delta F \approx 0.2\ \%$. Ovšem z hlediska statistické signifikance je metoda MLLR oproti FTSIC výrazně lepší, neboť byla její menší chybovost potvrzena na úrovni $\alpha_0 = 0.1\%$.



Obrázek 5.4: *Histogram chyb časového určení pozic správně nalezených bodů změny metodou MLLR na testovací části databáze S-ART. Histogram je dělen po intervalech 10 ms.*

Dalším zajímavým pohledem na vyhodnocení navržených metod binárního dělení je přesnost časového určení správně nalezených hranic. Histogram těchto chyb pro metodu MLLR je uveden na obrázku 5.4. Srovnání parametrů histogramů ostatních metod pak nalezneme v tabulce 5.2. Není překvapující, že se tyto parametry příliš neliší, neboť pozice každého bodu je určována u všech tří metod stejnou rovnicí (5.6). Drobné rozdíly nelze přičítat rozdílu v kvalitě dílčích metod. Z tabulky 5.2 vyplývá, že detektor změn mluvčích založený na principu binárního dělení určí 2/3 resp. 95 % všech změn s přeností větší než 40 resp. 250 ms [2]. Míra δ_{10} byla zavedena v souvislosti s použitou parametrizací, kdy je vektor příznaků získáván každých 10 ms. Z tohoto pohledu lze cca 42.5 % změn považovat za určené zcela přesně.

5.2.2 Databáze FS-ART a MS-ART

Úspěšnost detekce změny mluvčího z pohledu porovnání dílčích metod binárního dělení potvrdilo výsledky testů na databázi S-ART, tj. signifikantně nejlepším byl

[2] Pro srovnání: 250 ms je doba trvání některých delších fonémů.

Princip	K	$\Delta_{2/3}$ [ms]	$\Delta_{0.95}$ [ms]	δ_{10} [%]
MLLR	72.5	40	250	42.53
FTSIC	570	40	240	42.63
FPWSIC	2.05	40	220	42.64

Tabulka 5.2: *Tabulka parametrů histogramů chyb určení časových pozic správně detekovaných bodů změny pro dílčí metody aplikované na testovací část databáze S-ART.*

opět přístup pomocí MLLR. Vyhodnocení úspěšnosti všech tří přístupů na databázích FS-ART a MS-ART lze nalézt v tabulkách B.1 a B.2. Srovnání výsledků získaných pro čistě ženskou, čistě mužskou a smíšenou databázi metodou MLLR je uvedeno v tabulce 5.3.

Část	Trénovací		Testovací		
Databáze	K	F_{max} [%]	F [%]	R [%]	P [%]
FS-ART	68.5	93.25	92.78	92.54	93.02
MS-ART	71.5	95.68	95.57	95.79	95.39
S-ART	72.5	96.51	96.27	95.73	96.82

Tabulka 5.3: *Srovnání úspěšnosti detekce změny řečníka metodou MLLR pro čistě ženskou* FS-ART, *čistě mužskou MS-ART a smíšenou databázi S-ART.*

Z této tabulky je patrný pokles úspěšnosti detekce na obou „unisex" databázích oproti databázi smíšené. Tento jev není překvapivý, neboť skutečnost, že u osob stejného pohlaví jsou rozdíly v hlasových charakteristikách menší, je zjevná. S tímto zřejmě souvisí i pokles optimální hranice kritického regionu, čímž se nastavuje vyšší „citlivost" detektoru. Zajímavým jevem je prudší pokles úspěšnosti FS-ART vs. S-ART ($\Delta F \approx 3.5\%$) oproti poklesu MS-ART vs. S-ART ($\Delta F \approx 0.7\%$). Můžeme usuzovat, že MFCC příznaky zajišťují větší separabilitu u mužských hlasů než u ženských. Na druhou stranu je toto pouze domněnka a provedené testy nelze považovat za důkaz.

5.2.3 Databáze ART

Důvodem tvorby databáze ART bylo mít možnost spolehlivě odhadnout volné parametry detektorů tak, aby je bylo možné používat v praxi. Z přehledu výsledků uvedených v tabulce 5.4 vyplývají dvě nepříjemné skutečnosti. Obě jsou zapříčiněny tím, jak se trénovací i testovací data blíží reálným signálům. Nejprve si povšimněme rostoucí optimální hranice kritického regionu odhadovaného na trénovacích datech. Tento vzestup je zapříčiněn tím, že reálnější signály obsahují různé typy rušení, které, nezačíná-li synchronně s počátkem promluvy, způsobuje chybu typu inzerce. Další inzerce jsou produkovány v místech, kde jsou např. delší úseky ticha. Vzhledem k tomu, že při trénování maximalizujeme *F-rate*, je přirozenou re-

akcí trénovacího algoritmu potlačit množství inzercí zvýšením kritické hranice K. Toto snížení citlivosti detektoru má ovšem také přímý dopad na množství delecí, neboť přestanou být detekovány některé méně významné změny mluvčích.

Databáze	Trénovací		Testovací		
Princip	K	F_{max} [%]	F [%]	R [%]	P [%]
MLLR	76.0	92.84	93.36	93.25	93.48
FTSIC	690	92.83	93.16	92.89	93.43
FPWSIC	2.20	91.59	91.83	90.68	93.01

Tabulka 5.4: *Vyhodnocení úspěšnosti detekce změny řečníka metod MLLR, FTSIC, FPWSIC na databázi ART.*

5.2.4 Databáze COST 278

Vzhledem k tomu, že databáze COST 278 není natolik rozsáhlá, aby ji bylo možné rozdělit na dostatečně velkou trénovací a testovací část, přejal autor metodiku vyhodnocení užívanou v rámci projektu COST 278 - viz [COST 278, Interspeech 2005]. Testování navržených metod probíhalo podle dvou scénářů. První scénář předpokládal trénování systému na externí databázi a následné testování na celé databázi COST. Pro natrénování systému autor využil databáze ART, základní shrnutí výsledků je znázorněno v tabulce 5.5, podrobnější přehled lze získat v tabulce B.3.

Databáze	ART		COST 278		
Princip	K	F_{max} [%]	F [%]	R [%]	P [%]
MLLR	76.0	92.84	68.27	84.56	57.24
FTSIC	690	92.83	69.65	82.88	60.07
FPWSIC	2.20	91.59	70.80	79.06	64.10

Tabulka 5.5: *Vyhodnocení úspěšnosti detekce změny řečníka metod MLLR, FTSIC, FPWSIC trénovaných na databázi ART a testovaných na DB COST 278.*

Z tabulky 5.5 je patrný znatelný pokles úspěšnost u všech tří metod o cca $\Delta F \approx 20\%$. Je způsobený především poklesem míry *precision*, což odpovídá prudkému nárůstu chyb typu inzerce. Důvody tohoto jevu jsou z části vysvětleny v sekci 5.2.3 věnované databázi ART, z části jsou způsobeny tím, že databáze je neustále ve stádiu vzniku a obsahuje značné množství chyb. Zajímavějším momentem je absolutně opačné pořadí dílčích metod vzhledem k úspěšnosti, než jaké bylo dosaženo při tzv. *matched testech*[3] uvedených v předchozích částech. Zdá se, že metody FPWSIC a FTSIC vykazují vyšší odolnost proti nesourodým trénovacím a testovacím podmínkám než metoda MLLR. Na druhé straně si je však

[3]Test, kdy trénovací a testovací data jsou obdobného charakteru.

nutné uvědomit, že počet testovaných položek $P = 57$ je vcelku nízký. Vzhledem k tomu, že test signifikance za daných podmínek není příliš přesvědčivým důkazem, neodvažuje se autor z těchto výsledků vyvozovat žádný jednoznačný závěr.

Zvýšení věrohodnosti výsledků podle druhého scénáře bylo dosaženo pomocí principu rotace trénovacích a testovacích dat. Algoritmus byl natrénován vždy pro jednu komponentu databáze a následně otestován na zbylé části. Tento postup byl opakován pro všechny komponenty, tj. 10 krát. Úspěšnost detektoru pak byla vyhodnocena jako průměr dílčích výsledků a jejich hodnoty jsou uvedeny v tabulce 5.6. Podrobnější přehled lze nalézt v příloze - viz tabulky B.4, B.5 a B.6.

Část	Trénovací		Testovací		
Princip	ϕK	ϕF_{max} [%]	ϕF [%]	ϕR [%]	ϕP [%]
MLLR	90.95	74.16	70.74	74.53	68.62
FTSIC	1057	73.51	70.52	75.13	67.72
FPWSIC	2.30	73.31	70.31	72.39	69.47

Tabulka 5.6: *Výsledky cyklického testu na databázi COST 278.*

Z těchto testů vychází nejlépe opět metoda MLLR následovaná FTSIC a FPWSIC. U metod MLLR a FTSIC lze pozorovat zlepšení (2.5%, 0.8%) oproti předchozímu testu, kdy byly trénovány na externích datech. U metody FPWSIC nastává naopak mírné zhoršení (0.5%), což naznačuje nekonzistenci odhadu její kritické hranice.

5.3 Shrnutí

Z pohledu praktického využití detektoru změn mluvčího založeného na metodě binárního dělení jsou zajímavé údaje o úspěšnosti detekce změn a výpočetních nárocích. Výpočetní náročnost všech tří porovnávaných verzí je zhruba stejná a závisí jak na délce celého signálu, tak na počtu změn. Na běžném osobním počítači (P IV, 2.4 GHz) a používaných databázích byl každý signál rozsegmentován v průměru 10 krát rychleji, než byla jeho doba trvání. Paměťové nároky jsou determinovány především poli z_1 a z_2 - viz (5.7). Jejich velikost závisí na použité parametrizaci a délce signálu. Při použité parametrizaci MFCC, kdy byl příznakový vektor o rozměru 12 emitován každých 10 ms, je to zhruba 40 kB na jednu sekundu záznamu. Z pohledu statistického je nejúspěšnější verzí metody binárního dělení princip MLLR, nejhorší pak FPWSIC. Z praktického hlediska je však mezi nimi tak malý rozdíl, že je lze považovat za stejné.

KAPITOLA 6

METODA GLOBÁLNÍ MAXIMALIZACE BIC

Všechny ostatní metody popisované v této práci jsou založeny na předpokladu, že analyzovaná část signálu obsahuje pouze jeden bod změny. Multiple change-point problém je pak řešen rozkladem na sérii single change-point úloh za pomoci vhodných metod. Je patrné, že tyto metody zajišťují pouze lokálně optimální segmentace signálu. Z tohoto důvodu se autor pokusil nadefinovat problém detekce více bodů změny jako globální optimalizační úlohu s jasným kritériem, jímž je nalezení nejlepší sekvence stavů skrytého markovovského modelu. Vzhledem ke skutečnosti, že předem není znám počet stavů ani parametry dílčích stavů, pokusil se autor tuto nepříjemnost řešit pomocí dobře známých aproximací směřujících ke globální maximalizaci Bayesovského informačního kritéria.

Z teoretických úvah nastíněných v části 3.4.4 vyplývá, že optimální sekvenci bodů změn lze získat nalezením takové segmentace $\mathbf{t}_i^S$, jež bude maximalizovat logaritmickou věrohodnost dat, potažmo BIC:

$$\log p(\mathbf{x}|\mathbf{S}_i^S) \approx \mathcal{BIC}(\mathbf{x}|\mathbf{S}_i^S) = \sum_{s=1}^{S} \ell(t_i^s, t_i^{s-1}) - \lambda \frac{Sc}{2} \log T, \tag{6.1}$$

kde $\ell(t_i^s, t_i^{s-1})$ je dáno rovnicí (3.74) a c vztahem (3.76). První člen rovnice (6.1) vyjadřuje věrohodnost segmentace $\mathbf{t}_i^S$. Jeho maximalizací pro daný počet mluvčích S získáme body změn, tj. segmentaci $\mathbf{t}_{opt}^S$. Segmentaci $\mathbf{t}_{opt}^S$ určíme pro všechny možné počty mluvčích[1] $S = 1, ..., S_{max}$. Aplikace druhého členu rovnice (6.1) pak umožní vybrat správný řád segmentace S a tím i optimální segmentaci $\mathbf{t}_{opt}$.

