

AKUSTICKÉ LISTY

České akustické společnosti
www.czakustika.cz

ročník 16, číslo 1

březen 2010

Obsah

Usnesení Valné hromady České akustické společnosti	3
1. duben 2010 – Den uvědomění si hluku (International Noise Awareness Day)	4
Tvorba rozpoznávače plynulých promluv v českém jazyce standardními nástroji HTK Creation of Continuous Speech Recogniser for Czech Using Standard HTK Tools <i>Josef Rajnoha, Václav Procházka a Petr Pollák</i>	5
Prínos Alana D. Blumleina v oblasti stereofonického zvukového záznamu Contribution of Alan Dower Blumlein to Stereophonic Sound Recording <i>Pavol Brezina</i>	11

ČESKÁ AKUSTICKÁ SPOLEČNOST

Usnesení Valné hromady České akustické společnosti, konané dne 21. ledna 2010 v prostorech Fakulty elektrotechnické ČVUT

Valná hromada ČsAS bere na vědomí:

1. zprávu o činnosti Rady ČsAS;
2. zprávy o činnosti jednotlivých odborných skupin a o jejich dalším zaměření;
3. zprávu o přípravě 80. akustického semináře;
4. zprávu o výsledcích revize hospodaření společnosti;
5. výsledky voleb do Rady společnosti a výsledky voleb předsedů odborných skupin;
6. zprávu o usnášenišchopnosti Valné hromady, 48 / 83;
7. informace o plnění úkolů a poslání Akustických listů a vyzývá členy k zasílání příspěvků.

Pro funkční období roku 2010 byli v jednotlivých odborných skupinách zvoleni:

- A. Obecná, lineární a nelineární akustika
předseda – M. BEDNAŘÍK zástupce – M. ČERVENKA
- B. Ultrazvuk a akustické emise
předseda – J. PLOCEK zástupce – R. BÁLEK
- C. Hluk a stavební akustika
předseda – J. KOZÁK zástupce – M. MELLER
- D. Zpracování a záznam akustických signálů
předseda – T. SALAVA
- E. Psychoakustika, fyziologická akustika a akustika hudby a řeči
předseda – volba bude provedena dodatečně
- F. Elektroakustika
předseda – Z. KEŠNER zástupce – B. SÝKORA

Valná hromada ČsAS schvaluje:

1. zprávu o činnosti Rady za kalendářní rok 2009 a uděluje Radě absolutorium;
2. zprávu o hospodaření společnosti za kalendářní rok 2009;
3. výši členských příspěvků na rok 2010 (400 Kč pro členy, 100 Kč pro studenty a důchodce);
4. činnost Rady a odborných skupin v roce 2009.

Valná hromada ukládá nově zvolené Radě společnosti na kalendářní rok 2010:

1. pokračovat v odborné a organizační činnosti i v zahraničních kontaktech, v rozvíjení spolupráce s Českou matricí technickou, Slovenskou akustickou společností, společností Elektra, českou sekcí AES, Českým centrem IEE, československou sekcí IEEE, ICA, EAA a I-INCE;
2. věnovat pozornost pořádání odborných akcí a pravidelných seminářů odborných skupin;
3. nadále rozvíjet vydávání Akustických listů;
4. usilovat o konání kongresu Euronoise v roce 2012.

Valná hromada ukládá nově zvoleným předsedům odborných skupin na kalendářní rok 2010:

1. publikovat informace o připravovaných aktivitách skupin v Akustických listech a na webové stránce s předstihem tak, aby se zájemci mohli včas na akce přihlašovat.

Valná hromada doporučuje Radě ČsAS:

1. pravidelně se zabývat činností a plánem akcí odborných skupin;
2. pravidelně se zabývat plánem a zaměřením konaných akustických konferencí;
3. poskytovat možnost finančních výhod členům společnosti, např. nižšími sazbami vložného na akcích pořádaných společností.

Výsledky voleb do Rady České akustické společnosti:

Počet vydaných platných hlasovacích lístků:		48
Odevzdáno platných hlasovacích lístků:		48
předseda:	O. JIŘÍČEK	48
místopředseda:	V. KUNZL	48
sekretář:	M. BROTHÁNEK	48
hospodář:	O. KUDĚJOVÁ	48
revizní komise:	J. KOZÁK	48
	T. HELLMUTH	48
	D. POTUŽNÍKOVÁ	48

Návrh usnesení sestavila návrhová komise ve složení M. Meller a P. Urban.

Vážení kolegové,

jak se můžete dočíst v usnesení Valné hromady otištěném v tomto čísle Akustických listů, i v letošním roce zůstávají členské příspěvky nezměněny. Pro výdělečně činné členy tedy činí 400 Kč a pro důchodce a studenty 100 Kč. Pokud jste ještě nezaplatili, je nejlepším způsobem platba převodem z účtu na účet, tedy z Vašeho účtu v kterékoli bance na účet ČsAS, Komerční banka, a. s., Praha 6, číslo účtu 17838061/0100. Platbu můžete provést ze svého soukromého účtu, který nese název Vašeho příjmení. V případě, že platíte z jiného účtu, použijte prosím variabilní symbol ve tvaru XXX09, kde XXX je Vaše osobní číslo, které naleznete nad svým jménem na obálce. Dalším možným způsobem je složení hotovosti na přepážce Komerční banky. Na kterékoli pobočce Komerční banky můžete zaplatit své členské příspěvky složením hotovosti na účet číslo 17838061/0100. Variabilní symbol je shodný jako v předešlém případě. Finančně oboustranně nejméně výhodným způsobem je platba na poště složenkou typu A (zelená). I zde je variabilní symbol shodný jako v předešlých případech.

Marek Brothánek
sekretář společnosti

1. duben 2010 – Den uvědomění si hluku (International Noise Awareness Day)

Není to „april“. Letos Zelený velikonoční čtvrtek připadá na 1. dubna 2010. Jako každý rok podle křesťanské tradice „uletí zvony do Říma“ a je – mělo by být – ticho! Většina lidí nemá ráda hluk, ale skoro nikdo si ho „neuvědomuje“. Samozřejmě pokud nás nezbudí, nebo naopak nenechá usnout. Hluk je doprovodným jevem naší činnosti. Z hlediska lidského není ani důležité, jestli jde o hladinu akustického tlaku 50 nebo 60 dB. Záleží na ohledu k našim „bližním“. Jestliže si hluk okolo nás uvědomíme, budeme ohleduplnější, hluchému místu se vyhneme, nebo dokonce „ztišíme“ reprodukováný zvuk, pomůžeme nejen sobě, ale i ostatním zvýšíme jejich imunitu vůči nemocem všeho druhu, stresu především. Jako multiprofesní sdružení „uvědomujících si“ dovolujeme si vyzvat Vás, profesionální akustiky, zvukaře a odborníky na hluk, abyste se i vy připojili a třeba podpořili osvětu v tomto směru. Např. v restauraci, prodejně požádejte na Zelený čtvrtek o vypnutí nebo ztišení hudby.

Jan Stěnička
předseda Sdružení Hudekos

Tvorba rozpoznávače plynulých promluv v českém jazyce standardními nástroji HTK

Josef Rajnoha, Václav Procházka a Petr Pollák

Katedra teorie obvodů, ČVUT–FEL, Technická 2, 166 27 Praha 6
email: {rajnojos;prochva1;pollak}@fel.cvut.cz

This paper describes the procedure of LVCSR construction for Czech language which is based on basic tools from HMM Toolkit (HTK). Standard design of English LVCSR explained in HTK documentation is supplemented here by language specific peculiarities related to Czech LVCSR creation. The paper brings the overview of particular steps required for the creation of a system which can be easily used as a first step in LVCSR research. Although it is not an optimal solution, especially from the point of view of achieved speed and accuracy, the usage of HTK tools provides high flexibility in the testing of different modifications of particular LVCSR modules. The paper also describes basically the training of context-dependent cross-word triphone HMMs and statistical language model generation with possible optimization of its performance. All scripts for LVCSR training and testing which are described in this paper are publicly available. Finally, the experiments on parameter setting for recognition time and accuracy balance are presented. Concerning the speed, proposed system gives currently real-time factor in the range between 1.5 and 2 with acceptable accuracy for medium-sized vocabulary recognition task.

1. Úvod

Téma rozpoznávání mluvené řeči strojem je již spoustu let předmětem výzkumu na mnohých pracovištích po celém světě. S rostoucí výkonností přístrojů, které člověka obklopují při běžných činnostech (PC, mobilní telefony), roste i složitost implementovaných technologií a náročnost jejich uživatelů. Dřívější rozpoznávače základních povelů jsou dnes střídány diktovacími stroji, přepisujícími automaticky plynulou promluvu do textové podoby například pro účely titulování, tvorby poznámek či přepisu záznamů z oblasti soudnictví, zdravotnictví apod. Mnoho takových systémů je již využíváno v praxi, a to nejen pro angličtinu (např. [1]), ale i pro méně rozšířené a mnohdy více „ohebné“ jazyky, mezi něž řadíme i češtinu [2–4].

