Počítačové zpracování řeči

Přednáška 9

Programové prostředí HTK

Zkušenosti z předchozí úlohy

- Seznámili jste se s principy pro základní trénování HMM a rozpoznávání. Nejsou nijak složité. V Matlabu je lze je zrealizovat v kódech do 50 řádek.
- Úspěšnost vycházela přibližně stejná jako u DTW, výpočet ale mnohonásobně rychlejší (s 1 modelem se pracuje mnohem rychleji než s desítkami referencí).
- U rozsáhlejšího slovníku by rozdíl v úspěšnosti a rychlosti mezi HMM a DTW byl výrazně ve prospěch HMM.

Jak ještě dále zvýšit úspěšnost HMM?

- Trénováním na více datech
- Modelováním rozložení příznakových vektorů více gaussovkami (směsí)
- Trénováním parametrů modelů založeném na pravděpodobnostním přiřazení framů ke stavům (dosud jsme uvažovali pevnou vazbu každého framu pouze na 1 stav) – Baumův-Welchův algoritmus
- Výše uvedené postupy už vyžadují výrazně složitější kódy, jejichž implementaci ponecháme na odbornících – využijeme platformu HTK

HTK – Hidden Markov Model Toolkit

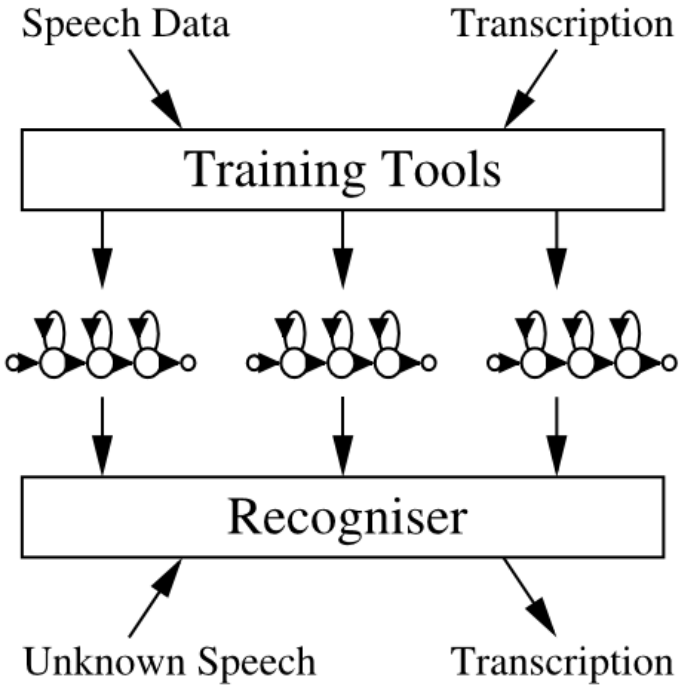
- Soubor programů a nástrojů pro práci s HMM. Výrazným způsobem usnadňuje vývoj systémů rozpoznávání řeči (zejména spojité)
- Vyvíjen od roku 1989 na University of Cambridge (UK)
- Po registraci na http://htk.eng.cam.ac.uk/register.shtml je možné stáhnout si zdrojové i přeložené programy a manuál k nim (HTKbook)
- Je používán na mnoha univerzitách v celém světě a je tudíž citován v mnoha tisících vědeckých článků
- Existuje rozsáhlá komunita uživatelů a při problémech lze rychle najít řešení na různých diskusních fórech
- Je použitelný v různých operačních systémech (Linux, Windows, …)
- Hodí se i na další úlohy zaměřené na sekvence (psané písmo, DNA, ..)
- Pozor – HTK nemá rád diakritiku (takže např. názvy fonémů nebo názvy modelů či souborů nesmí obsahovat jiné znaky než ASCII a ani mezery)