6.1 Návrh algoritmu

Zřejmě nejefektivnější implementací této optimalizační úlohy je přístup založený na principu *dynamického programování*- viz např. [Huang & kol., 2001]. Algoritmus rozdělíme na dvě základní části:

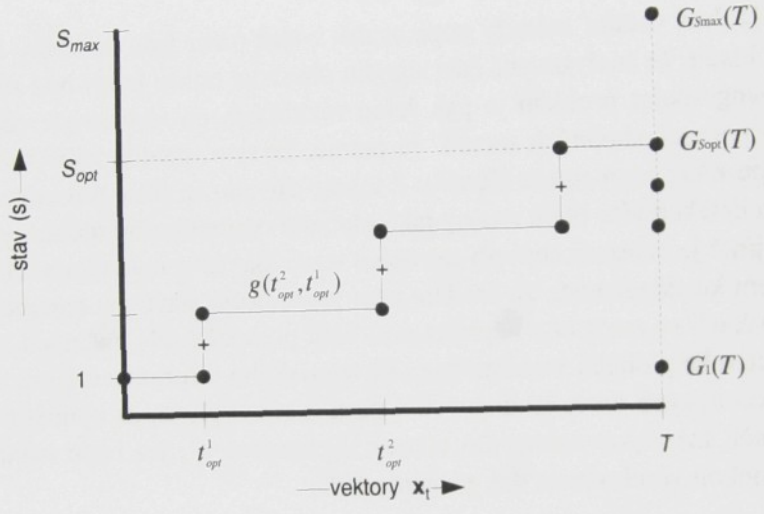
[1] Celý audio signál mohl být produkován buď jedním mluvčím nebo v mezním případě $S_{max} = T$ mluvčími, tj. každý příznakový vektor by teoreticky příslušel jiné osobě.

dopředná část zajistí maximalizaci prvního členu rovnice (6.1) přes všechny možné segmentace $\mathbf{t}_i^S$.

zpětná část vybere nejlepší řád segmentace S_{opt} a identifikuje odpovídající body změny $\mathbf{t}_{opt}$.

6.1.1 Dopředná část

Maximalizaci prvního členu rovnice (6.1) lze chápat jako hledání nejziskovějších cest z bodu $(1, 1)$ do bodů (T, s) roviny znázorněné na obrázku 6.1, kde s nabývá hodnot $1, ..., S_{max}$.



Obrázek 6.1: *Demonstrace nejziskovější cesty pro optimální segmentaci* $\mathbf{t}_{opt}$.

V souladu s principy dynamického programování zavedeme parciální zisk cesty z bodu $(t_i^{s-1}, s-1)$ do bodu (t_i^s, s) dle rovnice (3.74)

$$g(t_i^s, t_i^{s-1}) = \frac{t_i^{s-1} - t_i^s}{2} \left[\log |\Sigma| + d + d \log (2\pi)\right], \tag{6.2}$$

kde Σ je kovarianční matice dat $\mathbf{x}_s = \{x_{t_i^{s-1}+1}, \ldots, x_{t_i^s}\}$ a d rozměr příznakového vektoru. Vertikálními spojnicemi označíme operaci sčítání. Abychom byli schopni odhadnout kovarianční matici s určitou spolehlivostí, musíme určit minimální délku segmentu T_{min}, což automaticky determinuje maximální počet segmentů S_{max}. Nyní nadefinujeme dvě pole. Pole $G_s(t)$ bude obsahovat hodnotu globálního zisku cesty od počátku do současného časového okamžiku t procházejíce stavy $1, \ldots, s$. Pole pak $P_s(t)$ bude uchovávat pozici konce předchozího segmentu.

Uplatníme-li princip dynamického programování, pak pro každý časový okamžik t a každý možný počátek segmentu τ vypočteme zisk $g(t, \tau)$. Poté pro každý stav s najdeme takovou pozici τ, jež maximalizuje současný globální zisk $G_s(t)$ a

uložíme ji do $P_s(t)$. Bude-li T celková délka akustického signálu, lze dopřednou část zapsat následovně:

Cyklus:

for $t = T_{min}, \ldots, T$
  $G_1(t) = g(t, 0)$;
  $P_1(t) = 0$;
  for $s = 2, \ldots, S_{max}$
    $G_s(t) = -\infty$;
    for $\tau = (s-1)T_{min}, \ldots, t - T_{min}$
      $G_{max} = G_{s-1}(\tau) + g(t, \tau)$;
      if $G_{max} > G_s(t)$
        $G_s(t) = G_{max}$;
        $P_s(t) = \tau - 1$;
      end
    end
  end
end

6.1.2 Zpětná část

Po provedení dopředné části algoritmu získáme dílčí segmentace $\mathbf{t}_{opt}^{s}$ maximalizující zisk pro případy $s = 1, \ldots, S_{max}$ segmentů. Optimální řád segmentace vybereme dle vztahu

$$S_{opt} = \arg\max_{\forall s} \left[G_s(T) - \lambda \frac{sc}{2} \log T \right], \tag{6.3}$$

kde c je definováno vzorcem (3.76). Posloupnost bodů změny $\mathbf{t}_{opt}^{S_{opt}}$ získáme pomocí tzv. zpětného trasování:

Zpětné trasování:

$t_{opt}^{S_{opt}} = T$;
for $s = S_{opt}, \ldots, 1$
  $t_{opt}^{s-1} = P_s(t_{opt}^{s})$;
end

6.1.3 Efektivní implementace

Ke snížení počtu numerických operací nutných k výpočtu determinantu lze využít stejných principů jako u metody binárního dělení - viz část 5.1.3. Dalším efektivním způsobem snížení množství numerických operací je redukce počtu prohledávaných cest při průchodu modelem.

První možností je dynamický odhad maximálního řádu segmentace. Předpokládejme, že jsouce v čase t, můžeme odhadnout řád segmentace v čase $t-1$ z rovnice

$$S_{high}(t-1) = \arg\max_{\forall s} \left[G_s(t-1) - \lambda_{min} \frac{sc}{2} \log(t-1) \right], \tag{6.4}$$

kde λ_{min} je spodní odhad parametru λ. Jelikož řád segmentace v čase $t-1$ je roven $S_{high}(t-1)$, je patrné, že v čase t nemůže být aktuální řád segmentace více než o jednu větší. V dopředné části algoritmu tedy lze průběžně odhadovat S_{max} z rovnice:

$$S_{max}(t) = S_{high}(t-1) + 1. \tag{6.5}$$

Daleko závažnějším problémem, než je omezení maximálního počtu stavů, je lineární nárůst výpočetní náročnosti s rostoucím časem t, což je způsobem slabými restrikcemi hloubky prohledávaného prostoru, tj. hodnotami, jež může nabývat proměnná τ:

$$\tau = (s-1)T_{min}, \ldots, t - T_{min}. \tag{6.6}$$

Pro každý čas t je tedy třeba vypočíst zisk $g(t,\tau)$ takřka t-krát. Jelikož však víme, že čím vyšší zvolíme parametr λ, tím nižší bude řád segmentace (tzn. bude delší očekávaná délka segmentů), můžeme zavést omezení na maximální hloubku prohledávaného prostoru volbou horní hranice parametru λ_{max}. Zjistíme minimální řád segmentace v čase $t-1$

$$S_{low}(t-1) = \arg\max_{\forall s} \left[G_s(t-1) - \lambda_{max} \frac{sc}{2} \log(t-1) \right] \tag{6.7}$$

a budeme opět předpokládat, že v čase t nebude vyšší než $S_{low}(t-1)+1$. Za těchto podmínek můžeme zavést restrikci na možné hodnoty, jichž může nabývat τ:

$$\tau = P_{S_{low}(t-1)+1}(t-1), \ldots, t - T_{min}. \tag{6.8}$$

Tento druh prořezávání umožňuje významné snížení výpočetních nároků.

6.1.4 Efektivní odhad parametru λ - trénování algoritmu

Velkou výhodou metody globální maximalizace je snadná trénovatelnost, neboť vše podstatné se spočte již při dopředné části algoritmu. Stačí zvolit meze odhadu jediného volného parametru $\lambda \in \langle \lambda_{min}, \lambda_{max} \rangle$, jež slouží zároveň pro snížení výpočetní náročnosti, a spustíme dopřednou proceduru. Poté si zvolíme počet hodnot

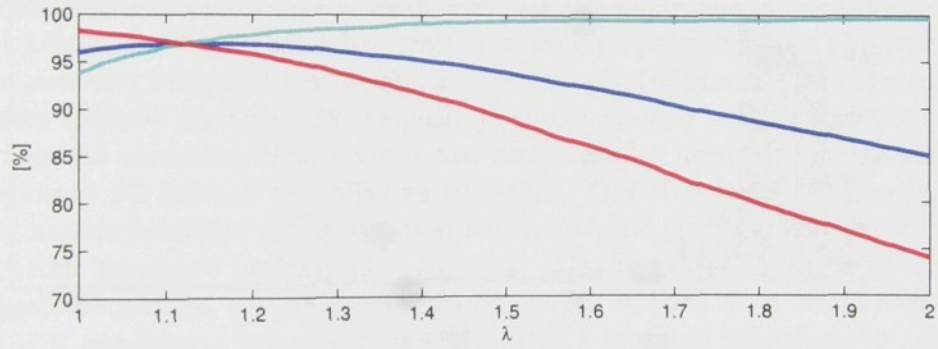
k mezi λ_{min} a λ_{max} a nadefinujeme krok $\Delta k = (\lambda_{max} - \lambda_{min})/k$. Zjistíme i-tý optimální řád segmentace

$$S_{opt}(i) = \arg\max_{\forall s} \left[G_s(T) - (\lambda_{min} + i\Delta k)\frac{sc}{2}\log T \right] \tag{6.9}$$

a spustíme fázi zpětného trasování, jež neklade takřka žádné výpočetní nároky. Tento postup zopakujeme pro $i = 0, \ldots, k$ a pro všechny položky trénovací databáze $p = 1, \ldots, P$. Poté provedeme globální vyhodnocení přes všechny položky P pro každou hodnotu i. Za optimální nastavení algoritmu budeme považovat to i, jež poskytne nejvyšší hodnotu míry *F-rate*.

6.2 Experimentální výsledky

Z obrázku 6.2 je patrné, že průběh měr *recall(R)* , *precision(P)* a *F-rate(F)* je velmi podobný odpovídajícímu grafu získaného z trénovacích dat databáze S-ART metodou binárního dělení. Pro ostatní databáze jsou průběhy křivek trénování metody natolik podobné, že je zbytečné je zde dále uvádět.



Obrázek 6.2: *Graf závisloti měr recall (červěně), precison (zeleně) a F-rate (modře) na hranici kritického regionu K pro metodu globální maximalizace BIC a trénovací část databáze S-ART.*

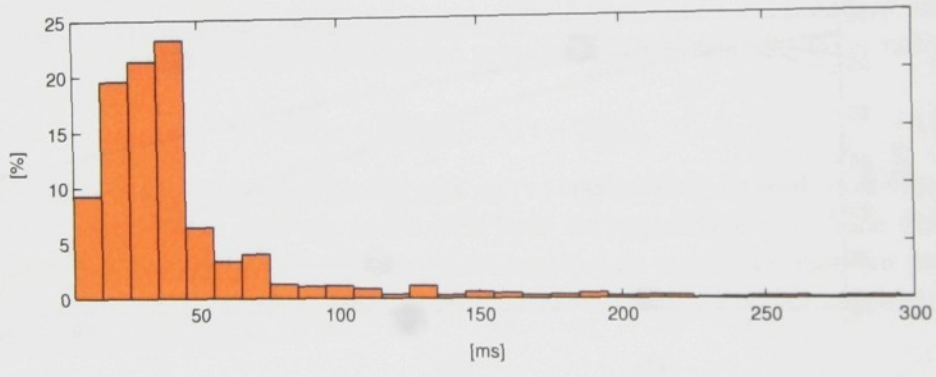
Z hlediska vyhodnocení metody globální maximalizace BIC (GMBIC) jsou zajímavé údaje shrnuté v tabulce 6.1, kde jsou uvedeny výsledky testů na dílčích databázích. Pro porovnání zde také nalezneme výsledky nejlepší verze metody binárního dělení - metody MLLR. Z této tabulky je patrné, že metodou GMBIC lze dosahovat lepších výsledků než metodou MLLR, což je potvrzeno konzistentním zvýšením míry *F-rate* na všech využívaných databázích.