Obecně lze princip rozpoznávání spojitě řeči s velkým slovníkem (LVCSR – Large Vocabulary Continuous Speech Recognition) symbolicky zobrazit např. schématem na obr. 1. Jednotlivé funkční bloky zahrnují problematiku z mnoha oblastí, od fyziologie hlasového traktu přes analýzu akustických signálů až po lingvistický rozbor jazyka. Základní princip je založen na hledání nejvěrohodnější kombinace modelů základních akustických elementů z pohledu obecného jazykového modelu spolu s výslovnostním slovníkem nesoucím informaci o posloupnosti jednotlivých akustických elementů v jednotlivých slovech. K modelování akustických jednotek bývá nejčastěji použito skrytých Markovových modelů (Hidden Markov Model – HMM), jazykový model rozpoznávače s velkým slovníkem je nejčastěji reprezentován statistickým modelem pravděpodobnosti výskytu určité kombinace slov v daném jazyce.

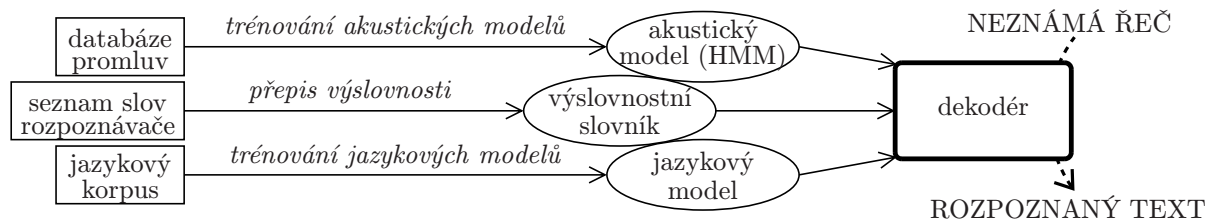
Implementace LVCSR systémů představují velmi složitou komplexní aplikaci, která sdružuje všechny potřebné kroky do jednoho celku, a především zahrnují mnoho optimalizací, umožňujících ve svém důsledku realizaci rozpo-

znávání v reálném čase, viz např. vznikající projekty [5]. Pro výzkumné účely je však často potřeba analyzovat či modifikovat v rámci funkčního LVCSR systému pouze jednotlivé moduly, a proto pro účely snadnějšího přístupu k dílčím funkčním blokům je použití některého z volně dostupných modulárních systémů velmi výhodné, ačkoliv bude většinou jeho výkon výrazně menší.

K nejčastěji používaným systémům patří sada nástrojů HTK (HMM ToolKit) [9], existují ale i další dostupné systémy jako Sphinx [6], Julius [7] či ISIP [8]. S ohledem na komplexnost úlohy rozpoznávání spojitě řeči vznikly i další nástroje, které doplňují systém HTK a pracují s odpovídajícím formátem dat. Z nich je nutno jmenovat zejména dekodér *HDecode*, nahrazující nástroj *HVite* pro účely rozpoznávání řeči s velkým slovníkem.

Přestože byl k nástrojům HTK vytvořen rozsáhlý manuál (*HTKBook*) [10] popisující jednotlivé nástroje a teoretické pozadí problematiky rozpoznávání řeči a přestože obsahuje i návodný postup, v němž je popsána tvorba základního rozpoznávače řeči pro anglický jazyk, není realizace vývojového funkčního systému zcela snadným procesem. Především v situaci, kdy uživatel z důvodu použití jiného jazyka či formátu některých dat nepostupuje zcela standardním způsobem, může se výsledná realizace výrazně lišit od standardního postupu.

Tento článek přináší návod pro tvorbu rozpoznávače spojitých promluv, který plně využívá nástrojů HTK, se zdůrazněním specifických požadavků pro realizaci rozpoznávače spojitě řeči pro češtinu s možným použitím obecně i pro jiný jazyk. Výstupem je i soubor skriptů zahrnující kompletní sadu nástrojů od podpory přípravy řečových dat pro trénování modelů až po vlastní dekodér, realizující proces rozpoznávání spojitě řeči s velkým slovníkem v základní podobě.



Obrázek 1: Obecné blokové schéma rozpoznávače spojitě řeči

Článek rozšiřuje návod přípravy rozpoznávače řeči pro angličtinu uvedený v *HTKBook* a doplňuje též zprávu [11], která se zabývá přípravou akustických modelů a tvorbou jednoduchého rozpoznávače povelů v češtině. Proto se tento příspěvek zaměřuje především na rozdílné, resp. doplňující kroky, které nejsou uvedeny v těchto dvou pramenech a které jsou případně specifické pro konstrukci LVCSR pro češtinu. Hlavním cílem je doplnit standardní postup tvorby rozpoznávače o konkrétní poznatky a zkušenosti, které souvisí s použitím jiného jazyka, a ucelit jednotlivé kroky postupu tvorby rozpoznávače spojitých promluv pomocí dekodéru *HDecode*.

Článek je rozdělen do čtyř základních částí popisujících tvorbu akustických modelů, přípravu slovníku výslovnostních variant, tvorbu jazykového modelu a nakonec realizaci vlastního rozpoznávače.

2. Příprava akustických modelů

Původní návod k HTK i zpráva [11] obsahují podrobný popis tvorby HMM monofonů (kontextově nezávislých modelů fonémů) a trifonů (modely fonémů zahrnující kontextovou závislost zprava i zleva). V dalším textu jsou proto zmíněny jen významné úpravy postupu, jež jsou dány použitým jazykem a použitím *cross-word* trifonů s mezislovní závislostí, které vyžaduje použitý dekodér *HDecode*.

2.1. HMM monofonů

Základní tvorbu HMM modelů monofonů v českém jazyce shrnuje práce [11]. V té je také zvolena sada kódů popisujících jednotlivé fonémy. Značky fonémů je nutné volit co nejjednodušší, aby nebyly použitými nástroji nesprávně interpretovány. Není vhodné používat značky obsahující písmena s diakritikou obsažená ve standardně používaných kódových sadách ISO8859-2, CP-1250 či Unicode. Avšak ani použití abecedy Sampa není vhodné, neboť dochází k dezinterpretaci kódů obsahujících znak „\“, který je používán při kódování nestandardních znaků (např. znaků s diakritikou). Místo kódu „J\“ pro foném „ď“ volíme tedy raději kód „dd“. S podobnou logikou jsou vytvořeny značky i pro ostatní fonémy, takže výsledná sada obsahuje pouze neakcentované znaky „a-z“. S pomocí těchto kódů jsou zapsány i výslovnostní přepisy slov obsažených ve slovníku rozpoznávače (viz sekce 4).

2.2. HMM kontextově vázaných fonémů – trifonů

Pro rozpoznávání spojitě řeči je nezbytné pracovat s kontextově závislými fonémy – trifony. Tyto akustické ele-

menty získáme doplněním pravého a levého kontextu. HTK ovšem nabízí dva způsoby přístupu k modelování kontextové závislosti s ohledem na zahrnutou informaci o hranicích slov s mezislovní kontextem (*cross-word*) či pouze s vnitřním kontextem (*word-internal*), viz obr. 2. Návodný postup v *HTKBook*, stejně jako zpráva [11] popisují postup tvorby trifonů s vnitřním kontextem, jež hranice slov zachovávají. Dekodér *HDecode* použitý v níže popisaném rozpoznávači spojitých promluv však pro svou činnost vyžaduje práci s trifony s mezislovní kontextem, proto bude v dalším textu popsán postup tvorby trifonů bez uvažování hranic slov.

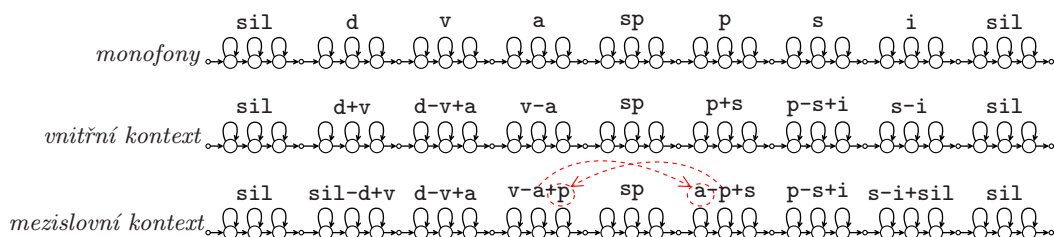
Před převodem HMM monofonů na modely trifonů je potřeba převést výslovnostní přepisy trénovacích promluv (*phone.mlf*) do trifonové podoby tak, jak je popsáno v *HTKBook*. K tomu lze použít nástroj *HLED* s níže uvedenou konfigurací pro tvorbu trifonů s mezislovní kontextem (*mktri.led*), kterou lze nalézt v ukázkových skriptech dodávaných s HTK (*HTK samples*).