HTK – základní koncept

Dvě nejdůležitější úlohy při rozpoznávání řeči

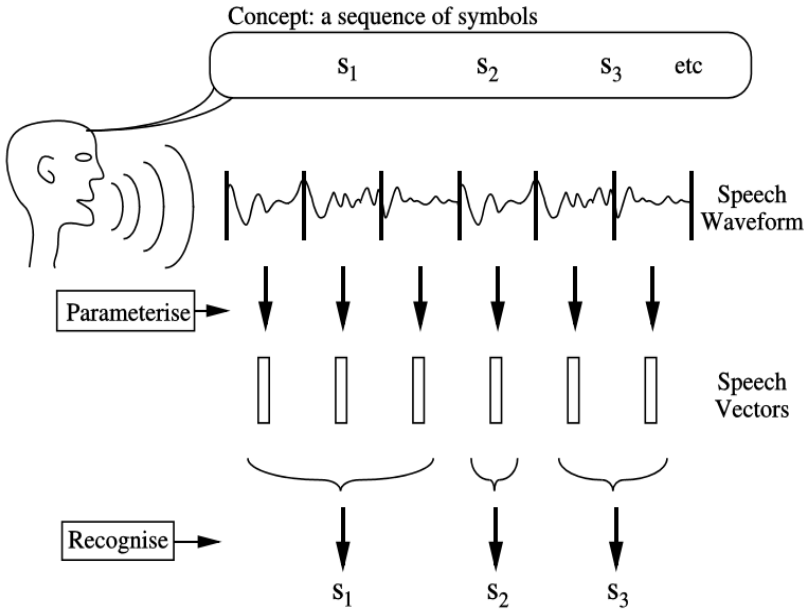
- Trénování modelů (akustických a u spojité řeči i jazykových)
- Testování – ověření a kvantitativní vyhodnocení rozpoznávání

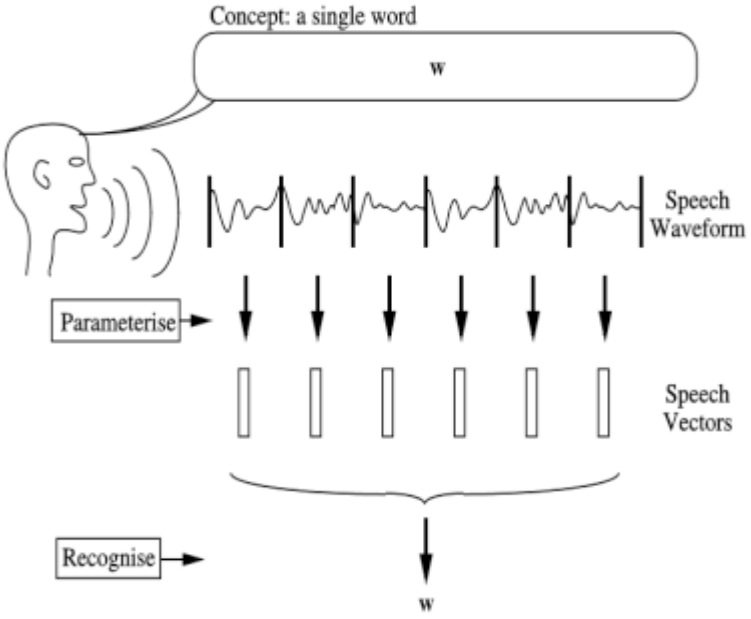
Pro obě jsou nutné nahrávky řeči a jejich přepisy



HTK – princip rozpoznávání řeči

Rozpoznávání izolovaných slov a spojité řeči (sekvence)