Test na databázi S-ART přinesl pouze nepatrné zvýšení úspěšnosti detekce bodu změny, $\Delta F \approx 0.5\%$. Ze statistického testu ale vyplývá, že se jedná o signifikantní zlepšení - dokonce na úrovni signifikance $\alpha_0 = 0.1$ %. Poněkud rozporuplných výsledků však bylo dosaženo při porovnání přesnosti detekce bodů změny z hlediska jejich časových poloh. Z tabulky 6.2 je možno vypozorovat, že dochází

Část	Trénovací		Testovací			MLLR
Databáze	λ	F_{max} [%]	F [%]	R [%]	P [%]	F [%]
S-ART	1.10	96.88	96.76	97.26	96.27	96.27
MS-ART	1.15	95.99	95.99	96.12	95.86	95.27
FS-ART	1.03	93.79	93.50	93.83	93.18	92.78
ART	1.31	93.75	94.20	94.32	94.09	93.36
ART-COST	1.31	93.75	70.03	87.43	58.41	68.27
COST-COST	1.80	77.13	72.99	74.63	73.03	70.74

Tabulka 6.1: *Vyhodnocení úspěšnosti detekce změny řečníka metodou GMBIC. Řádek ART-COST označuje test na databázi COST 278 dle scénáře s trénováním na externích datech, COST-COST značí cyklický test.*

k výraznému zlepšení míry $\Delta_{2/3}$ o 70 ms. Na druhou stranu však razantně klesá absolutní přesnost detektoru δ_{10}. Zatímco metodou MLLR se podařilo absolutně přesně detekovat 42.63 % změn, metodou GMBIC pak pouhých 9.29 %. Tento jev je ilustrován na obrázku 6.3.



Obrázek 6.3: *Histogram chyb časového určení pozic správně nalezených bodů změny metodou GMBIC na testovací části databáze S-ART. Histogram je dělen po intervalech 10 ms.*

Vyhodnocení metody GMBIC na databázích MS-ART a FS-ART potvrzuje výsledky dosažené metodou binárního dělení (popsané v části 5.2.2), tj. mírně zhoršenou detekovatelnost změn mluvčích stejného pohlaví. Nicméně i na těchto datech lze pozorovat mírné zlepšení oproti MLLR - $\Delta F \approx 0.7\%$ - potvrzené testem statistické signifikance.

Pro vyhodnocení úspěšnosti detekce změn na reálných nahrávkách databáze COST 278 byly v části 5.2.4 nadefinovány dva druhy testů. Test, kdy byla pro trénování užita externí data (databáze ART), přinesl zlepšení detektoru o $\Delta F \approx 1.8\%$ - viz řádek ART-COST tabulky 6.1. Z řádku COST-COST téže tabulky je patrné, že u testu založeného na rotaci trénovacích a testovacích dat došlo ještě k výraznějšímu zlepšení oproti metodě MLLR - $\Delta F \approx 2.2\%$. Podrobnější údaje

Metoda	λ/K	$\Delta_{2/3}$ [ms]	$\Delta_{0.95}$ [ms]	δ_{10} [%]
GMBIC	1.10	40	180	9.29
MLLR	72.5	40	250	42.63

Tabulka 6.2: *Tabulka parametrů histogramů chyb určení časových pozic správně detekovaných bodů změny pro dílčí metody aplikované na testovací část databáze S-ART.*

o testech metody GMBIC na databázi COST 278 jsou uvedeny v tabulkách B.7 a B.8.

6.3 Shrnutí

Vyhodnocením metody globální maximalizace BIC bylo potvrzeno, že přístup, kdy je multiple-change point problém řešen jako globální optimalizační úloha, přináší lepší výsledky (viz srovnání s výsledky partnerů spolupracujících v rámci projektu COST 278 [COST 278, Interspeech 2005]), než přístup přes dílčí lokální optimalizace. Tohoto zlepšení však bylo dosaženo na úkor razantního nárůstu jak výpočetních, tak paměťových nároků. I při efektivní implementaci metodou dynamického programování spotřebuje tento algoritmus při parametrizaci s 12 příznaky emitovanými každých 10 ms v průměru 0.5 MB operační paměti na 1 s záznamu. Růstu spotřeby paměti s délkou signálu je možné zabránit implementací formou kruhového zásobníku, čímž je však nutné zavést restrikci maximální možné délky segmentu. Při omezení této délky na 10 minut je GMBIC zhruba 10x pomalejší než metoda binárního dělení. Rychlost lze samozřejmě také výrazně ovlivnit nastavením parametrů prořezávání λ_{max}, λ_{min}. Ačkoliv je GMBIC metoda velmi výpočetně náročná a její výsledky nejsou zas tak výrazně lepší než u metody MLLR, perspektiva této metody tkví především v možnosti jednoduchého zabudování apriorní pravděpodobnosti segmentace formou distribuční funkce popř. histogramu délek segmentů, což je u metod využívajících přístupu pomocí testování hypotéz podstatně komplikovanější.

KAPITOLA 7

METODA S ADAPTIVNÍM OKNEM

Tato kapitola se zabývá velmi účinnou a přitom výpočetně nenáročnou metodou detekce více bodů změny, jíž autor nazývá metoda s adaptivním oknem, neboť je založena na detekci bodu změny v analyzujícím okně, jehož počátek i konec se neustále adaptuje tak, aby bylo možno rozhodovat o bodu změny s maximální možnou jistotou. Vychází přitom z dobře známého algoritmu publikovaného poprvé v [Chen & kol., 1998]. Tato metoda má však tři volné parametry a bohužel autoři ani jejich následovníci nepopisují možnosti jejich efektivního odhadu. Proto autor této práce navrhl takovou její modifikaci, že dva ze tří parametrů pak již nemají na úspěšnost metody takřka žádný vliv a třetí parametr lze odhadnout pomocí metody binárního dělení.

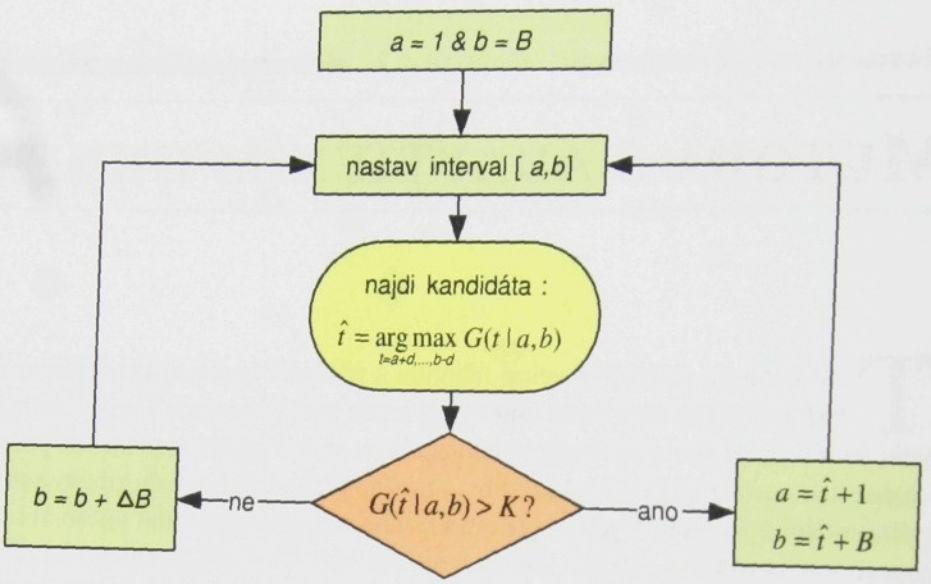
7.1 Originální algoritmus

Originální algoritmus metody s adaptivním oknem (AWIN) schematicky znázorňuje obrázek 7.1. Proměnnými $[a, b]$ označíme počátek a konec aktuálně analyzované části signálu. Dále zavedeme inicializační délku okna B, koeficient rozšíření ΔB a proměnnou $G(t|a, b)$ označíme zisk asociovaný s bodem změny t. Způsob jeho výpočtu je stejný jako u metody binárního dělení a definici lze nalézt v části 5.1.2.

Celý algoritmus funguje na velmi jednoduchém principu. Inicializujeme okno o velikosti B a umístíme ho na počátek signálu. V daném okně najdeme takový bod $\hat{t}$, jenž bude maximalizovat zisk $G(\hat{t}|a, b)$ a označíme ho jako *kandidáta* na bod změny. Je-li kandidát potvrzen, tj. $G(\hat{t}|a, b) > K$, lze ho považovat za bod změny, počátek analyzujícího okna je přesunut do polohy $\hat{t} + 1$ a velikost okna je nastavena na hodnotu B. Není-li kandidát $\hat{t}$ potvrzen, je okno rozšířeno o délku ΔB a celý postup se opakuje.

7.2 Modifikovaný algoritmus

Modifikovaný algoritmus (MAWIN) funguje na velmi obdobném principu jako jeho základní verze. V aktuálně analyzované části signálu však není hledána pouze jediná změna, ale ta změna, která se nachází nejblíže počátku, čímž lze s úspěchem takřka eliminovat vliv parametrů B a ΔB na úspěšnost detekce. V reálné imple-



Obrázek 7.1: *Schéma originální metody s adaptivním oknem.*

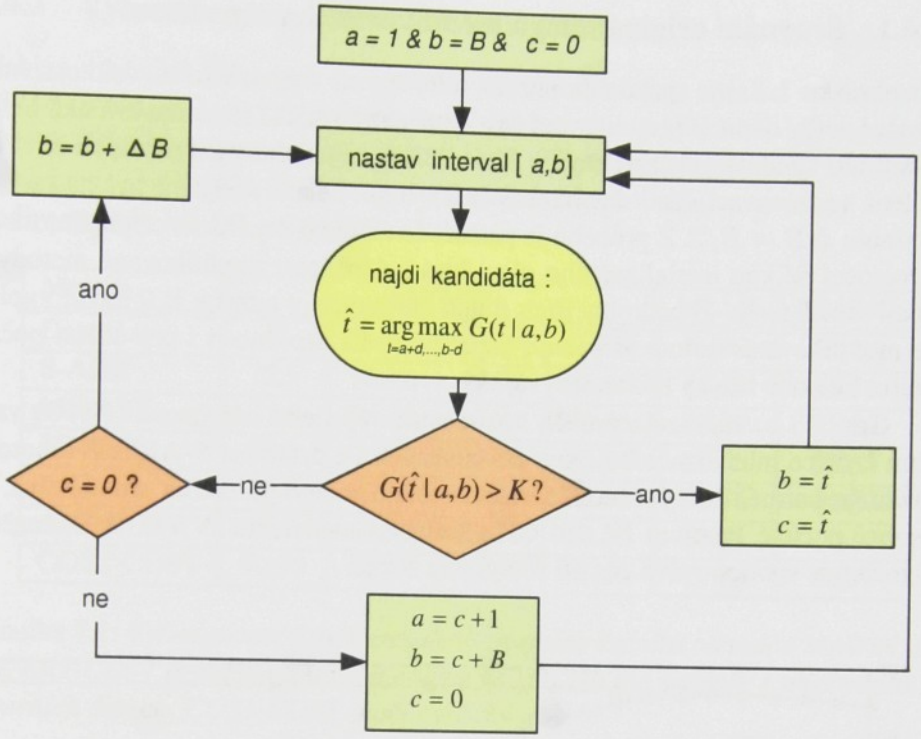
mentaci pak tato změna znamená přidání jedné podmínky a jedné proměnné c - viz obrázek 7.2.