Word-internal:	Cross-word:
WB sil	NB sp
WB sp	TC
TC	IT
	CH sp * sp *
	CH sil * sil *

Na rozdíl od word-internal převodu, kde se určí modely pauz *sp* a *sil* jako oddělovače slov (příkaz *WB*), je nutné naopak pro *cross-word* trifony tento oddělovač v prvním kroku ignorovat (*NB*). Po převodu (*TC*) je pro vzniklý trifonový přepis potřeba opět vrátit zpět informaci o pauzách mezi slovy (*CH*), ovšem bez trifonového kontextu (*IT*). Výsledkem je přepis, kde se na okrajích slov nevyskytují bifony, ale kontext plynule přechází do dalšího slova, viz trifon *v-a+p* na konci slova „dva“ v následujícím přepisu.

slova:	word-internal trifony:	cross-word trifony:
sil	sil	sil
dva	d+v	sil-d+v
sp	d-v+a	d-v+a
psi	v-a	v-a+p
sil	sp	sp
	p+s	a-p+s
	p-s+i	p-s+i
	s-i	s-i+sil
	sil	sil

V dalších krocích jsou již standardním způsobem svázány některé stavy modelů jednotlivých trifonů pro snížení



Obrázek 2: Schematické naznačení kontextové závislosti a nezávislosti fonémů

složitosti a velikosti modelů. Rovněž je možné zvýšit efektivitu modelování rozdělením hustoty pravděpodobnosti na více směsí (*mixtures*).

Pro snazší implementaci adaptačního procesu pomocí nástrojů HTK není voleno v popisovaných modelech dělení vektoru příznaků na menší vektory (*streams*).

3. Příprava jazykového modelu

Jazykový model vystihuje obecně vzájemnou pozici slov ve větě a jeho složitost významně ovlivňuje nejen rozpoznávací skóre systému, ale také jeho výkon. Proto je při jeho tvorbě potřeba hledat kompromis mezi složitostí modelu a efektivitou rozpoznávacího procesu.

Nejjednodušší formou, jak popsat skladbu jazyka, je vytvoření gramatiky, podle které jsou jednotlivá slova promluvy skládána za sebou, se stejnou pravděpodobností. Tato volba je vhodná pro situace, kdy je použit velmi malý slovník. Gramatika tak nachází využití při tvorbě rozpoznávače povelů, hlásek či sekvence číslovek.

HTK obsahuje nástroje pro snadnou tvorbu takové gramatiky spolu s podrobným popisem a ukázkovým příkladem gramatiky pro hlasové vytáčení. Tak je vytvořena síť, jejímž průchodem jsou určeny možné podoby výstupního textu. Jelikož nelze nastavit pravděpodobnost přechodu mezi jednotlivými slovy sítě, jsou všechna použitá slova stejně pravděpodobná. To ovšem neodpovídá složení běžné spojitě promluvy.

Proto rozpoznávače plynulých promluv používají statistický jazykový model, který popisuje pravděpodobnost výskytu n -tice slov (n -gram) ve větě. Tato pravděpodobnost se získá na základě předložené sady trénovacích textů. To umožňuje vytvářet tematicky zaměřené modely pro jednotlivé oblasti použití rozpoznávače.

3.1. Trénovací materiál jazykového modelu

Pro natrénování kvalitního statistického jazykového modelu je potřeba předložit systému dostatečně reprezentativní soubor dat, která vystihují podobu jazyka. Přestože může být rozpoznávač použit pro specifickou aplikaci (soudní přepisy, medicínské výrazy apod.), trénovací materiál nemusí být takto přesně zaměřen, neboť a) může být později adaptován na specializovaných textech, b) umožňuje zahrnout a rozpoznávat i neznámá slova, která se nevyskytují v trénovací sadě.

Jako trénovací materiál slouží nejčastěji volně přístupné internetové zdroje, především zpravodajské servery. S ohledem na často různorodé zdroje textových dat je ovšem potřeba korpus pro trénování jazykového modelu připravit. To zahrnuje především:

- sjednotit často různorodé kódování češtiny,
- odstranit informace nesouvisející s významovým textem (HTML kód, reklamní bannery apod.),
- převést číslice či zkratky do slovní podoby (viz např. [14]),
- získaná data rozdělit na jednotlivé věty,
- rozdělit věty podle speciálních pravidel (přímá řeč, závorky),
- odstranit interpunkci, závorky či další speciální znaky (nedělitelné mezery, nepísmenné znaky),
- překontrolovat výsledný text vhodným nástrojem na kontrolu pravopisu,
- v souladu se standardním postupem doplnit značky pro začátek (`<s>`) a konec věty (`</s>`).

Vhodným kódováním je především takové, které používá pro dané znaky jen jeden kód (ISO8859-2, CP-1250), neboť nástroje HTK používají pro specifické znaky češtiny vlastní reprezentaci založenou právě na tomto kódu.

3.2. Trénování jazykového modelu

V souladu s prvními kroky tvorby jazykového modelu pro angličtinu je na začátku vytvořen prázdný model (nástroj *LNewMap*), který je následně doplněn statistikami z trénovacích dat. To znamená zahrnutí výskytu a četnosti všech unigramů, bigramů, příp. n -gramů vyšších řádů.

Poté musí následovat omezení slovníku. To může být založeno na informaci o četnosti výskytu slov v trénovací množině (např. 5 000 nejčastěji se vyskytujících slov), kterou lze získat setříděním dat ve vygenerované mapě *wmap*. Stejně tak lze použít i analýzy jiných zdrojů či jiná kritéria výběru. Výsledkem je vytvořená množina OOV (Out Of Vocabulary) slov, do které spadají všechna méně významná slova. Těm je v rámci dalšího trénování přiřazena jistá pravděpodobnost výskytu, aby mohla být zahrnuta ve slovníku pro rozpoznávač (viz sekce 4).

S použitím omezeného slovníku a předchozí mapy se vytvoří nová mapa a z ní již lze standardním postupem vytvořit jazykový model. Při této tvorbě se zadává parametr *cut-off*, kterým lze dále ovlivnit velikost modelu, neboť umožňuje odstranit z modelu položky (bigramy, trigramy) s nízkou četností výskytu.

Stupeň modelu	Velikost modelu (MB)	Čas (min)	Přesnost rozpoznání (%)	real-time faktor
2-gram	30	19,2	70,7	1,25
3-gram	85	21,8	66,8	1,42

Tabulka 1: Porovnání doby a přesnosti rozpoznávání pro různý stupeň modelu

Tak jako v průběhu trénování akustických modelů můžeme sledovat postupný vývoj modelů v jednotlivých fázích jejich přípravy testováním na testovací databázi, lze pro dané nastavení parametrů jazykového modelu sledovat jeho vlastnosti, tedy schopnost modelovat text, který je předložen. K tomu účelu se sleduje tzv. perplexita (viz např. [10]), která odhaduje pravděpodobnost prisouzenou předloženému textu daným modelem. Pro účely srovnání jazykových modelů s různými parametry pomocí perplexity (*p-plex*) byl v následujících experimentech použit text povídky F. Kafky „Proměna“ obsahující 13 500 slov, z toho 4 320 unikátních.

3.3. Volitelné parametry LM

V průběhu tvorby jazykového modelu lze nastavením některých parametrů ovlivnit výsledné vlastnosti modelu. Především se jedná o stupeň *n*-gramů, míru vyhlazení modelu a velikost trénovacího slovníku.

Volba stupně modelu

Při volbě stupně jazykového modelu záleží na rozsahu dat pro trénování a složitosti cílového jazyka. Nejčastější volbou pro úlohu rozpoznávání řeči s velkým slovníkem je použití bigramového či trigramového modelu.

Lze očekávat, že bigramový model bude dosahovat menší velikosti na úkor přesnosti modelování struktury jazyka, a tedy i nižší výsledné přesnosti rozpoznávání spojitě řeči. To platí v situaci, kdy je trénovací množina dostatečně reprezentativní materiál pro daný jazyk. Především pro inflektivní jazyky je ovšem tato závislost velmi podstatná. Trigramový model tak může přinést zvýšení přesnosti rozpoznávání v řádu jednotek procent, avšak za cenu výraznějšího nárůstu výpočetní náročnosti [12].

Experimenty ale ukazují [13], že příspěvek trigramového modelu k přesnosti rozpoznávání může být velmi malý či dokonce negativní, což dokládá i experiment s jazykovým modelem trénovaným na textech z internetových zpravodajských serverů o celkovém objemu cca 310 MB (640 k unikátních výrazů). Uvedené množství dat je v porovnání s jinými pracemi (2 M unikátních výrazů [14]) velmi nízké. Dosažené výsledky na úloze se 139 promluvami o celkové délce 15 minut v tab. 1, kdy je při použití trigramů dokonce snížena přesnost rozpoznávače, tak potvrzují význam trénovací množiny pro tvorbu jazykového modelu.

Omezení slovníku

Omezením slovníku může dojít k významnému snížení velikosti jazykového modelu, ovšem na úkor jeho přesnosti [14–16]. Je-li omezení provedeno na základě četnosti slov,

	cut-off	OOV (%)	2-gram		3-gram	
			p-plex	2-gramy	p-plex	3-gramy
60 k	1	12,68	1302	4,7 M	1473	10,4 M
	3	12,68	1323	3,2 M	1441	6,7 M
	6	12,68	1322	2,0 M	1316	3,2 M
100 k	1	7,34	2063	5,5 M	2380	11,0 M
	3	7,34	2084	3,7 M	2301	7,0 M
	6	7,34	2060	2,2 M	2071	3,3 M
200 k	1	4,65	2615	6,0 M	3008	11,2 M
	3	4,65	2651	3,9 M	2910	7,1 M
	6	4,65	2607	2,3 M	2620	3,3 M

Tabulka 2: Srovnání jazykových modelů s různým omezením slovníku pro předložený text

lze takto také eliminovat některé překlepy či méně obvyklá slova obsažená v trénovacím materiálu (nářečí, slang apod.).