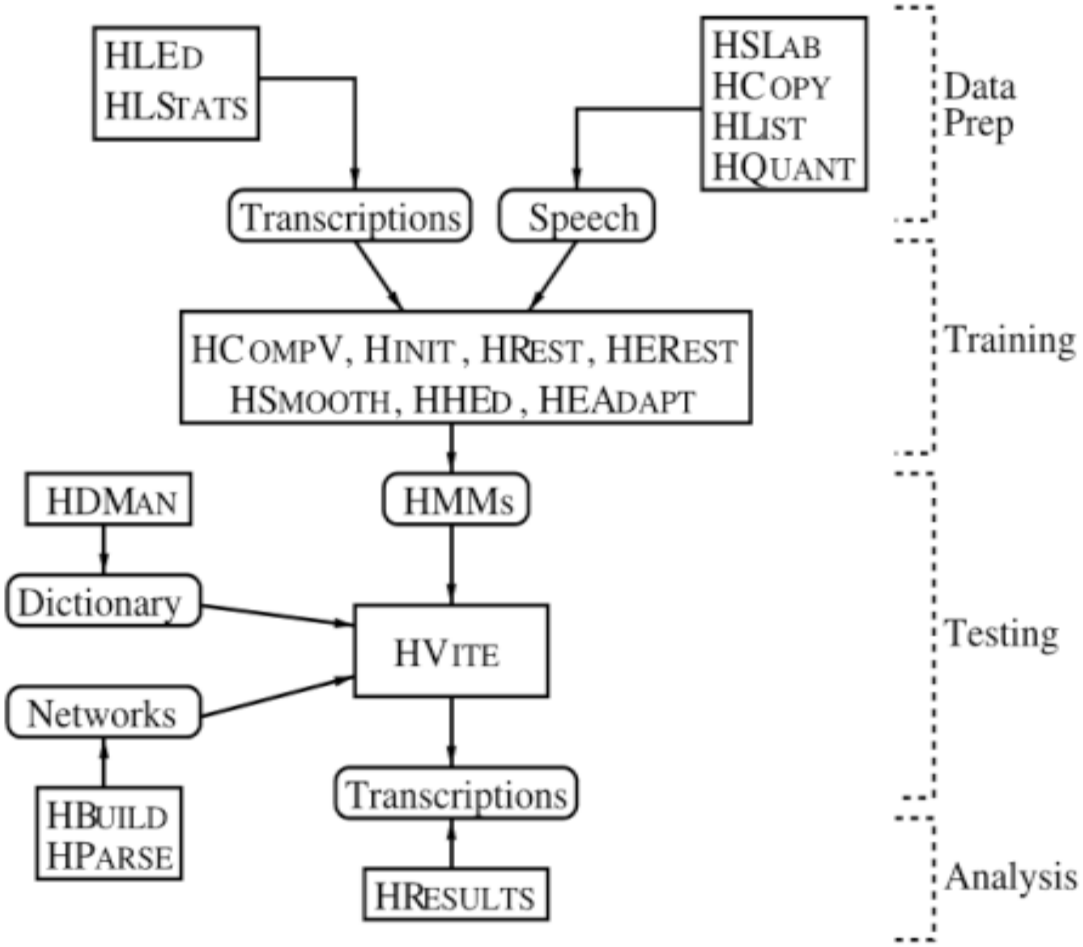


Prostředí HTK je zaměřeno na **rozpoznávání spojité řeči** (sekvence slov či hlásek)

Izolovaná slova jsou brána jako speciální případ (jako sekvence ticho – slovo – ticho)

HTK – nástroje

Prostředí nabízí několik desítek nástrojů (programů) pro různé dílčí úlohy, např. zpracování textů, příprava slovníku a gramatiky, parametrizace řeči, trénování modelů, rozpoznávání, vyhodnocování



HTK – klíčové nástroje

HCopy (HParam) – program pro **parametrizaci** signálu,
typ a počet příznaků se nastavuje pomocí konfiguračního souboru

HParse – program pro vytváření rozpoznávací **sítě slov** (specifikuje která slova mohou následovat po kterých, parsing)

HRest, HERest – programy pro **trénování** (reestimaci) HMM

HVite – program pro **rozpoznávání** (Viterbiho dekodér)

HResults – program pro **vyhodnocování** experimentů

HTK – slovník a slovní síť

Slovník definuje seznam slov a z jakých (dílčích) jednotek se skládají

Příklady: slovník pro rozpoznávání číslic vytvořený z celoslovních a hláskových jednotek

slovo	jednotka
NULA	nula
JEDNA	jedna
DVA	dva
….	
DEVET	devet
SENT-END	[] sil
SENT-START	[] sil

symbol pro ticho

slovo	jednotky
NULA	n u l a
JEDNA	j e d n a
DVA	d v a
…	
DEVET	d e v j e t
SENT-END	[] sil
SENT-START	[] sil

Příští semestr

Gramatika – symbolický popis povolených sekvencí slov

$digit = JEDNA | DVA | TRI | CTYRI | PET | SEST | SEDM | OSM | DEVET | NULA;