Algoritmus inicializujeme stejně jako v předchozím případě a najdeme kandidáta na bod změny. Poté mohou nastat dva různé případy. Kandidát není potvrzen a rozšíříme analyzující okno. Je-li kandidát $\hat{t}$ potvrzen, zmenšíme okno na velikost $[a, \hat{t}]$ a najdeme v něm dalšího kandidáta. Bude-li tento kandidát potvrzen, opět okno zmenšíme. V opačném případě jsme nalezli bod změny, jenž byl z hlediska původního okna nejvíce vlevo, posuneme analyzující okno do tohoto bodu a nastavíme jeho inicializační velikost B.

7.3 Efektivní implementace a otázka trénování

Jelikož výpočetní náročnost celého algoritmu je opět dána především počtem numerických operací nutných k výpočtu determinantů kovariančních matic, lze i u této metody využít ke snížení počtu nutných operací stejných principů, popsaných v části 5.1.3, jako u metody binárního dělení.

Otázku odhadu kritické hranice K pro potvrzení či zamítnutí kandidáta bodu změny lze vyřešit velmi snadno, neboť popisovaná metoda využívá k detekci bodu změny stejného principu jako metoda binárního dělení. Pro optimální nastavení parametru K tedy s výhodou poslouží trénovací algoritmus metody binárního dělení (viz 5.1.4), čímž se stává nedílnou součástí metody s adaptivním oknem.



Obrázek 7.2: *Schéma modifikované metody s adpativním oknem.*

7.4 Experimentální výsledky

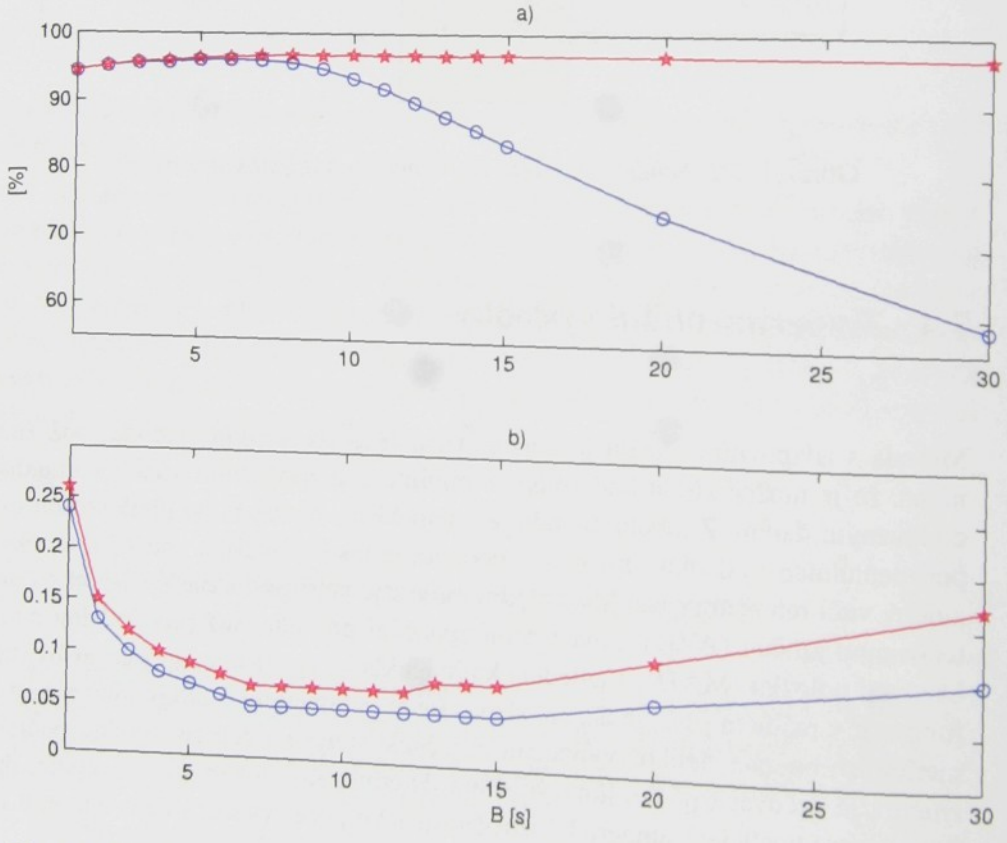
Metoda s adaptivním oknem je koncipována jako tzv. *on-line metoda*, což znamená, že je možné získat bod změny s minimálním zpožděním vůči kontinuálně dodávaným datům. Z tohoto důvodu je v této části - na rozdíl od předchozích experimentálních výsledků - metoda vyhodnocena také z hlediska zpoždění detekce změny vůči reálnému času. Mezi sledované údaje patří průměrné zpoždění na detekovanou změnu (AVD) a maximální zpoždění průměrované přes každou databázovou položku (MXD). Vzhledem k předpokladu, že on-line algoritmus by měl fungovat v reálném čase, je dalším sledovaným parametrem metody množství numerických operací. Jelikož výpočetní náročnost souvisí s délkou signálu, počtem změn a je určována především výpočtem determinantů kovariančních matic, definuje autor koeficient numerické náročnosti jako počet výpočtů determinantů na změnu a délku signálu (NNO).

Vzhledem k úzké spjatosti s metodou binárního dělení, pro niž bylo dosaženo nejlepších výsledků užitím verze MLLR, byla testována metoda s adaptivním oknem pouze v téže verzi, tj. zisk $G(t|a,b)$ byl počítán dle vztahu (5.1).

7.4.1 Srovnání originálního a modifikovaného algoritmu

Na obrázku 7.3 a) je znázorněn průběh míry *F-rate* v závislosti na velikosti inicializační délky okna B pro testovací část databáze S-ART. Hodnota kritické hranice zamítnutí kandidáta na bod změny $K = 72.5$ byla natrénována metodou binárního dělení na trénovací části databáze S-ART a koeficient rozšíření byl nastaven na hodnotu $\Delta B = B/2$. Z průběhu je patrné, že zatímco úspěšnost originální metody s rostoucí délkou inicializačního okna klesá, úspěšnost modifikované metody naopak mírně roste. Z podrobnějších údajů shrnutých v tabulce B.9 navíc vyplývá, že modifikovaná metoda dosahuje nepatrně vyšší úspěšnosti i pro oblast počátku grafu, kde obě křivky takřka splývají.

Graf 7.3 b) zobrazuje průběh koeficientu výpočetní náročnosti (NNO) vzhledem k délce inicializačního okna B - získaného z databáze S-ART. Z tohoto obrázku je patrné, že výpočetní náročnost obou metod je takřka srovnatelná. Pro osobní počítač Pentium IV, 2.4 GHz hodnota koeficientu $NNO = 1$ značí, že algoritmus spotřebovává cca 50 % výkonu stroje.



Obrázek 7.3: *Srovnání originální (modře) a modfikované (červeně) metody s adaptivním oknem. Graf a) zobrazuje průběh míry F-rate pro různé volby velikosti okna B. Graf b) ilustruje závislost NNO na B.*

7.4.2 Vyhodnocení modifikovaného algortimu

Přehled výsledků získaných na jednotlivých databázích je uveden v tabulce 7.1. Při těchto testech byla kritická hranice K odhadována pomocí MLLR metody binárního dělení, velikost inicializačního okna B byla nastavena na 30 s a hodnota koeficientu rozšíření $\Delta B = 15$ s. Pro srovnání jsou v tabulce uvedeny také výsledky metody MLLR a GMBIC.

Metoda	MAWIN				MLLR	GMBIC
Databáze	K	R [%]	P [%]	F [%]	F [%]	F [%]
S-ART	72.5	95.76	97.96	96.85	96.27	96.76
MS-ART	71.5	95.86	97.04	96.46	95.27	95.99
FS-ART	68.5	92.64	94.80	93.71	92.78	93.50
ART	76.0	93.14	94.88	93.90	93.36	94.20
ART-COST	76.0	84.33	59.11	69.50	68.27	70.30
COST-COST	90.95	74.44	70.33	71.65	70.74	72.99

Tabulka 7.1: *Vyhodnocení úspěšnosti detekce změny řečníka metodou MAWIN. Řádek ART-COST označuje test na databázi COST 278 dle scénáře s trénováním na externích datech, COST-COST značí cyklický test.*

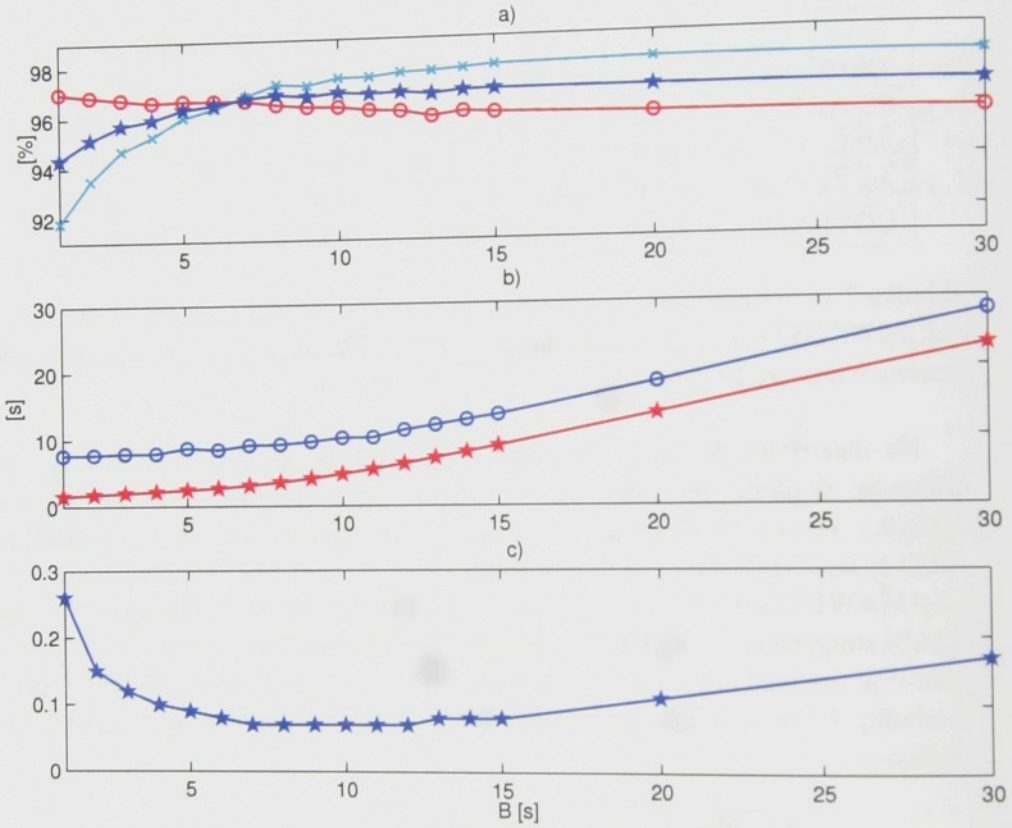
Na databázích S-ART, MS-ART a FS-ART lze pozorovat statisticky signifikantní zlepšení oproti metodě MLLR. Při srovnání s metodou GMBIC jsou výsledky metody MAWIN z hlediska statistické signifikance stejné (FS-ART, S-ART) či lepší (MS-ART) na úrovni signifikance $\alpha_0 = 0.1$ %. Vyhodnocením metody MAWIN na databázích ART a COST 278 pak zjistíme, že tyto výsledky jsou signifikantně lepší oproti MLLR a naopak signifikantně horší oproti GMBIC. Zajímavé je srovnání přesnosti detekce bodů změny z pohledu jejich časového určení. Z tabulky 7.2 je patrné, že metoda MAWIN v tomto směru poskytuje nejlepší výsledky.