Návodný postup v *HTKBook* přináší ukázkou, v níž je slovník oříznut na 5 000 slov, což je ale pro účely tvorby kvalitnějšího modelu češtiny nedostačující. S využitím výše zmíněných trénovacích dat přesahuje podíl OOV slov pro takto omezený slovník 32%. Tabulka 2 ukazuje srovnání bigramového a trigramového modelu pro úroveň 60 k, 100 k a 200 k slov, které odpovídají minimální četnosti výskytu slova 10, 30 a 100 ve výše uvedené sadě dat ze zpravodajských serverů.

Je zřejmé, že velikost slovníku přispívá především k snížení podílu OOV slov. Charakter perplexity neumožňuje poměřovat modely podle omezení slovníku (průměrná pravděpodobnost v modelu s větším slovníkem je nižší, perplexita pak vychází vyšší).

Redukce počtu *n*-gramů

Při tvorbě modelu lze snížit složitost modelu odstraněním *n*-gramů, které mají velmi nízkou četnost v trénovací sadě dat. Nástroj *LFoF* s konfiguračním parametrem `LPCALC: TRACE=3` generuje tabulku, která ukazuje počet *n*-gramů, které v jazykovém modelu zůstanou pro danou hodnotu *cut-off*. Tím lze před vlastní tvorbou modelu získat představu o rozsahu zjednodušení.

V tabulce 2 jsou shrnuty perplexity a počet *n*-gramů pro různou hodnotu *cut-off*. Na rozdíl od volby velikosti slovníku nepřispívá *cut-off* výrazně ke změně perplexity, zatímco změna počtu *n*-gramů, a tím velikost celého modelu, exponenciálně klesá.

3.4. Další možnosti vylepšení

Během tvorby jazykového modelu lze realizovat další kroky, které mohou vést ke zvýšení efektivity modelu. Mezi časté úpravy patří:

Klastrování: Setříděním slov do menšího počtu skupin (tříd), u nichž předpokládáme stejný kontextový význam, lze dosáhnout významného snížení počtu *n*-gramů, a tím i velikosti modelu. Sada nástrojů HTK umožňuje provést proces automatického klastrování do daného počtu tříd na základě statistiky z trénovacích dat. Toto dělení je ovšem velmi závislé právě na trénovacím materiálu, což může

vést na velmi neefektivní dělení, které sníží schopnost modelu popsat daný jazyk. Proto se namísto automatického klastrování provádí také dělení např. podle slovních druhů a dalších kritérií. Víme-li například, že se některé předložky pojí jen s určitým pádem a jsou ve větě vždy před podstatným jménem, lze tyto seskupit do jedné třídy a poté již v n -gramech pracovat pouze s touto třídou.

Omezení krátkých slov: Vytvořením modelu bez krátkých slov (především předložky) lze zamezit jejich nadhodnocení v rámci modelu v důsledku jejich zvýšené četnosti [17], které může vést ke zvýšení chybovosti rozpoznávače vkládáním těchto slov. Jejich zahrnutí do rozpoznávacího procesu zajistí přítomnost ve výslovnostním slovníku a vyhlazovací algoritmus při trénování jazykového modelu.

4. Výslovnostní slovník

Slovník v jazykovém modelu zahrnuje nejčastější výrazy, které se v jazyce vyskytují, a vystihuje tak nejčastější gramatické struktury. Naproti tomu slovník pro dekodér definuje množinu všech slov, jež se mohou vyskytnout na výstupu rozpoznávače a spolu s tím i jejich výslovnost.

Připomeňme, že při tvorbě jazykového modelu v sekci 3 byl počet slov zahrnutých do modelu omezen, čímž vznikla skupina OOV slov s určitou pravděpodobností výskytu. To umožňuje použít daný jazykový model i na slova, která nebyla v trénovacím korpusu (případně v omezeném slovníku pro trénování jazykového modelu) obsažena, a to právě definicí slova ve výslovnostním slovníku. Ten může navíc obsahovat i různé výslovnostní varianty, které definují možnou odlišnou výslovnost některých slov, viz např.

...		
čtyři	cc t i rr i	výslovnost „čtyři“
čtyři	cc t i r i	„čtyry“
čtyři	ss t i rr i	„štyři“
čtyři	ss t i r i	„štyry“
...		

Specifickým rysem dekodéru *HDecode* je automatické přidávání modelu pauzy za každé slovo. Ve výslovnostním slovníku tedy v tomto případě není možné za slovy používat oddělovací modely krátké pauzy, jak tomu je při použití dekodéru *HVite*. Navíc musí být ve slovníku obsažena výslovnostní varianta pro začátek a konec promluvy, standardně nastavené jako `<s>` a `</s>`. Tyto značky lze v konfiguračním souboru pro *HDecode* předefinovat.

5. Vlastní rozpoznávač spojitých promluv

Je-li připraven funkční model akustických jednotek, jazykový model a výslovnostní slovník, je možné přistoupit k vlastní konstrukci rozpoznávače. Jako dekodér bude tedy použit diskutovaný dekodér *HDecode*, určený pro použití s velkým slovníkem a komplexnějším jazykovým modelem.

5.1. HDecode a jeho vlastnosti

HDecode je nástroj, který může pracovat pouze v konkrétní konfiguraci a s určitými omezeními. To bylo disku-

toováno již výše ve fázi přípravy modelů. Základní vlastnosti tohoto dekodéru je možné shrnout v následujících bodech:

- o pracuje výhradně s modely trifonů s mezislovním kontextem,
- o modely pauz je nutné značit kódy `sil` a `sp`, tyto modely jsou pak automaticky vkládány dekodérem,
- o v daném okamžiku pracuje dekodér s nejvyšším trigramovým jazykovým modelem.

Pro běh dekodéru je možné dále nastavit parametry, které mohou omezit použité prostředky během výpočtů. Tím lze bez větší ztráty přesnosti získat významné snížení výpočetní náročnosti. Především se jedná o kleštění modelů (parametr `-t`) a slov (parametr `-v`) s nízkou pravděpodobností výskytu. Uvedené parametry udávají prahovou hodnotu logaritmu pravděpodobnosti, pod kterou jsou použité mezivýsledky zanedbávány a odstraněny z výpočetního cyklu. To umožňuje omezit nadbytečné výpočty pro málo pravděpodobné stavy a slova. Vliv těchto parametrů na dobu a přesnost ASR systému ukazuje obr. 3.

5.2. Spuštění rozpoznávače spojitých promluv

S výše uvedenými modely a slovníkem je již možné spustit proces rozpoznávání spojitých promluv pomocí nástroje *HDecode*. Jak může nastavení parametrů dekodéru uvedených v sekci 5.1 ovlivnit přesnost a dobu rozpoznávání, ukazuje následující experiment.

Výpočet byl proveden na stroji Intel Core2 1,86 GHz, 2 GB RAM. Výsledky jsou porovnány na úrovni přesnosti rozpoznávání slov $Acc = (N - D - S - I) / N \cdot 100\%$, kde N , D , S a I odpovídají počtu všech rozpoznávaných, smažaných, zaměněných a chybně vložených slov.

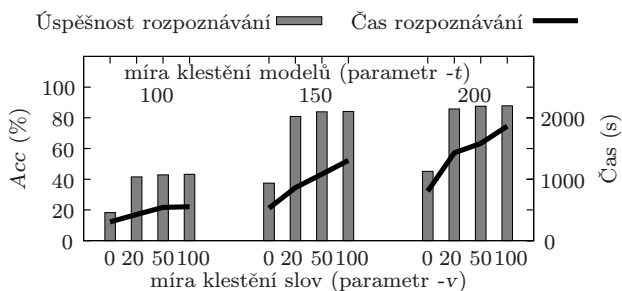
S uvedenými parametry dosahuje rozpoznávač vysokých dob výpočtu (real-time faktor okolo 1–3). K realizaci rozpoznávače spojitých promluv v reálném čase by bylo ovšem zapotřebí tento čas co nejvíce snížit. Toho lze docílit uvedenými návrhy na úpravu modelů (např. různé počty mixtures, změny v jazykovém modelu, ...) a také parametry dekodéru zmíněnými v sekci 5.1.

6. Závěr

Článek popisuje kroky vedoucí k tvorbě rozpoznávače spojitých promluv v češtině s využitím nástrojů HTK. Postup využívá standardních postupů v souladu s manuálem *HTKBook*, které jsou doplněny s ohledem na koncové využití systému pro rozpoznávání s velkým slovníkem.

Rozpoznávač vzniklý uvedeným postupem je základní konstrukcí pro experimenty v oblasti rozpoznávání řeči. Vzhledem k možnosti dalších úprav jednotlivých částí systému oproti komerčně realizovaným celkům zaměřeným na nízkou výpočetní náročnost a vysokou efektivitu je tento systém vhodný pro uplatnění především ve výzkumu.

Publikované experimenty srovnávají vliv základních parametrů jazykového modelu a nastavení dekodéru. Na



Obrázek 3: Vliv parametrů dekodéru na přesnost a dobu rozpoznávání

úloze rozpoznávání spojitě řeči se středně velkým slovníkem bylo dosaženo úspěšnosti až 87,8 % a real-time faktor 2, resp. 83,9 % a RT faktor 1,2.