(SENT-START ($digit) SENT-END) promluva musí obsahovat právě 1 číslici

---------- alternativně----------

(SENT-START (<$digit>) SENT-END) promluva může obsahovat 1 nebo více číslic

Slovní síť – interní popis mezislovních přechodů

HParse grammar wordnet

HTK – nahrávky

Nahrávky můžeme realizovat v libovolném SW,
je nutné dodržet stejný formát (WAV, vzork. frekvenci, počet bitů, mono)

Ke každé nahrávce potřebujeme další **stejnojmenné** soubory

- textový soubor s příponou TXT,
  obsahující **slova**, která byla řečena, tedy např.
  JEDNA

- textový soubor s příponou LAB (není nutný u testovacích nahrávek)
  obsahující **jednotky**, které se v nahrávce vyskytují, tedy např.
  sil ticho (model pro ticho, angl. silence)
  jedna
  sil

HTK – parametrizace (1)

Z nahrávky typu WAV vytvoří soubor obsahující příznakové vektory v jednotlivých framech

Předdefinované typy:

FBANK - log. energie v P spektrálních pásmech

MFCC – kepstrální koeficienty

…..

Rozšíření o dynamické příznaky a příp. další příznaky:

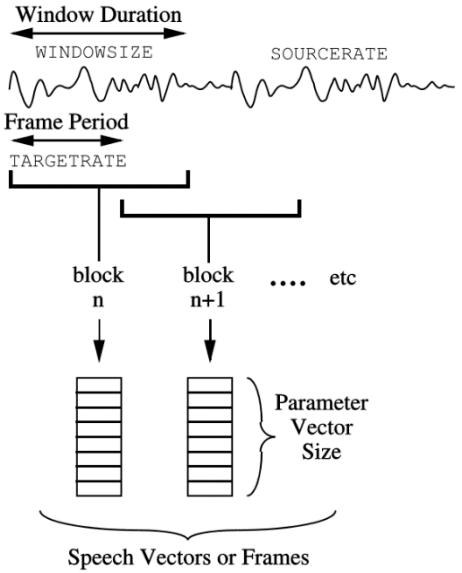
_D - 1. derivace

_A - 2. derivace

_0 log. energie v celém spektru

Konfigurační soubor pro typ FBANK:

```
SOURCEFORMAT = WAV
TARGETKIND = FBANK
TARGETRATE = 100000.0     perioda framu (x 100ns)
SAVECOMPRESSED = F
SAVEWITHCRC = F
WINDOWSIZE = 250000.0     délka framu (x 100ns)
USEHAMMING = T
PREEMCOEF = 0.97
NUMCHANS = 16
ENORMALISE = F
```



Provedení parametrizace

HCopy -T 1 -C ParamConfig-FBANK -S param.list

HTK – parametrizace (2)

Parametrizace vlastním programem do vlastních příznaků je možná

Musíme napsat vlastní program, který

- vytvoří **binární soubor**
- na jeho úvod napíše hlavičku – 4 čísla
- vypočítá a uloží příznakové vektory pro celou nahrávku

Hlavička:

počet framů - INT 4 byty
TARGETRATE - INT 4 byty
počet bajtů na 1 frame = počet příznaků x počet bytů na příznak - INT 2 byty
kód typu příznaků - USER (9) - INT 2 byty