Metoda	λ/K	$\Delta_{2/3}$ [ms]	$\Delta_{0.95}$ [ms]	δ_{10} [%]
MAWIN	72.5	30	180	45.96
MLLR	72.5	40	250	42.63
GMBIC	1.10	40	180	9.29

Tabulka 7.2: *Tabulka parametrů histogramů chyb určení časových pozic správně detekovaných bodů změny pro dílčí metody aplikované na testovací část databáze S-ART.*

Z obrázku 7.4 a), kde jsou zobrazeny průběhy měr *F-rate*, *recall* a *precision* v závislosti na délce inicializačního okna B pro databázi S-ART, plyne, že míra *F-rate* se dostává do saturace pro $B \approx 14$ s. Tato hodnota pravděpodobně souvisí s průměrnou délkou segmentu databáze S-ART, jež je 6.31 s - viz tabulka 4.2. Chceme-li tedy detekovat změny on-line a zároveň s co nejvyšší úspěšnosti, měla

by být délka inicializačního okna minimálně dvojnásobkem očekávané délky segmentu. Tento požadavek je pochopitelný, neboť jedině tímto způsobem zajistíme, že rozhodnutí o změně nebude „v průměru" učiněno dříve, než je k dispozici maximální možné množství relevantních dat. Zároveň z výše uvedeného logicky plyne, že průměrné zpoždění detekce změny AVD pro toto nastavení by mělo nabývat hodnoty zhruba poloviny incializační délky B, což je potvrzeno grafem 7.4 b). Přesná hodnota je uvedena v tabulce B.9. Okolo tohoto bodu se také pohybuje nastavení, jež klade nejmenší výpočetní nároky - viz graf 7.4 c).



Obrázek 7.4: *Vyhodnocení on-line vlastností modifikované metody s adaptivním oknem v závislosti na délce inicializačního okna B:* **a)** *průběh měr recall (červeně), precision (zeleně) a F-rate (modře);* **b)** *průběh maximálního MXD (modře) a průměrného AVD (červeně) zpoždění;* **c)** *průběh výpočetní náročnosti NNO.*

7.5 Shrnutí

Výsledkem těchto experimentů popisovaných v této kapitole je tedy zjištění, že metodou MAWIN lze získat nepatrně lepší výsledky než MLLR a stejné či zanedbatelně horší výsledky ve srovnání s metodou GMBIC. Zároveň metoda MAWIN poskytuje nejpřesnější odhad pozic bodů změny. Výpočetní nároky algoritmu

jsou nízké, neboť průměrné zatížení PC Pentium IV s 2.4 GHz procesorem se při on-line implementaci pohybovalo kolem 5 % výkonu stroje. Paměťové nároky algoritmu jsou nízké a souvisí s aktuálně analyzovanou délkou signálu. Při parametrizaci, kdy je každý vektor příznaků rozměru 12 získáván každých 10 ms, odpovídá množství spotřebované paměti 40 kB na sekundu analyzujícího okna. Vzhledem k těmto faktům se zdá být metoda MAWIN nepraktičtějším z přístupů ověřovaných v této práci.

KAPITOLA 8

ZÁVĚR

Zřejmě nejpopulárnějším přístupem k řešení úlohy detekce změny řečníka je tzv. *metoda fixních oken*. Při její implementaci se však autor setkal s celou řadou problémů, jejichž řešením se odborná literatura příliš nezabývá. Největším z nich je především otázka odhadu parametrů této metody, neboť jich je příliš mnoho a pravděpodobně neexistuje způsob jejich rozumného odhadu. Velké množství parametrů je zapříčiněno zejména nutností implementace detektoru lokálních maxim, což také není právě triviální úloha. Méně užívaným přístupem je tzv. *metoda s adaptivním oknem*, která takový detektor ke své činnosti nepotřebuje. Ani u této metody se však z literatury nelze dozvědět nic o možnostech trénování. Další problematickou částí je její faktická implementace (stejně jako u metody fixních oken), kdy je sice prezentována jako metoda založená na testování hypotéz pomocí Schwarzova (Bayesova) informačního kritéria (FPWSIC), ale ve výsledku není používána zcela v souladu s teoretickými předpoklady. Výsledky publikované v této práci ilustrují, že metodami, jež aplikují teorii testování hypotéz ve správném znění (MLLR, FTSIC), lze dosáhnout lepších výsledků.

V této práci se autor zabýval třemi alternativními způsoby detekce změn řečníka. U všech navrhl jejich efektivní implementaci, algoritmus snadného odhadu volných parametrů a otestoval je na několika typech databází. První z navržených přístupů, *metoda binárního dělení*, není sice novou metodou, ale v oblasti detekce změny řečníka pravděpodobně ještě nebyla nikdy použita. Autor ji implementoval ve třech různých verzích. Verze MLLR je založená na přístupu pomocí testování jednoduchých hypotéz, verze FTSIC na přístupu přes testování kompozitních hypotéz, jež vede na řešení využívající Schwarzova (Bayesova) informačního kritéria. Poslední verze, FPWSIC, byla testována z důvodu její oblíbenosti mezi odbornou veřejností. Jelikož je metoda binárního dělení v jiných oblastech používána již od roku 1981, není překvapující, že poskytuje velmi rozumné výsledky - a to pro všechny její testované verze. Nicméně z provedené analýzy vyšla nejlépe metoda MLLR, když na zidealizované databázi S-ART bylo detekováno takřka 96 % všech existujících změn (*recall*) a téměř 97 % detekovaných změn bylo nalezeno správně (*precision*). Na reálné databázi COST 278 dopadla detekce změn o poznání hůře. Bezchybně bylo detekováno necelých 75 % všech existujících změn a pouze 69 % z nalezených změn bylo určeno správně.

Jelikož metoda binárního dělení je pouze šikovnou aplikací single-change point analýzy na multiple change-point problém, pokusil se autor navrhnout metodu na-

zvanou *globální maximalizace BIC*, která je definována jako globální jednoprůchodová metoda pro přímé řešení multiple change-point problému. GMBIC sice poskytuje nepatrně lepší výsledky než MLLR, cenou za zlepšení je však obrovský nárůst výpočetní a paměťové náročnosti. Její perspektiva tkví především v možnosti velmi jednoduše zahrnout do procesu detekce apriorní pravděpodobnost délek segmentů, což metody založené na testování hypotéz příliš neumožňují. Snížená funkčnost detektorů bez explicitně užívané apriorní pravděpodobnosti se projevuje především v okamžiku, jestliže je distribuční funkce apriorní pravděpodobnosti délek segmentů vícemodální, tj. jedná-li se např. o segmentaci nahrávky typu rozhovor, kdy se střídá krátká otázka moderátora s dlouhou odpovědí hosta.

Z praktického pohledu se nejvýhodnějším přístupem jeví *metoda s adaptivním oknem*, respektive její *modifikovaná* verze MAWIN. Metoda s adaptivním oknem je dobře známa již několik let, leč není příliš populární, zřejmě díky neobjasněnému způsobu odhadu jejích tří volných parametrů. S touto nepříjemností se autor vypořádal takovou její modifikací, že dva ze tří parametrů přestávají mít na úspěšnost detekce vliv. Zároveň je v práci experimentálně prokázáno, že třetí volný parametr lze efektivně odhadnout pomocí metody binárního dělení. Velkou výhodou této metody je jak její on-line pojetí, tj. minimální zpoždění detekce bodu změny vůči reálnému času, tak její nízká výpočetní a paměťová náročnost. Úspěšnost metody s adaptivním oknem se pak pohybuje mezi dvěma výše uvedenými přístupy. Zároveň vyniká nejvyšší přesností detekce bodů změny z hlediska jejich časových pozic.

Cílem této práce bylo prozkoumat metody vhodné k detekci změny řečníka z pohledu algoritmického, nikoliv z pohledu vhodných příznaků. Nalezení vhodných příznaků lze považovat za zcela samostatnou problematiku, která je řešena především v úloze identifikace a verifikace řečníka. Zajímavým výsledkem uskutečněných experimentů z hlediska používaných MFCC příznaků je horší detekovatelnost změn žena-žena oproti změnám muž-muž.

Neznalého čtenáře by mohl zarazit prudký pokles úspěšnosti detekce změn při testech na reálné databázi oproti výsledkům z uměle míchaných idealizovaných databází, kdy došlo ke snížení míry *F-rate* o cca 25 %. Velký pokles je způsoben především nárůstem chyb typu inzerce, tedy falešnou detekcí změny řečníka. Tyto chyby jsou způsobeny zejména detekcí bodů, kdy přes sebe mluví několik mluvčích, do řeči hraje hudba či je na pozadí jiný aditivní, často nestacionární, šum. Je zřejmé, že detektory založené na testování změn parametrů stochastického procesu takovou změnu považují za významnější, než je změna mluvčího. Trénováním algoritmu na těchto datech, tj. maximalizací míry *F-rate*, pouze zvyšujeme hodnotu volného parametru, díky čemuž přestane detektor zachycovat některé méně významné změny mluvčích a tím roste zase počet chyb typu delece. Dalším důvodem sníženné úspěšnosti detekce změn mluvčího je také skutečnost, že testování proběhlo bez předřadného detektoru řeč/neřeč, tj. algoritmy byly trénovány jako detektory změn mluvčího, testovány jako obecné detektory akustických změn a následně vyhodnoceny opět jako detektory změn mluvčího. V neposlední řadě svou roli sehrála i kvalita databáze COST 278, která byla zvláště v některých jejích ná-

rodních částech poněkud sporná. S ohledem na praktické nasazení metod detekce změn si je rovněž zapotřebí uvědomit, že cílem této úlohy je rozdělení dlouhých záznamů řeči na kratší úseky, které mají z akustického hlediska homogenní charakter. Tato segmentace je nutná jak pro následnou identifikaci, tak i pro řádnou funkci rozpoznávače. Z tohoto hlediska nejsou navíc vložené body změny, tj. inzerce, takovým problémem, neboť nezpůsobují chybu při určování mluvčího a většinou ani nepoškodí činnost dekodéru spojité řeči. Opomenuté body změny, tj. delece, mají výrazně horší dopad na další zpracování.

Cesta ke zvýšení úspěšnosti detektoru změn mluvčího na reálných datech pravděpodobně povede přes nalezení robustnějších příznaků a tvorbu komplexnějších rozhodovacích systémů, kde bude součástí detektoru i např. detektor řeč/neřeč či modul identifikace řečníka.

Shrnutí přínosů k rozvoji vědního oboru

V práci je

- podán *jednotný výklad základních přístupů* k úloze detekce změny mluvčího v audio signálu na základě testování hypotéz o změně parametrů gaussovského procesu;
- provedeno *srovnání* běžně používaného přístupu *s alternativními přístupy*, jež plynou z teoretického rozboru;
- ověřena vyšší *spolehlivost autorem navržených alternativních přístupů*;
- potvrzena možnost aplikace *metody binárního dělení* na úlohu detekce vícenásobné změny mluvčího;
- navržena nová *metoda globální maximalizace BIC*;
- navrženo významné vylepšení on-line *metody s adaptivním oknem*;
- u všech metod popsán *způsob trénování* a způsob jejich *efektivní implementace*;
- provedeno *vyhodnocení úspěšnosti* na rozsáhlých databázích jak uměle připravených, tak i reálných vícejazyčných řečových záznamů, přičemž všechny navržené metody jsou buď lepší nebo přinejmenším stejně účinné jako metody, jež užívají partneři spolupracující v rámci projektu COST 278.