V rámci práce vznikla také sada skriptů, které realizují jednotlivé kroky tvorby rozpoznávače a které umožňují rychle vytvořit jednoduchý rozpoznávač spojitých promluv na bázi HTK. Tyto skripty jsou volně k dispozici na stránkách laboratoře [18].

Poděkování

Výzkum byl podporován granty GAČR 102/08/0707 „Rozpoznávání mluvené řeči v reálných podmínkách“, GAČR 102/08/H008 „Analýza a modelování biomedicinských a řečových signálů“ a výzkumným záměrem MSM 6840770014 „Výzkum perspektivních informačních a komunikačních technologií“.

Reference

- [1] Gales, M. J. F. – et al.: Progress in the CU-HTK broadcast news transcription system, *IEEE Transactions on Audio, Speech and Language Processing*, vol. 14, nr. 5, 2006, pp. 1513–1525.
- [2] Abad, A., Meinedo, H., Neto, J.: Automatic Classification and Transcription of Telephone Speech in Radio Broadcast Data. *Proceedings of the 8th international conference on Computational Processing of the Portuguese Language (PROPOR '08)*, Aveiro, Portugal, 2008, pp. 172–181.
- [3] Nouza, J., Žďánský, J., David, P., Červa, P., Kolorenc, J., Nejedlová, D.: Fully Automated System for Czech Spoken Broadcast Transcription with Very Large (300K+) Lexicon. *Proc. of International Conference on Speech and Language Processing (Interspeech 2005)*, Lisboa, Portugal, Sept. 2005.
- [4] Nouza, J., Silovský, J., Žďánský, J., Červa, P., Kroul, M., Chaloupka, J.: Czech-to-Slovak Adapted Broadcast News Transcription System. *Proceedings of the 9th Annual Conference of the International Speech Communication Association, (Interspeech'08)*, pp. 2683–2686, Brisbane, Australia, Sept. 2008.
- [5] Stránka projektů – Laboratoř počítačového zpracování řeči. Ústav informačních technologií a elektromiky, FM, TUL. Liberec. <https://www.ite.tul.cz/speechlab/index.php/vyzkum/projekty.html>.
- [6] Lamere, P., Kwok, P., Gouvea, E., Raj, B., Singh, R., Walker, W., Warmuth, M., Wolf, P.: The CMU SPHINX-4 speech recognition system. *Proceedings of the IEEE Intl. Conf. on Acoustics, Speech and Signal Processing*. Hong Kong 2003.
- [7] Lee, A., Kawahara, T., Shikano, K.: “Julius – an open source real-time large vocabulary recognition engine.” *Proc. European Conference on Speech Communication and Technology (EUROSPEECH'01)*, pp. 1691–1694, 2001.
- [8] Sundaram, R., Ganapathiraju, A., Hamaker, J., Picone, J.: Isip 2000 Conversational Speech Evaluation System. In: NIST Evaluation of Conversational Speech Recognition over the Telephone, 2000.
- [9] Young, S., Evermann, G., Gales, M., Hain, T., Kershaw, D., Liu, X.-Y., Moore, G., Odell, J., Ollason, D., Povey, D., Valtchev, V., Woodland, P.: The hidden Markov model toolkit (HTK) version 3.4.1, 2006. <http://htk.eng.cam.ac.uk/>.
- [10] Young, S. – et al.: *The HTK Book (for HTK Version 3.4)*. Cambridge University Engineering Department, 2009.
- [11] Novotný, J.: *Trénování a využití kontextově vázaných HMM modelů fonémů*. Výzkumná zpráva, ČVUT, Praha, srpen 2002.
- [12] Rotovnik, T., Maučec, M. S., and Kačič, Z. Large vocabulary continuous speech recognition of an inflected language using stems and endings. *Speech Communication*. 49, 6 (June 2007), 437–452.
- [13] Nouza, J., Nejedlová, D., Žďánský, J., Kolorenc, J.: Very Large Vocabulary Speech Recognition System for Automatic Transcription of Czech Broadcast. *Proceedings of ICSLP 2004*, Jeju Island, Korea, October 2004, pp. 409–412, 2004.
- [14] Nejedlová, D.: *Creation of lexicons and language models for automatic broadcast news transcription*. Dizertační práce, Technical University in Liberec, Liberec, 2006.
- [15] McTait, K., Adda-Decker, M.: The 300k LIMSI German Broadcast News Transcription System. *Proc. of European Conference on Speech Communication and Technology, (Eurospeech 2003)*, Geneva, Sept. 2003, pp. 213–216.
- [16] Ircing, P.: *Large vocabulary continuous speech recognition of highly inflectional language (Czech)*. Dizertační práce, University of West Bohemia, Faculty of Applied Sciences, Pilsen, Czech Republic, 2003.
- [17] Drábková, J.: *Tvorba jazykového modelu založeného na třídách*. Dizertační práce, Technical University in Liberec, Liberec, 2005.
- [18] Stránky Laboratoře zpracování řeči SpeechLab. <http://noel.feld.cvut.cz>

Prínos Alana D. Blumleina v oblasti stereofonického zvukového záznamu

Pavol Brezina

UKF PF KH, Tr. A. Hlinku 1, 949 01 Nitra, Slovensko
e-mail: pbrezina@ukf.sk

This paper deals with basic principles of stereo with an emphasis placed on two channel stereo system. It focuses on coincident microphone technique invented by Alan Dower Blumlein in 1931 and describes how it works. Using of Blumlein stereo pair is demonstrated on few practical examples of recording different musical instruments. Some examples of recording techniques used by renowned sound engineer Bruce Swedien are also described.

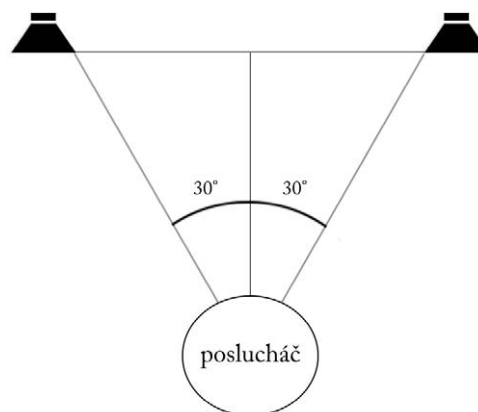
1. Alan Dower Blumlein

Alan Dower Blumlein (29. jún 1903 – 7. jún 1942) bol anglický elektrotechnik známy vďaka mnohým vynálezom a patentom v oblastiach telekomunikácií, zvukového záznamu, stereofónie, televíznej a radarovej techniky. V roku 1923 mu bol po štúdiu na City and Guilds College udeľený titul BSc (bakalár vied). Po štúdiu začal pracovať vo firme International Western Electric Corporation a v roku 1924 prešiel do firmy Standard Telephones and Cables, kde sa venoval výskumu v oblasti diaľkovej telekomunikácie v rámci Európy. Nadanie a novátorstvo sa prejavili u Blumleina už v mladom veku a v 23 rokoch mu bola udeľená cena za prínos. V roku 1929 začal Blumlein pracovať pre jednu z najstarších gramofónových firiem v Británii – Columbia Graphophone Company (neskôr EMI). Jeho prvou úlohou bolo vytvoriť prístroj slúžiaci na výrobu gramofónových platní tak, aby obišiel vtedy známy patent Bellových laboratórií. V roku 1930 navrhol spolu s Herbertom Holmanom niekoľko elektrodynamických mikrofónov, ktoré používali na nahrávanie v EMI a BBC. Okolo roku 1930 začína vzrastať Blumleinov záujem o stereofónny záznam zvuku a jeho reprodukciu. Ten neskôr priniesol výsledky v podobe prevratných patentov v oblasti stereofónie a výroby stereofónnych zvukových nosičov. V priebehu roku 1931 vytvoril Blumlein základnú ideu stereofónneho zvuku a nazval ju „binaurálny zvuk“. Myšlienka sa zrodila začiatkom roku 1931 pri návšteve kina. Blumlein si všimol, že zvuk reči hercov vychádzajúci z reproduktorov nezodpovedá ich aktuálnemu umiestneniu na filmovom plátne. Rozhodol sa preto, že vytvorí systém prenosu zvuku, ktorý bude „sprevádzať“ hercov pohyb na filmovom plátne. V princípe išlo o to, že ak sa herec nachádza napríklad na pravej strane filmového plátna, zvuk prichádza prevažne z pravého reproduktora. Tým sa dosiahne vyššia vierohodnosť lokalizácie filmového priestoru. Túto prevratnú myšlienku predstavil Blumlein svojmu poradcovi v EMI Isaacovi Shoenbergovi ešte v lete roku 1931. Prvé poznámky k danej téme začal Blumlein spisovať 25. septembra 1931 a patent niesol názov *Improvements in and relating to Sound-transmission, Sound-recording and Sound-reproducing Systems* (Vylepšenia v oblasti sys-

témov zvukového prenosu, zvukového záznamu a zvukovej reprodukcie). Patent bol akceptovaný 14. júna 1933 pod číslom 394325 [1]. Približne v tom istom čase prebiehal podobný výskum stereofónie v Bellových laboratóriách pod vedením Harveyho Fletchera. Fletcher sa zameriaval na využívanie trojkanálového systému, zatiaľ čo Blumlein pracoval s dvojkanálovým. Otázka počtu kanálov sa neskôr ukázala ako kľúčová pre ďalší vývoj stereofónie. Skôr, ako prejdeme k podrobnejšiemu popisu mikrofónovej stereotechniky vynájdenej Blumleinom, je potrebné ozrejmiť základnú myšlienku stereofónneho snímania zvuku.