DATA

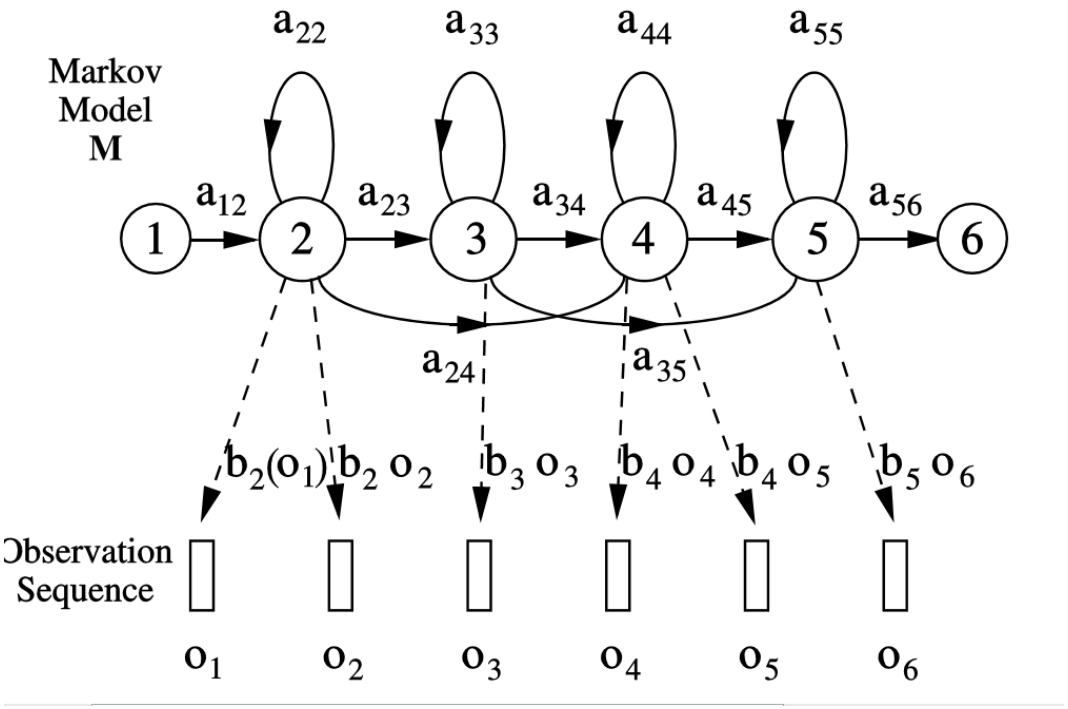
float čísla

HTK – struktura modelu (1)

HMM v HTK může mít různou strukturu, nejčastěji tzv. levo pravý model

Model má volitelný počet stavů, přičemž 1. a poslední jsou pouze pomocné
(terminologie – třístavový model v HTK má ve skutečnosti 5 stavů, včetně pomocných)

Typy přechodů mezi stavy: setrvání, přechod, přechod s přeskokem (málokdy se používá)



HTK – struktura modelu (2)

Struktura HMM se definuje pomocí tzv. prototypu

Příklad 3-stavový model

```
~o <VecSize> 16 <FBANK>          počet příznaků, typ příznaků
~h "proto"                       sem pak přijde později název modelu
<BeginHMM>
<NumStates> 5                    počet stavů, včetně pomocných (zde tedy 3 + 2)
<State> 2                        číslo stavu  (pomocné stavy se neuvádí)
<Mean> 16                        délka vektoru středních hodnot příznaků (tj. počet příznaků)
0.0 0.0 0.0 ... 0.0              šablona
<Variance> 16                    délka vektoru rozptylů příznaků (diagonála kovar. matice)
1.0 1.0 1.0 ... 1.0              šablona
<State> 3
<Mean> 16
0.0 0.0 0.0 ... 0.0
<Variance> 16
1.0 1.0 1.0 ... 1.0
<State> 4
<Mean> 16
0.0 0.0 0.0 ... 0.0
<Variance> 16
1.0 1.0 1.0 ... 1.0
<TransP> 5                       matice přechodů s vyznačením povolených přechodů
0.0 1.0 0.0 0.0 0.0              z 1. (fiktivního) stavu pouze do 2. stavu
0.0 0.6 0.4 0.0 0.0              v 2. mohu setrvat nebo přejít 3. stavu
0.0 0.0 0.6 0.4 0.0              v 3. mohu setrvat nebo přejít 4. stavu
0.0 0.0 0.0 0.7 0.3
0.0 0.0 0.0 0.0 0.0
<EndHMM>
```

Uvedený 3-stavový model se používá pro hlásky a pro ticho, pro slova je vhodný např. 8-stavový

HTK – trénování modelů

Trénování je iterativní proces. V rámci inicializace se nastaví počáteční hodnoty parametrů modelů, které se v dalších iteracích vylepšují.

Inicializaci lze provést pomocí programu **HCompV** a reestimace se dělají pomocí programu **HERest**.