Shrnutí přínosů pro praxi

Všechny v práci navržené metody byly postupně testovány v systému pro automatický přepis televizního zpravodajství, vyvíjeného na TUL. Detektor založený na modifikované metodě s adaptivním oknem v současné době tvoří nedílnou součást

tohoto transkripčního systému a umožňuje jeho plnou automatizaci. Jeho implementací a zařazením do procesu automatického přepisu odpadla jedna z nejnamáhavějších činností, jíž je ruční segmentace záznamů. Tím, že je segmentace uskutečňována na základě změn mluvčích, mohly být do systému zařazeny také moduly pro automatickou identifikaci řečníka a adaptaci na charakteristiky jeho hlasu.

Literatura

Citovaná literatura

[Ajmera & kol., 2004] Ajmera, J., McCowan, I., Bourlard, H., *Robust Speaker Change Detection*, IEEE Signal Processing Letters, Vol. 11, No. 8, August 2004.

[Chen J., 2000] Chen, J., Gupta, A. K., *Parametric Statistical Change Point Analysis*, Birkäuser, Boston, 2000.

[Chen & kol., 2001] Chen, S. S., Eide E., Gales F. J. M., Gopinath A. R., Kanvesky D., Olsen P., *Automatic Transcription of Broadcast News*, IBM T. J. Watson Research Center, Yorktown Heights, NY 10598, 2001.

[Chen & kol., 1998] Chen, S. S., Gopalakrishnan, P. S., *Speaker, Environment and Channel Change Detection and Clustering via the Bayesian Information Criterion*, IBM T.J. Watson Research Center, Yorktown Heights, NY, Technical Report, 1998.

[Chickering & Heckerman, 1996] Chickering, D. M., Heckerman, D., *Efficient approximations for the marginal likelihood of Bayesian networks with hidden variables*, Technical report MSR-TR-96-08, Microsoft Research, 1996.

[COST 278, Interspeech 2005] Žibert, J., Mihelič, F., Martens, J. P., Meinedo, H., Neto, J., Docio, L., Garcia-Mateo, C., David, P., Zdansky, J., Pleva, M., Cizmar, A., Žgank, A., Kačič, Z., Teleki, C., Vicsi, K., *The COST 278 Broadcast News Segmentation and Speaker Clustering Evaluation - Overview, Methodology, Systems, Results*, In Proceedings of 9-th International Conference on Speech Communications and Technology Interspeech 2005, pp. 629-632, Lisboa (Portugal), 2005.

[Delacourt & Wellekens, 2000] Delacourt P., Wellekens, C. J., *DISTBIC: A Speaker-based Segmentation for Audio Data Indexing*, In Speech Communication, Vol. 32, No 1-2,2000.

[ETSI, 2003] ETSI ES 202 211 V1.1.1 (2003-11), on-line dokument na serveru `www.etsi.org`.

[Ferguson, 1980] Ferguson, J. D., *Variable duration models for speech*, In Proc. Symposium on the Application of Hidden Markov Models to Text and Speech, pp. 143-179, October 1980.

[Horváth, 1993] Horváth, L., *The Maximum Likelihood Method for Testing Changes in the Parameters of Normal Observations*, Annals of Statistics, Vol. 21, No 2., pp. 671-680,1993.

[Huang & kol., 2001] Huang, X., Acero, A., Hon, H., *Spoken Language Processing*, Prentice Hall PTR, 2001.

[Jiříček, 2002] Jiříček, O., *Úvod do akustiky*, Vydavatelství ČVUT, Praha, 2002.

[Kotz & kol., 1992] Kotz, S., Johnson, N. L., *Breakthroughs in Statistics*, Vol. 1, Springer-Verlag, New York, 1992.

[Kwon & Narayanan, 2002] Kwon, S., Narayanan, S.: *Speaker Change Detection Using a New Weighted Distance Measure*, ICSLP 2002, Denver (USA), 2002.

[Lauro & kol, 2002] Lauro, C., Antoch, J., Vinzi, V. E., Saporta G., *Multivariate Total Quality Control*, Physica-Verlag, Heidelberg, 2002.

[Lehmann, 1986] Lehmann, E. L., *Testing Statistical Hypotheses*, 2^{nd} edition, Wiley & Sons, New York, 1986.

[Lu & Zhang, 2002] Lu L., Zhang, H., *Speaker Change Detection and Tracking in Real-Time News Broadcasting Analysis*, In Proceedings of the 10^{th} ACM international conference on Multimedia, Juan-les-Pins, France, 2002.

[Pietquin & kol., 2001] Pietquin, O., Couvreur, L., Couvreur, P., *Applied Clustering for Automatic Speaker-Based Segmentation of Audio Material*, Belgian Journal of Operations Research, Statistics and Computer Science (JORBEL) Special Issue : OR and Statistics in the Universities of Mons, Vol. 41, No. 1-2,01.,2001.

[Press & kol., 2002] Press, W. H., Teukolsky, S. A., Vetterling, W. T., Flannery, B. P., *Numerical Recipes in C - The Art of Scientific Computing*, Second Edition, Cambridge University Press, New York, 2002.

[Schwarz, 1978] Schwarz, G., *Estimating the Dimension of a Model*, Annals of Statistics, Vol. 6, pp. 461–464, 1978.

[Vandecatseye & Martins, 2003] Vandecatseye A., Martens, J. P., *A Fast, Accurate and Stream-based Speaker Segmentation and Clustering Algorithm*, Eurospeech 2003, Geneva (Switzerland) , 2003.

[Vandecatseye & kol, 2004] Vandecatseye, Martens, Neto, Meinedo, Garcia-Mateo, Dieguez, Mihelic, Zibert, Nouza, David, Pleva, Protopapas, Papageorgiou, *The COST278 pan-European database for research Broadcast News transcription*, LREC 2004, Lisbon, Portugal, 2004.

[Vostrikova, 1981] Vostrikova, L. Ju., *Detecting Disorder in Multidimensional Random Processes*, Soviet Mathematics Doklady, 1981.

[Zdansky & kol., ICSLP 2004] Zdansky, J., David, P., Nouza, J., *An Improved Preprocessor for the Automatic Transcription of Broadcast News Audio Stream*, In Proceedings of 8^{th} International Conference on Spoken Language Processing ICSLP 2004, JeJu (South Korea), 2004.

[Zdansky, Radioengineering 2005] Zdansky, J., *Detection of Acoustic Change-Points in Audio Streams and Signal Segmentation*, Radioengineering, Vol. 14, No. 1, pp. 37–40, April 2005.

[Zdansky & Nouza, Interspeech 2005] Zdansky, J., Nouza J., *Detection of Acoustic Change-Points in Audio Records via Global BIC Maximization and Dynamic Programming*, In Proceedings of 9^{th} International Conference on Speech Communications and Technology Interspeech 2005, pp. 669–672, Lisboa (Portugal), 2005.

[Zhou & Hansen, 2005] Zhou, B., Hansen, J. H. L., *Efficient Audio Stream Segmentation via Combined T^2 Statistic and Bayesian Information Criterion*, IEEE Transaction on Speech and Audio Processing, Vol. 13, No. 4, July 2005.

Ostatní použitá literatura

[1] Bernardo, J. M., Smith, A. F. M., *Bayesian Theory*, John Wiley & Sons, England, 1994.

[2] Cichocki, A., Amari, S., *Adaptive Blind Signal and Image Processing*, John Wiley & Sons, England, 2002.

[3] Eckel, B., *Myslíme v jazyku C++*, Grada Publishing, Praha, 2000.

[4] Kriš'ák, M., *Základy teórie stochastických procesov*, AX INZERT, Bratislava, 1998.

[5] Kemeny, J. G., Snell, J. L., Thompson, G. L., *Úvod do finitní matematiky*, SNTL, Praha, 1971.

[6] Nadler, M., Smith, E., P., *Pattern Recognition Engineering*,John Wiley & Sons, England, 1993.

[7] Oetiker, T., Partl, H., Hyna, I., Schlegl, E., *The Not So Short Introduction to LaTeX 2_ε*, on-line dokument na serveru sdružení CTAN, `www.ctan.org`.

[8] Rybička, J., *LaTeXpro začátečníky*, KONVOJ, Brno, 1999.

[9] Satrapa, P., *Perl pro zelenáče*, Neocortex, Praha, 2001.

[10] Seber, G. A. F., *Multivariate Observations*, John Wiley & Sons, England, 1984.

[11] Schlesinger, M. I., Hlaváč, V., *Deset přednášek z teorie statistického a strukturního rozpoznávání*, Vydavatelství ČVUT, Praha, 1999.

[12] Tong, I. L., *The Multivariate Normal Distribution*, Springer-Verlag, New York, 1990.

[13] Uhlíř, J, Sovka, P., *Číslicové zpracování signálů*, vydavatelství ČVUT, Praha, 2002.

[14] Zayezdny, A., Tabak, D., Wulich, D., *Engineering Applications of Stochastic Processes*, John Wiley & Sons, England, 1989.

Seznam vlastních prací

[1] Zdansky, J., Nouza, J., *Detection of Acoustic Change-Points in Audio Records via Global BIC Maximization and Dynamic Programming*, In Proceedings of 9^{th} International Conference on Speech Communications and Technology Interspeech 2005, pp. 669–672, Lisboa (Portugal), 2005.

[2] Nouza, J., Zdansky, J., David, P., Cerva, P., Kolorenc, J., Nejedlova, D., *Fully Automated System for Czech Spoken Broadcast Transcription with Very Large (300K+) Lexicon*, In Proceedings of 9^{th} International Conference on Speech Communications and Technology Interspeech 2005, pp. 1681–1684, Lisboa (Portugal), 2005.

[3] Žibert J., Mihelič F., Martens J. P., Meinedo H., Neto J., Docio L., Garcia-Mateo C., David P., Zdansky J., Pleva M., Cizmar A., Žgank A., Kačič Z., Teleki C., Vicsi K., *The COST 278 Broadcast News Segmentation and Speaker Clustering Evaluation - Overview, Methodology, Systems, Results*, In proceedings of 9^{th} International Conference on Speech Communications and Technology Interspeech 2005, pp. 629–632, Lisboa (Portugal), 2005.

[4] Nouza, J., Červa, P., Žďánský, J., Kolorenč, J., David, P., *Towards Automatic Transcription of Parliament Speech*, In proc. of 16^{th} Conference on Electronic Speech Signal Processing, pp. 237–244, Prague, September 2005.

[5] Zdansky, J., *Detection of Acoustic Change-Points in Audio Streams and Signal Segmentation*, Radioengineering, Vol. 14, No. 1, pp. 37–40, April 2005.

[6] Zdansky, J., *Novel Algorithm for Speaker Segmentation of TV Broadcast News*, In Proc. of Radioelektronika 2005, April 2005, Brno, Czech Republic.

[7] Zdansky, J., David, P., Nouza, J., *An Improved Preprocessor for the Automatic Transcription of Broadcast News Audio Stream*, In Proc. of ICSLP 2004, pp. 1065–1068, Jeju (South Korea), October 2004.

[8] Nouza, J., Nejedlova, D., Zdansky, J., Kolorenc, J., *Very Large Vocabulary Speech Recognition System for Automatic Transcription of Czech Broadcast*, In Proc. of ICSLP 2004, pp. 409–412, Jeju (South Korea), October 2004.

[9] Nouza, J., Zdansky, J., David, P., *Fully Automated Approach to Broadcast News Transcription in Czech Language*, In Proc. of Text, Speech and Dialogue, pp. 401–408, Springer-Verlag, Berlin, 2004.

[10] Zdansky, J., Kroul, M., *Semi-Automatic Non-speech Events Database Formation*, In 8th International Student Conference on Electrical Engineering - POSTER 2004 [CD-ROM], Prague, May 2004.

[11] Zdansky, J., David, P., *Automatic Audio Segmentation of Tv Broadcast News*, In Proc. of Radioelektronika 2004, pp. 358-361, Bratislava (Slovak Republic), April 2004.