2. Základná charakteristika stereofónie

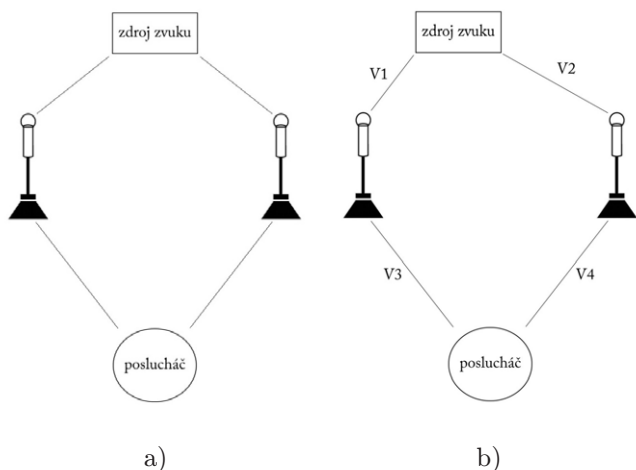
Štandardná schéma stereofónneho systému pozostáva z dvoch mikrofónov a dvoch reproduktorov, pričom každý z mikrofónov je zapojený do jedného reproduktora. Naskytá sa ale otázka, ako je možné, že poslucháč vníma zvuk priestorovo, keď má k dispozícii iba dva zdroje zvuku – ľavý a pravý reproduktor? Odpoveď musíme hľadať v spôsobe, akým vníma reprodukováný zvuk ľudský mozog. Ak je poslucháč umiestnený v štandardnej pozícii ako na obrázku 1, tak stred jeho hlavy delí oba reproduktory imaginárnou čiarou. Tá sa nachádza presne v strede 60° optimál-



Obrázok 1: Umiestnenie poslucháča v ideálnom reprodukčnom stereosystéme

neho uhla poslucháča k oboj reproduktorom. Štandardná pozícia teda vytvára rovnostranný trojuholník, pričom poslucháč a oba reproduktory tvoria jeho rohy.

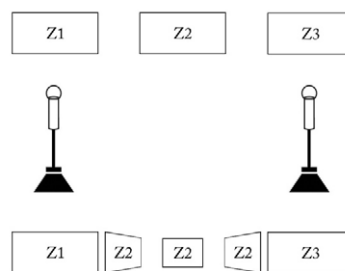
Pokiaľ pustíme do oboch reproduktorov identický signál (obrázok 2a), sila vyžarovaného zvuku bude rovnaká pre každý z nich. Keďže je poslucháč vzdialený rovnako od ľavého aj pravého reproduktora a intenzita zvuku oboch reproduktorov je rovnaká, nenastáva žiadne oneskorenie zvuku na ceste k poslucháčovi.



Obrázok 2: (a) Pozícia zaznamenaného zvuku je v stereosystéme identická s pozíciou reprodukováného zvuku; (b) Záznam pozície zvukového zdroja sa prenáša k poslucháčovi po dvoch cestách – V1, V3 a V2, V4

V takomto prípade vznikne ilúzia, že zvuk prichádza z bodu v presnom strede medzi oboma reproduktormi. Tento efekt sa nazýva virtuálny stred (phantom center). Obrázok 2b nám znázorňuje situáciu, keď je nahrávaný zdroj zvuku bližšie k jednému z mikrofónov, a teda nastáva zmena v intenzite a časové oneskorenie zvuku. Keďže je zdroj zvuku bližšie k ľavému mikrofónu, intenzita v pravom mikrofóne bude nižšia a zvuk sa k nemu dostane neskôr. K poslucháčovi sa potom dostáva zvuk po dvoch odlišných cestách – V1 a V3 pre ľavé ucho a V2 a V4 pre pravé ucho (kvôli zjednodušeniu príkladu zámerne vynechávame skutočnosť, že reálne prichádza zvuk z ľavého reproduktora aj do pravého ucha a z pravého reproduktora do ľavého ucha, čo je jednou z najväčších komplikácií sterea). Pokiaľ je vzdialenosť V3 a V4 rovnaká, vzdialenosť medzi zdrojom zvuku a mikrofónmi zostane pre poslucháča zachovaná. Pokiaľ si poslucháč zachová štandardnú pozíciu, bude počuť zdroj zvuku výraznejšie v ľavom uchu. Pokiaľ sa ale vychýli zo štandardnej pozície a vzdialenosť V3 a V4 nebude rovnaká, zvuk dorazí k poslucháčovi v rozdielnych časoch. Tým vznikne deformácia pôvodného stereofónneho obrazu, takzvanej stereo panorámy.

K narušeniu stereo panorámy môže prísť aj nesprávnym rozložením mikrofónov. Pokiaľ sú mikrofóny od seba vzdialené príliš a majú snímať tri zvukové zdroje, ako je

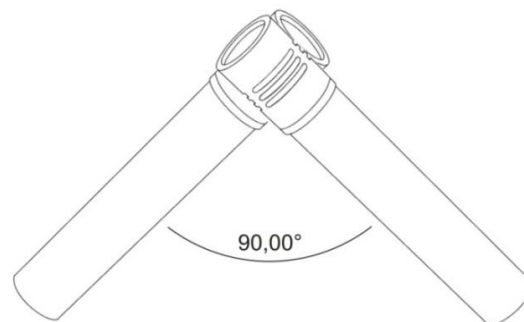


Obrázok 3: Efekt prázdneho stredy

zobrazené na obrázku 3, zvukové zdroje umiestnené bližšie k mikrofónom budú dominantné, zatiaľ čo vzdialenejšie budú výrazne tichšie zaznamenané. Zdroj zvuku umiestnený v strede mikrofónov je v porovnaní s ostatnými zdrojmi ďalej od mikrofónov, preto bude jeho zvukový signál zosnímaný omnoho tichšie. Pri reprodukcii takto nasnímaného priestoru vznikne efekt prázdneho stredy (hole in the middle effect). Zvukový zdroj, ktorý by mal znieť v strede, bude kombinovaný s ostatnými, výrazne dominantne zaznamenanými zdrojmi, a z tohto dôvodu bude jeho zvuk neprirodzený a zle počuteľný. Efekt prázdneho stredy sa snažili vyriešiť v roku 1930 Bellove laboratória tým, že pridali tretí, centrálny prenosový kanál. Namiesto dvoch reproduktorov musel mať poslucháč k dispozícii tri, čo sa ujalo iba začiatkom 50. rokov. Centrálny reproduktor sa začal vo väčšej miere využívať v kinách a v súčasnosti sa vracia späť k poslucháčom v podobe viackanálových domácich kín.

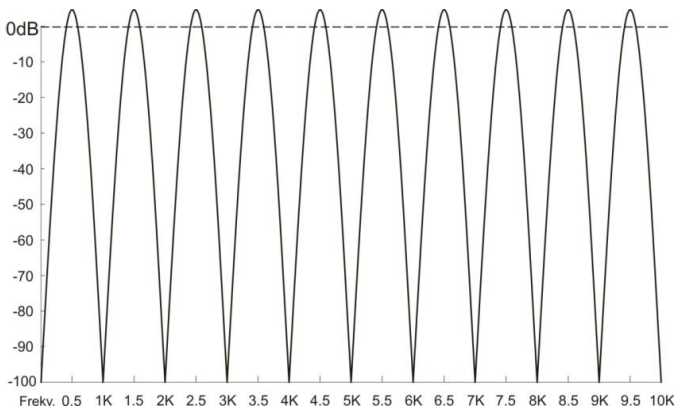
3. Koincidenčná mikrofónová technika

Alan Blumlein vyriešil problém efektu prázdneho stredy vynájdením prvej mikrofónovej stereotechniky, pričom ako sme už skôr uviedli, pracoval iba s dvoma kanálmi a mikrofónmi. Základom Blumleinovej stereotechniky je koincidenčná mikrofónová technika (obrázok 4).



Obrázok 4: Koincidenčné rozloženie mikrofónov (technika XY)

Tá vznikne umiestnením dvoch mikrofónov k sebe najbližším možným spôsobom, čo zminimalizuje rozdiel časov, za ktoré zvuk príde k oboch mikrofónom. Tým sa eliminuje problém vzniku fázového posunu (obrázok 4). Mikrofóny sú pritom umiestnené tak, že membrána jedného mikrofónu je nad membránou druhého, čiže si v horizontálnej rovine nebránia a zvierajú 90° uhol. Takéto rozloženie mikrofónov nazývame koincidenčný pár. Každý z mikrofónov sníma jednu stranu stereofónneho obrazu a vzhľadom na to, že intenzita prichádzajúceho zvuku je jediná hodnota, ktorá je pre každý z mikrofónov iná, koincidenčný pár sa zvykne nazývať aj intenzitné stereo (intensity stereo). Pri využívaní koincidenčných, ale aj iných stereo techník je nutné dbať na kompatibilitu s mono systémami. Využívanie stereo reprodukčných systémov je síce dnes už bežné, no napriek tomu je mono kompatibilita nutná napríklad pre vysielanie v rádiu či televízii. Ešte aj v súčasnosti existuje mnoho poslucháčov, ktorí používajú mono distribučné systémy. Využívanie koincidenčných mikrofónových techník má oproti iným technikám ešte jednu dôležitú výhodu. Tým, že sú mikrofóny od seba vzdialené iba minimálne, eliminuje sa takmer úplne takzvaný hrebeňový efekt (comb-filtering). Ten vzniká kombináciou priameho a oneskoreného zvuku prichádzajúceho z rovnakého smeru k jednému alebo viacerým mikrofónom, ktoré sú od seba vzdialené zhruba 8,5 cm za sebou v osi prichádzajúceho zvuku. Niektoré frekvencie sa pritom zmiešavajú vo fáze, zatiaľ čo iné sa zmiešavajú v protifáze (obrázok 5).

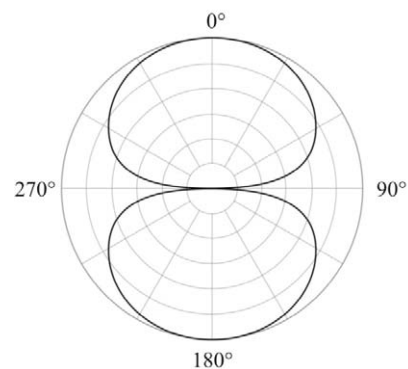


Obrázok 5: Hrebeňový efekt. Krivky nad hranicou 0 dB znázorňujú skreslenie signálu

Pri počúvaní na stereo systéme je hrebeňový efekt takmer nepočuteľný, pretože zvuk prichádza z dvoch reproduktorov (v niektorých situáciách je hrebeňový efekt vnímateľný ako priestorová atmosféra, no v skutočnosti ide o skreslenie intenzity a doby, za ktorú zvuk príde k mikrofónu). Pokiaľ ale prehráme daný zvukový materiál na mono systéme, zvuk prichádza iba z jedného reproduktora a hrebeňový efekt sa stáva výrazne počuteľným a nepríjemne skresleným. Umiestnenie mikrofónov pri nahrávaní je preto dôležité. Chybné vytvorený zvukový záznam totiž nie je možné v neskoršej fáze spracovania zvuku uspokojivo opraviť.

4. Blumleinov stereopár

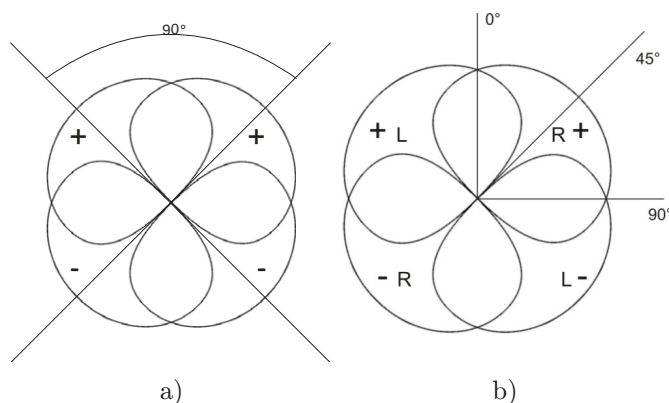
Ako sme už skôr spomínali, Blumlein používal pri nahrávaní aj reprodukcii dva kanály a jeho systém záznamu zvuku pozostáva z dvoch mikrofónov. Koincidenčné rozloženie mikrofónov zostáva zachované, ale každý z mikrofónov má osmičkovú charakteristiku snímania zvuku, známu aj pod názvami bidirectional alebo figure-of-eight. Skôr, ako prejdeme k popisu Blumleinovho stereopáru, je potrebné priblížiť osmičkovú charakteristiku mikrofónu. Ide o najcitlivejšiu mikrofónovú charakteristiku, pretože reaguje na zmenu akustického tlaku na oboch stranách membrány (obrázok 6). Pokiaľ zvuk prichádza v osi membrány, či už spredu (pozícia 0°) alebo odzadu (pozícia 180°) membrány, bude snímaný v rovnakej sile. Čím viac sa zvuk



Obrázok 6: Osmičková charakteristika

bude blížil k pozíciám 270° a 90° , tým viac bude slabnúť sila jeho zosnímania. Teoreticky, ak by zvuk prichádzal iba z osi 270° alebo 90° , mikrofón by zvuk nezaznamenal, pretože by na membránu nebol vyvinutý žiaden akustický tlak. Ďalšou dôležitou vlastnosťou osmičkovej charakteristiky je absolútna polarita, čo znamená, že predná strana membrány (pozícia 0°) vytvára kladný elektrický signál a zadná strana (pozícia 180°) záporný. Ako neskôr zistíme, absolútna polarita zohráva dôležitú úlohu v Blumleinovom stereopáre. Alan Blumlein prišiel s nápadom umiestniť dva mikrofóny osmičkovej charakteristiky k sebe tak, aby sa ich osi pretínali v mieste minimálnej citlivosti snímania (obrázok 7a).

Osi mikrofónov teda zvierajú 90° uhol a prednú časť stereopáru tvoria kladné polarities oboch mikrofónov, zatiaľ čo zadnú tvoria záporné polarities. 90° natočenie mikrofónov zároveň zaručuje, že celá oblasť okolo mikrofónov bude rovnomerne zosnímaná. Stereopár musí byť voči jednému statickému zvukovému zdroju otočený stredom, teda pozíciou 0° a nikdy nie pozíciou 45° , ako býva často chybné uvádzané (obrázok 7b). Zvuk prichádzajúci z pozície 0° je distribuovaný rovnomerne do ľavého aj pravého kanálu a pri reprodukovani ho poslucháč vníma ako stred panoramatického priestoru medzi oboma reproduktormi. Ak by sa zvukový zdroj začal pohybovať smerom k hranici 45° , intenzita zvuku v pravom kanáli by sa začala stávať dominantnou oproti intenzite v ľavom kanáli. Ak by zvukový



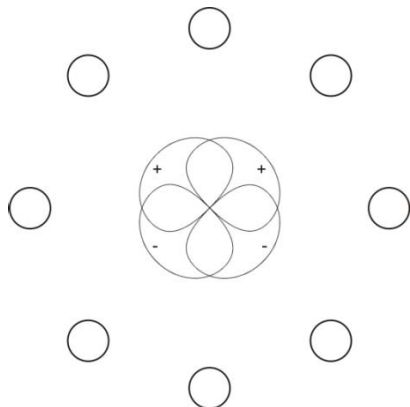
Obrázok 7: (a) Blumleinov stereopár; (b) Prednú stranu Blumleinovho stereopáru určuje pozícia 0°

zdroj dosiahol úroveň 45° , pravý kanál bude disponovať maximálnou intenzitou zvuku, zatiaľ čo intenzita v ľavom kanáli bude na minimálnej až nulovej úrovni. Keďže pomer nasnímanej intenzity zvuku sa rovnakým spôsobom prenáša aj do reproduktorov, poslucháč je schopný určiť, z ktorej strany zvuk k mikrofónom prichádzal. Dospiaľ sme sa venovali prednej strane mikrofónov s kladnou polaritou, ale zvuk v priestore sa dostáva k mikrofónom aj zo zadnej strany a tu sa práve potvrdzuje význam absolútnej polaritu. Keď polovica Blumleinovho stereopáru s kladnou polaritou zachytáva priamy zvukový signál, logicky potom zadná časť mikrofónov zachytáva odrazy pôvodného zvuku od okolitého prostredia. Faktor odrazov zvuku je kľúčovým pri vytváraní komplexného stereoobrazu, pretože určuje jeho hĺbku a charakter. Práve na základe odrazov vie poslucháč identifikovať priestor, v ktorom sa zvukový zdroj nachádzal, a bez ich prítomnosti by bol zaznamenaný zvuk plochý a deformovaný. Dôležitú rolu pritom zohráva pomer priameho a odrazeného zvuku. Koincidenčná charakteristika Blumleinovho stereopáru umožňuje experimentovať s polohou mikrofónov tak, aby bola úroveň priameho signálu vyrovnaná s úrovňou dozvukového signálu. Hľadanie polohy mikrofónov si často vyžaduje množstvo času, ako aj akusticky vhodný priestor na nahrávanie. Odrazy zvuku sa totiž dostávajú k mikrofónom nielen z prednej a zadnej strany, ale aj z bočných strán, kde je zvuk snímaný oblasťou kladnej polaritu jedného mikrofónu a zároveň oblasťou zápornej polaritu druhého mikrofónu. Môže tak vzniknúť fázový posun v prípade, že mikrofóny nie sú umiestnené vo vhodnej pozícii voči zvukovému zdroju, alebo sa nachádzajú v akusticky neprimeraných podmienkach. Okrem opačnej polaritu má zadná strana Blumleinovho stereopáru aj opačné rozloženie ľavého a pravého kanála (obrázok 7a). Pri ambientnom zvuku je ale opačná pozícia kanálov takmer zanedbateľná a nepočuteľná, zvlášť keď sa zvuk zmiešava s intenzitne dominantnou prednou stranou stereopáru. Rozdiel by sme počuli iba v prípade, že by bol pred zadnou stranou mikrofónov umiestnený iný zvukový zdroj v približne rovnakej vzdialenosti ako zdroj z prednej strany.