[12] Zdansky, J., Nouza, J., *Experimental Optimization of the Continuous Speech Recognition System*, In Proc. of 13^{th} Czech-German Workshop „Speech Processing", pp. 129-134, Prague, September 2003.

[13] Zdansky, J., *Přínos apriorní informace o dialektu pro HMM systémy rozpoznávání češtiny*, Analýza a zpracování signálů IV, Sborník seminářů katedry teorie obvodů, Praha, březen 2003.

[14] Zdansky, J., *Gender dependency of mel-frequency cepstral coefficients*, In 7^{th} International Student Conference on Electrical Engineering - POSTER 2003 [CD-ROM], Prague, May 2003.

[15] Zdansky, J., *Optimalizace struktury HMM*, Analýza a zpracování signálů III, Sborník seminářů katedry teorie obvodů, ISBN 80-01-02726, Praha, 2003.

DODATEK A

VYBRANÁ MATEMATICKÁ ODVOZENÍ

A.1 Odhad ML parametrů vícerozměrného normálního rozložení

Důvodem zařazení této přílohy do disertační práce je skutečnost, že v příliš velkém množství publikací se vyskytuje chybně odvozený vztah pro hodnotu maxima věrohodnosti (A.19).

Předpokládejme, že data

$$\mathbf{x} = \{x_1, x_2, \ldots, x_T\}, \tag{A.1}$$

byla generována nezávislými veličinami X_t. Dále předpokládejme, že tyto veličiny mají stejnou distribuční funkci - d-rozměrné normální rozložení, tj. $\{F_{X_t}(x)\}_1^T \equiv N_d(\mu, \Sigma)$. Vícerozměrná normální veličina X je popsána hustotou pravděpodobnosti

$$f_X(x) = \frac{1}{\sqrt{(2\pi)^d |\Sigma|}} \exp\left[-\frac{1}{2}(x-\mu)^t \Sigma^{-1}(x-\mu)\right], \tag{A.2}$$

kde $d \times d$ rozměrná matice Σ a d-rozměrný vektor μ jsou parametry normální veličiny. Věrohodnost dat $\mathbf{x}$ je pak dána součinem

$$\ell(\mu, \Sigma) = \prod_{t=1}^{T} f_X(x_t). \tag{A.3}$$

Při estimaci parametrů metodou maximální věrohodnosti hledáme takové hodnoty $\hat{\mu}$, $\hat{\Sigma}$, jež maximalizují součin (A.3). Dosazením (A.2) do (A.3) získáme věrohodnost

$$\ell(\mu, \Sigma) = (2\pi)^{-Td/2} |\Sigma|^{-T/2} \exp\left[-\frac{1}{2}\sum_{t=1}^{T}(x_t-\mu)'\Sigma^{-1}(x_t-\mu)\right]. \tag{A.4}$$

Jelikož aplikací logaritmu, neboť je monotónní rostoucí funkcí, se na výsledku maximalizační úlohy nic nemění, je výhodnější hledat maximum logaritmické věrohodnosti:

$$\log \ell(\mu, \Sigma) = L(\mu, \Sigma) = c - \frac{1}{2} T \log |\Sigma| - \frac{1}{2}\sum_{t=1}^{T}\left[(x_t-\mu)'\Sigma^{-1}(x_t-\mu)\right] \tag{A.5}$$

kde $c = -\frac{1}{2}Td\log(2\pi)$.

Podmínkou pro to, aby funkce $L(\mu, \Sigma)$ nabyla maxima, je nulovost derivací podle všech parametrů. Nejprve funkci zderivujeme podle jednotlivých složek vektoru μ:

$$\begin{aligned}\frac{\partial L(\mu,\Sigma)}{\partial \mu_s} &= \frac{1}{2}\sum_{t=1}^{T}\left\{\left[\Sigma^{-1}(x_t-\mu)\right]_s + \left[(x_t-\mu)'\Sigma^{-1}\right]_s\right\} \\ &= \left[\Sigma^{-1}\left(\sum_{t=1}^{T}x_t - T\mu\right)\right]_s, \end{aligned} \tag{A.6}$$

kde $[\cdot]_s$ značí výběr s-té složky vektoru. Z rovnice (A.6) plyne, že má platit:

$$\left[\Sigma^{-1}\left(\sum_{t=1}^{T}x_t - T\mu\right)\right]_s = \mathbf{0}. \tag{A.7}$$

Předpokládejme, že matice Σ je pozitivně definitní, tj. i $\Sigma^{-1} > \mathbf{0}$. Pak může rovnice platit pouze za podmínky $\sum_{t=1}^{T}x_t - T\mu = 0$, z čehož vyplývá, že ML odhad parametru μ bude dán aritmetickým průměrem pozorování:

$$\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T}x_t. \tag{A.8}$$

Nyní zbývá zderivovat $L(\mu, \Sigma)$ podle prvků matice Σ. Neboť $|\Sigma| = 1/|\Sigma^{-1}|$, tj. $\log|\Sigma| = -\log|\Sigma^{-1}|$, bude výhodnější derivovat podle prvků matice $A = \Sigma^{-1}$:

$$\frac{\partial L(\hat{\mu}, A^{-1})}{\partial A_{s,t}} = \frac{T}{2}\frac{1}{|A|}\frac{\partial|A|}{\partial A_{s,t}} - \frac{1}{2}\sum_{t=1}^{T}\left[(x_t-\hat{\mu})'\right]_s\left[(x_t-\hat{\mu})\right]_t. \tag{A.9}$$

Potřebujeme tedy vypočíst derivaci $\frac{\partial|A|}{\partial A_{s,t}}$. Vyjdeme ze vzorce pro výpočet determinantu a rozepíšeme ho podle s-tého řádku:

$$|A| = \sum_{k=1}^{d}(-1)^{s+k}A_{s,k}|A|^{s,k}, \tag{A.10}$$

kde $A_{s,k}$ je prvek matice A na pozici s, k a $|A|^{s,k}$ je determinant matice vzniklé z matice A vypuštěním s-tého řádku a k-tého sloupce. Parciální derivace determinantu (A.10) podle prvku matice $A_{s,t}$ bude nenulová pouze tehdy, když index k v sumě bude nabývat hodnoty t, tj.

$$\frac{\partial|A|}{\partial A_{s,t}} = (-1)^{s+t}|A|^{s,t}. \tag{A.11}$$

Využijme-li předpisu pro výpočet inverzní matice

$$A^{-1}_{t,s} = \frac{(-1)^{s+t}|A|^{s,t}}{|A|}, \tag{A.12}$$

vztahu (A.11) a dosadíme-li je do (A.9), obdržíme výslednou parciální derivaci ve tvaru:

$$\frac{\partial L(\hat{\mu}, A^{-1})}{\partial A_{s,t}} = \frac{T}{2} A_{t,s}^{-1} - \frac{1}{2} \sum_{t=1}^{T} \left[(x_t - \hat{\mu})' \right]_s \left[(x_t - \hat{\mu}) \right]_t . \tag{A.13}$$

Aplikací podmínky

$$\frac{\partial L(\hat{\mu}, A^{-1})}{\partial A_{s,t}} = 0 \quad \forall s, t \tag{A.14}$$

obdržíme

$$A_{t,s}^{-1} = \frac{1}{T} \sum_{t=1}^{T} \left[(x_t - \hat{\mu})' \right]_s \left[(x_t - \hat{\mu}) \right]_t , \tag{A.15}$$

z čehož plyne ML odhad disperzní matice normálního rozdělení

$$\hat{\Sigma} = \frac{1}{T} \sum_{t=1}^{T} (x_t - \hat{\mu})(x_t - \hat{\mu})' . \tag{A.16}$$

V tomto okamžiku zbývá pouze určit hodnotu funkce věrohodnosti v místě jejího maxima $L(\hat{\mu}, \hat{\Sigma})$. Vyjdeme z rovnice (A.5) a upravíme ji do tvaru:

$$\begin{aligned} L(\mu, \Sigma) &= c - \frac{1}{2} T \log |\Sigma| - \frac{1}{2} \sum_{t=1}^{T} \left[(x_t - \mu)' \Sigma^{-1} (x_t - \mu) \right] \\ &= c - \frac{1}{2} T \log |\Sigma| - \frac{1}{2} \sum_{t=1}^{T} \operatorname{tr} \left[(x_t - \mu)' \Sigma^{-1} (x_t - \mu) \right] \\ &= c - \frac{1}{2} T \log |\Sigma| - \frac{1}{2} \operatorname{tr} \left[\Sigma^{-1} \sum_{t=1}^{T} (x_t - \mu)(x_t - \mu)' \right] \end{aligned} \tag{A.17}$$

S využitím ML odhadu disperzní matice (A.16) zjistíme hodnotu maxima:

$$\begin{aligned} L(\hat{\mu}, \hat{\Sigma}) &= c - \frac{1}{2} T \log |\hat{\Sigma}| - \frac{1}{2} \operatorname{tr} \left[\hat{\Sigma}^{-1} T \hat{\Sigma} \right] \\ &= c - \frac{1}{2} T \log |\hat{\Sigma}| - \frac{Td}{2} . \end{aligned} \tag{A.18}$$

Odlogaritmováním pak získáme konečný vztah pro hodnotu věrohodnosti v jejím maximu:

$$\ell(\hat{\mu}, \hat{\Sigma}) = (2\pi)^{-Td/2} |\hat{\Sigma}|^{-T/2} e^{-Td/2} . \tag{A.19}$$

A.2 Nástin postupu odvození BIC

Předpokládejme, že máme k dispozici data $\mathbf{x} = x_1, x_2, \ldots, x_T$, jež jsou realizacemi náhodných veličin $\mathbf{X} = X_1, X_2, \ldots, X_T$. Dále předpokládejme, že je potřeba vybrat model $m = 1, \ldots, M$, jenž by nejlépe popisoval veličinu $\mathbf{X}$. Zvolíme-li jako kritérium výběru nejlepšího modelu věrohodnost dat, pak je cílem nalézt takový model m, jenž maximalizuje věrohodnost $p(\mathbf{x}|m)$. Problém je v tom, že tuto věrohodnost nelze vyčíslit, pročež je nutno do ní zakomponovat odpovídající vektor parametrů modelu - θ_m.

$$\begin{aligned} p(\mathbf{x}|m) &= \frac{p(\mathbf{x}, m)}{P(m)} = \frac{\int p(\mathbf{x}, \theta_m, m) d\theta_m}{P(m)} \\ &= \frac{\int p(\mathbf{x}|\theta_m, m) p(\theta_m, m) d\theta_m}{P(m)} \\ &= \int p(\mathbf{x}|\theta_m, m) p(\theta_m|m) d\theta_m \end{aligned} \tag{A.20}$$

Pro výpočet integrálu (A.20) využijeme *Laplaceovy metody*. Nadefinujeme funkci parametrů θ_m modelu m

$$g(\theta_m) = \log \left[p(\mathbf{x}|\theta_m, m) \cdot p(\theta_m|m) \right] \tag{A.21}$$

a hodnotu $\tilde{\theta}_m$, jež maximalizuje $g(\theta_m)$, nazveme *maximum a posteriori* (MAP) konfigurací parametrů θ_m pro daná data $\mathbf{x}$ a model m. Abychom získali aproximaci $g(\theta_m)$, použijeme *Taylorův rozvoj druhého řádu* v okolí bodu $\tilde{\theta}_m$

$$g(\theta_m) \approx \tilde{\theta}_m - \frac{1}{2}(\theta_m - \tilde{\theta}_m) A (\theta_m - \tilde{\theta}_m)', \tag{A.22}$$

kde $(\theta_m - \tilde{\theta}_m)'$ značí transpozici řádkového vektoru $(\theta_m - \tilde{\theta}_m)$, A je záporně vzatý *Hessián* funkce $g(\theta_m)$ v bodě $\tilde{\theta}_m$. Odlogaritmováním $g(\theta_m)$ obdržíme aproximaci součinu

$$\begin{aligned} p(\mathbf{x}|m) &= \int p(\mathbf{x}|\theta_m, m) p(\theta_m|m) d\theta_m \approx \\ &\approx p(\mathbf{x}|\tilde{\theta}_m, m) p(\tilde{\theta}_m|m) \exp\left[-\frac{1}{2}(\theta_m - \tilde{\theta}_m) A (\theta_m - \tilde{\theta}_m)' \right]. \end{aligned} \tag{A.23}$$

Jelikož aproximace (A.23) nápadně připomíná předpis pro hustotu pravděpodobnosti vícerozměrného normálního rozložení, lze při integraci využít faktu, že $\int f_X(x) dx = 1$, tj.

$$\int \exp\left[-\frac{1}{2}(\theta_m - \tilde{\theta}_m) A (\theta_m - \tilde{\theta}_m)' \right] d\theta_m = (2\pi)^{c/2} \sqrt{|A^{-1}|}. \tag{A.24}$$

Dosazením (A.24) do (A.23) a logaritmováním obdržíme tzv. *Laplaceovu aproximaci* ve tvaru

$$\log p(\mathbf{x}|m) \approx \log p(\mathbf{x}|\tilde{\theta}_m, m) + \log p(\tilde{\theta}_m|m) + \frac{c}{2} \log (2\pi) - \frac{1}{2} \log |A|, \tag{A.25}$$

kde c je rozměr modelu m, tj. počet jeho volných parametrů.