5. Využívanie Blumleinovho stereopáru v praxi

Alan Blumlein uskutočnil v roku 1934 niekoľko testovacích nahrávok pre nahrávaciu spoločnosť EMI, ktoré v roku 2008 digitálne remastroval Roger Beardsley. Pre verejnosť sa tak stali prístupné po vyše sedemdesiatich rokoch a predstavujú unikátnu zbierku prvých zvukových stereonahrávok. V súvislosti so zložitým procesom digitálneho prepisu Beardsley vyhlásil, že reprodukcia síce neprebela ideálne, ale boli zachytené všetky pôvodné informácie obsiahnuté v origináli [2]. Medzi testovacími nahrávkami sa nachádzajú aj videozáznamy, v ktorých Blumlein spolu so svojimi kolegami prezentuje pohyb zvuku v závislosti na pohybe osôb v miestnosti. Jedným z najznámejších videozáznamov je prechádzanie osôb z ľavej strany obrazovky na pravú, pričom nahlas počítajú. Zvukový záznam presne kopíruje pohyb osôb, teda postupne sa presúva z ľavého reproduktora do pravého. Zo zvukových nahrávok je najznámejšia Mozartova symfónia č. 41 C dur v interpretácii Londýnskeho filharmonického orchestru pod vedením Sira Thomasa Beechama. Nahrávka bola vytvorená v nahrávacích štúdiách Abbey Road pomocou Blumleinovej inovatívnej stereotechniky. Ešte v tom istom roku Blumlein prezentoval záznam symfónie vedeniu nahrávacej spoločnosti EMI, ktorá rozhodla o pozastavení výskumu v tejto oblasti. Dôvodom bolo vyhlásenie EMI o slabom komerčnom potenciáli patentu. Hneď na to bol Blumlein poverený výskumom v oblasti rozvoja televízie a následne radarovej techniky. Blumlein navrhol svoju jedinečnú mikrofónovú stereotechniku tak, aby zosnímala akustiku priestoru vierohodne. Jej využívanie si preto vyžaduje aj excelentné akustické podmienky, ktoré sú dostupné iba niektorým zvukovým inžinierom. Oveľa častejšie sa využívajú akusticky menej vhodné priestory (menšie nahrávacie miestnosti, nevyhovujúci pomer odrazivosti a pohltivosti akustických riešení), a preto siahajú zvukoví inžinieri po mikrofónových technikách s menšou náročnosťou na akustiku priestoru. O to intenzívnejšie sa ale v posledných rokoch o Blumleinových patentoch diskutuje a jeho stereotechnika si získava čoraz viac pozornosti v profesionálnej i amatérskej sfére. Jedným z najznámejších propagátorov Blumleinovho odkazu je legendárny zvukový inžinier Bruce Swedien. Je jedným z tých, čo vedia vyťažiť maximálny potenciál z Blumleinovej stereotechniky a aj vďaka nemu zažíva táto jedinečná mikrofónová technika svoju renesanciu. Vo svojej knihe *Make mine music* [3] uvádza príklady využívania Blumleinovho stereopáru na nahrávanie rôznych hudobných nástrojov. Jej najefektívnejšie využitie je pri snímaní väčších nástrojových zoskupení a ansámblov. Na často kladenú otázku, aký zvukový efekt použil pri nahrávaní speváckeho zboru pre pieseň Michaela Jacksona *Man in the mirror*, odpovedá – žiaden. Zbor bol nahrávaný použitím jedného Blumleinovho stereopáru bez dodatočnej aplikácie akýchkoľvek dozvukových efektov. Zaujímavý spôsob využitia pri nahrávaní speváckeho zboru je možné počuť v nahrávke skupiny Garneau

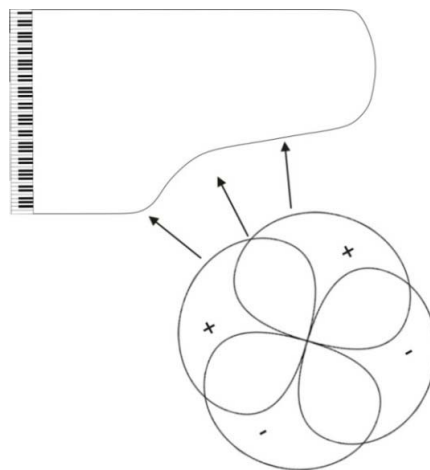
Keep the dream Alive, ktorú taktiež vytvoril Swedien. Pri nahrávaní zborových partov boli speváci rozmiestnení do kruhu okolo Blumleinovho stereopáru vo vzdialenosti približne jedného metra (obrázok 8).



Obrázok 8: Rozloženie spevákov okolo Blumleinovho stereopáru

Rovnomerne zaznamenaný zvuk poskytuje nevšedný zvukový zážitok a vytvára dojem, akoby speváci boli rozostavení okolo poslucháča. Zbor na nahrávke skupiny Garneau pozostáva iba z ôsmich spevákov, ale táto technika je dobre aplikovateľná aj na podstatne väčšie zborové zoskupenia. Okrem zborov sa Blumleinova stereotechnika osvedčila aj pri nahrávaní klavíra, akustickej a elektrickej gitary. Pri nahrávaní klavíra sa často používa mikrofónová technika AB v dobrej viere, že ide o verné zobrazenie stereoobrazu hudobného nástroja. Mikrofóny v pozícii AB sú k strunám klavíra príliš blízko a snímajú pomerne silnou intenzitou iba istú rezonančnú oblasť. Pri reprodukcii signálu v ľavom a pravom reproduktore vzniká efekt prázdneho streda. Z tohto dôvodu nejde pri technike AB o reprodukcii komplexného stereopriestoru, ale takmer dokonalého mona. Poslucháč má dojem akoby počúval dva nesúvisiace zvukové zdroje, každý z jedného reproduktora. V skutočnosti vníma poslucháč zvuk klavíra komplexnejšie. Ak sedí priamo pred klavirom, zvuk prichádza z jedného zvukového zdroja a zmiešava sa s odrazmi od okolia. Poslucháčovi je takéto vnímanie prirodzené, preto je dobré, ak sa zvuk klavíra dokáže zaznamenať tak, aby ho bolo možné vnímať prirodzene aj z nahrávky. Správna aplikácia Blumleinovho stereopáru zaznamenáva zvuk akustického klavíra vierohodnejším a prirodzenejším spôsobom ako ostatné mikrofónové techniky. Umiestnenie mikrofónov je znázornené na (obrázok 9).

Mikrofóny by mali byť v takej výške, aby sa nachádzali približne v strede medzi vekom klavíra a jeho rezonančnou doskou. Pole snímania mikrofónov tak obsiahne zvuk prichádzajúci z každej strany rezonančnej dosky a nie len z úzkeho priestoru pri kladivkách klavíra ako to bolo pri technike AB. Zároveň sú zaznamenané aj odrazy zvuku, čo dodáva nahrávke hĺbku a priestor. Snímanie pomocou Blumleinovho stereopáru zvyčajne v plnom rozsahu po-



Obrázok 9: Umiestnenie Blumleinovho stereopáru pri nahrávaní akustického klavíra

stačuje, no v špeciálnych situáciách sa môžu využiť aj prídavné mikrofóny. Obdobným spôsobom by sa postupovalo pri nahrávaní akustickej gitary a iných hudobných nástrojov.

6. Záver

Alan Blumlein sa zaslúžil svojimi objavmi a patentmi o vznik stereofónie tak, ako ju poznáme v dnešnej dobe. Jeho mikrofónová technika bola zároveň prvou, ktorá snímala priestorový zvuk do formátu stereo a zároveň slúžila ako východisko pre vznik ďalších mikrofónových techník. S nástupom viackanálových stereosystémov by sa mohlo zdať, že v klasickej stereofónii je už všetko prebádané a vyriešené. Nie je ale tomu tak. Oblasť najnovších skúmaní v oblasti stereotechniky sa sústreďuje na binaurálny zvukový záznam a reprodukcii, čo by mohlo posunúť úroveň kvality v oblasti stera na vyššiu zážitkovú úroveň pri počúvaní hudby. Ak by predstavitelia spoločnosti EMI vedeli predvídať masívny úspech stera a Alan Blumlein by pokračoval vo svojom výskume, dočkali by sme sa možno viacerých objavov a vylepšení. Faktom zostáva, že oblasť stereofónie je stále neuzavretá a mimoriadne zaujímavá pre ďalší výskum.

Reference

- [1] Blumlein, D., S.: *Improvements in and relating to Sound-transmission, Sound-recording and Sound-reproducing System*. Patent č. 394325 14. jún 1933, <http://www.doramusic.com/patents/394325.htm> [cit. 23.01.2010]
- [2] Shankleman, M.: *Early stereo recordings restored*. <http://news.bbc.co.uk/2/hi/technology/7537782.stm> [cit. 23.01.2010]
- [3] Swedien, B.: *Make mine music*, MIA Musikk, Norway, 2003