Z rovnice (A.25) získáme další zjednodušení tím, že v ní ponecháme pouze členy, jež rostou s T:

1. $\log p(\mathbf{x}|\tilde{\theta}_m, m)$ roste lineárně s T;
2. $\log |A|$ roste v závislosti na T jako $c \log T$;

Pro $T \rightarrow \infty$ lze navíc $\tilde{\theta}_m$ aproximovat hodnotou $\hat{\theta}_m$, která maximalizuje $p(\mathbf{x}|\theta_m, m)$, tj.ML konfigurací θ_m. Výsledkem je tedy aproximace věrohodnosti dat $\mathbf{x}$ pro daný model m, označovaná jako *Bayesovské informační kritérium*:

$$\log p(\mathbf{x}|m) \approx BIC(\mathbf{x}|m) = \log p(\mathbf{x}|\hat{\theta}_m, m) - \frac{c}{2} \log T \qquad \text{(A.26)}$$

DODATEK B

TABULKY

B.1 Metoda binárního dělení

Databáze	Trénovací		Testovací		
Princip	K	F_{max} [%]	F [%]	R [%]	P [%]
MLLR	68.5	93.25	92.78	92.54	93.02
FTSIC	480	92.21	91.61	90.15	93.11
FPWSIC	1.75	90.31	89.57	89.39	89.75

Tabulka B.1: *Výsledky získané metodou binárního dělení na databázi FS-ART.*

Databáze	Trénovací		Testovací		
Princip	K	F_{max} [%]	F [%]	R [%]	P [%]
MLLR	71.5	95.68	95.57	95.79	95.39
FTSIC	540	95.35	95.37	95.4	95.34
FPWSIC	2.05	92.91	93.15	91.82	94.52

Tabulka B.2: *Výsledky získané metodou binárního dělení na databázi MS-ART.*

Pricip	MLLR			FTSIC			FPWSIC		
Komponenta	F [%]	R [%]	P [%]	F [%]	R [%]	P [%]	F [%]	R [%]	P [%]
CZ	80.39	84.47	76.68	80.12	82.80	77.60	79.27	78.14	80.44
GR	64.50	81.88	53.21	66.00	80.58	55.89	67.82	79.13	59.34
GA	61.36	94.00	45.54	65.93	92.96	51.08	68.83	92.13	54.94
HU	71.59	81.45	63.86	73.44	80.51	67.51	72.87	74.91	70.94
PT	68.59	88.85	55.86	71.10	87.74	59.76	73.65	83.44	65.91
SI	67.18	88.03	54.32	68.64	87.29	56.56	68.96	84.71	58.15
SI2	58.83	83.57	45.39	61.36	80.50	49.57	61.95	74.37	53.08
SK	70.85	76.81	65.74	70.38	73.32	67.66	69.97	68.70	71.28
HR	60.37	90.56	45.28	62.22	89.79	47.60	64.82	88.05	51.29
BE	73.38	82.41	66.13	74.41	79.40	70.01	75.00	74.87	75.13
Σ	68.27	84.56	57.24	69.65	82.88	60.07	70.80	79.06	64.10

Tabulka B.3: *Výsledky získané metodou binárního dělení na databázi COST 278. Trénování metod bylo uskutečněno na databázi ART.*

Část	Trénovací		Testovací		
Komponenta	K	F_{max} [%]	F [%]	R [%]	P [%]
CZ	80.0	81.46	69.41	82.50	59.90
GR	95.5	71.69	72.30	71.66	72.96
GA	104.5	76.53	68.72	62.80	75.88
HU	82.5	73.55	71.62	81.84	63.67
PT	95.5	77.92	71.61	71.08	72.14
SI	93.0	73.46	72.51	72.89	72.14
SI2	92.0	67.25	73.06	74.29	71.87
SK	77.0	71.71	68.75	85.32	57.57
HR	107.5	71.97	68.06	61.08	76.85
BE	82.0	76.04	71.31	81.81	63.20
ϕ	90.95	74.16	70.74	74.53	68.62

Tabulka B.4: *Vyhodnocení cyklického testu na databázi COST 278 pro metodu binárního dělení ve verzi MLLR.*

Část	Trénovací		Testovací		
Komponenta	K	F_{max} [%]	F [%]	R [%]	P [%]
CZ	830	80.94	70.16	81.06	61.84
GR	9500	69.63	72.43	72.69	72.17
GA	990	72.55	71.34	69.67	73.10
HU	800	74.60	70.53	81.79	62.00
PT	1230	78.12	71.74	73.91	69.69
SI	1300	73.40	71.99	72.47	71.52
SI2	10200	66.67	72.57	77.77	68.01
SK	5800	70.74	67.26	85.37	55.48
HR	1980	71.55	65.87	56.33	79.32
BE	890	76.92	71.27	80.27	64.08
ϕ	1057	73.51	70.52	75.13	67.72

Tabulka B.5: *Vyhodnocení cyklického testu na databázi COST 278 pro metodu binárního dělení ve verzi FTSIC.*

Část	Trénovací		Testovací		
Komponenta	K	F_{max} [%]	F [%]	R [%]	P [%]
CZ	2.00	79.48	67.69	80.52	58.39
GR	2.90	70.92	71.94	70.79	73.13
GA	3.45	75.62	68.33	61.93	76.22
HU	2.30	73.11	71.07	78.83	64.70
PT	2.60	77.65	71.58	74.18	69.15
SI	3.00	71.89	71.44	68.82	74.27
SI2	2.95	66.93	71.87	70.16	73.67
SK	2.00	70.21	68.93	81.88	59.51
HR	3.50	71.53	68.39	61.57	76.91
BE	2.60	75.73	71.85	75.24	68.76
ϕ	2.30	73.31	70.31	72.39	69.47

Tabulka B.6: *Vyhodnocení cyklického testu na databázi COST 278 pro metodu binárního dělení ve verzi FPWSIC.*

B.2 Metoda globální maximalizace BIC

Komponenta	F-rate [%]	R [%]	P [%]
CZ	82.52	86.86	78.59
GR	66.63	84.63	54.94
GA	61.12	94.41	45.19
HU	75.62	87.75	66.43
PT	69.56	89.33	56.95
SI	68.67	90.61	55.28
SI2	60.61	84.68	47.20
SK	72.57	80.17	66.29
HR	61.12	93.45	45.41
BE	75.40	86.26	66.97
Σ	70.03	87.43	58.41

Tabulka B.7: *Výsledky získané metodou GMBIC na databázi COST 278. Trénování metod bylo uskutečněno na databázi ART.*

Část	K	F_{max} [%]	F [%]	R [%]	P [%]
CZ	1.54	83.66	72.41	84.17	63.54
GR	2.18	74.28	74.58	73.21	76.01
GA	3.14	81.33	68.25	58.49	81.93
HU	1.60	76.73	73.96	83.64	66.29
PT	2.17	79.38	74.10	72.93	75.32
SI	2.00	75.99	75.09	75.79	74.41
SI2	2.18	70.04	74.85	73.25	76.51
SK	1.70	73.98	75.12	82.82	68.73
HR	3.22	76.05	68.08	58.04	82.33
BE	1.56	79.88	73.41	83.94	65.23
ϕ	1.80	77.13	72.99	74.63	73.03

Tabulka B.8: *Vyhodnocení cyklického testu na databázi COST 278 pro metodu GMBIC.*

B.3 Metoda s adaptivním oknem

Mtd.	Modifikovaná						Originální	
B[s]	F[%]	R[%]	P[%]	NNO	AVD[s]	MXD[s]	F[%]	NNO
1	94.35	97.04	91.81	0.26	1.61	7.72	94.36	0.24
2	95.15	96.88	93.49	0.15	1.81	7.76	95.13	0.13
3	95.70	96.75	94.68	0.12	2.05	7.95	95.61	0.10
4	95.92	96.61	95.24	0.10	2.27	7.95	95.73	0.08
5	96.32	96.65	95.99	0.09	2.52	8.74	96.11	0.07
6	96.47	96.63	96.32	0.08	2.76	8.48	96.18	0.06
7	96.71	96.62	96.80	0.07	3.15	9.14	96.07	0.05
8	96.84	96.44	97.25	0.07	3.60	9.21	95.72	0.05
9	96.75	96.33	97.18	0.07	4.14	9.62	94.83	0.05
10	96.88	96.30	97.46	0.07	4.81	10.16	93.45	0.05
11	96.81	96.16	97.47	0.07	5.52	10.23	91.95	0.05
12	96.86	96.11	97.63	0.07	6.32	11.31	90.00	0.05
13	96.77	95.88	97.68	0.08	7.17	12.03	87.95	0.05
14	96.90	96.05	97.77	0.08	8.02	12.80	86.00	0.05
15	96.92	95.98	97.87	0.08	8.89	13.54	83.89	0.05
20	96.90	95.85	97.98	0.11	13.55	18.17	73.99	0.07
30	96.85	95.76	97.96	0.17	23.20	28.12	58.41	0.10

Tabulka B.9: *Závislost úspěšnosti metody MAWIN na délce inicializačního okna B a její srovnání s originální metodou AWIN.*

Komponenta	F-rate [%]	R [%]	P [%]
CZ	81.54	83.63	79.55
GR	66.62	82.85	55.71
GA	63.57	94.82	47.81
HU	71.83	80.63	64.76
PT	70.23	87.90	58.47
SI	68.33	88.58	55.61
SI2	61.10	83.57	48.15
SK	70.49	75.81	65.87
HR	62.04	91.33	46.98
BE	74.32	81.91	68.01
Σ	69.50	84.33	59.11

Tabulka B.10: *Výsledky získané metodou MAWIN na databázi COST 278. Trénování metod bylo uskutečněno metodou MLLR na databázi ART.*

Část	Trénovací		Testovací		
Komponenta	K	F_{max} [%]	F [%]	R [%]	P [%]
CZ	80.0	81.46	71.51	82.80	62.93
GR	95.5	71.69	72.75	71.31	74.25
GA	104.5	76.53	69.78	63.37	77.63
HU	82.5	73.55	72.61	81.75	65.31
PT	95.5	77.92	72.46	71.10	73.88
SI	93.0	73.46	73.02	72.56	73.48
SI2	92.0	67.25	73.40	73.79	73.01
SK	77.0	71.71	70.16	85.24	59.61
HR	107.5	71.97	68.73	61.13	78.49
BE	82.0	76.04	72.10	81.37	64.73
ϕ	90.95	74.16	71.65	74.44	70.33

Tabulka B.11: *Vyhodnocení cyklického testu na databázi COST 278 pro metodu MAWIN.*