Příklady volání:

HCompV -C TrainConfig-FBANK -f 0.01 -m -S train.scp -M hmm0 proto

HERest -C TrainConfig-FBANK -I source.mlf -t 250.0 150.0 1000.0 -S train.scp -H hmm0/hmmdefs -M hmm1 models0

Ve výše uvedených příkladech je

TrainConfig-FBANK … konfigurační soubor pro trénování

train.scp …. seznam zparametrizovaných trénovacích nahrávek

source.mlf …. soubor obsahující popis všech nahrávek odvozený ze souborů LAB

hmmdefs … soubor obsahující všechny natrénované modely v dané iteraci

HTK – rozpoznávání

Pro rozpoznávání se použije program **HVite**

Příklad volání:

HVite -H hmm6/hmmdefs -S test.scp -i recout.mlf -w wordnet -p -70.0 -s 0 dict models0

Ve výše uvedeném příkladu je

hmmdefs … soubor obsahující všechny natrénované modely v poslední 6. iteraci

test.scp …. seznam zparametrizovaných testovacích nahrávek

dict … slovník, wordnet …síť, models0 … seznam použitých modelů

Výstup je v souboru recout.mlf a vypadá následovně

"D:/HTK/DATA/0000_MVL/c0_p0000_s04.rec"
0 6700000 SENT-START -4.080025
6700000 11900000 NULA -877.901184
11900000 19800000 SENT-END 17.283134
.
"D:/HTK/DATA/0000_MVL/c1_p0000_s04.rec"
0 5900000 SENT-START -8.679775
5900000 13200000 JEDNA -1429.232910
13200000 19800000 SENT-END 22.320038

cas_od cas_do rozpoznáno_jako log_skore

HTK – vyhodnocování experimentů

Pro vyhodnocování se použije program **HResult**

Příklad volání:

HResults -e ??? SENT-START -e ??? SENT-END -t -I testref.mlf models0 recout.mlf

Ve výše uvedeném příkladu je

recout.mlf … výstup rozpoznávače, testref.mlf … soubor obsahující slova v každé nahrávce

Výstup vypadá následovně

```
Aligned transcription: D:/HTK/DATA/0000_MVL/c2_p0000_s04.lab vs D:/HTK/DATA/0000_MVL/c2_p0000_s04.rec
 LAB: DVA
 REC: NULA
Aligned transcription: D:/HTK/DATA/0000_MVL/c5_p0000_s04.lab vs D:/HTK/DATA/0000_MVL/c5_p0000_s04.rec
 LAB: PET
 REC: DEVET
….
====================== HTK Results Analysis =======================
  Date: Fri Mar 29 14:31:16 2019
  Ref : testref.mlf
  Rec : recout.mlf
------------------------ Overall Results --------------------------
SENT: %Correct=90.00 [H=45, S=5, N=50]
WORD: %Corr=90.00, Acc=90.00 [H=45, D=0, S=5, I=0, N=50]
===================================================================
```

HTK – rozpoznávání naživo

Program pro rozpoznávání **HVite** lze provozovat „naživo“
– tedy přímo mluvit do mikrofonu a sledovat výsledky rozpoznávání

Volání programu:

HVite -H hmm6/hmmdefs -C LiveConfig-FBANK16 -w wordnet -p -70.0 -s 0 dict models0

Je pouze nutné připravit konfig. soubor LiveConfig-FBANK16, a to tak že do TrainConfig-FBANK16 přidáme následující řádky:

```
# Waveform capture
SOURCERATE=625.0
SOURCEKIND=HAUDIO
SOURCEFORMAT=HTK
ENORMALISE=F
USESILDET=T
MEASURESIL=F
OUTSILWARN=T
```

Po spuštění programu jste nejprve požádáni, abyste řekli krátkou větu, během níž si rozpoznávač nastaví úroveň ticha a řeči. Pak už lze opakovaně říkat slova a sledovat zobrazený výsledek. Program se ukončí stiskem Ctrl+C

HTK – podrobný popis a ukázky

Podrobný popis celé platformy lze nalézt v manuálu HTKBook.

Detailní popis, jak připravit úlohu v této přednášce, je v souboru „HTK Trénování celoslovních modelů.doc“ na elearningu

Dále lze stáhnout soubor „HTK-example.zip“, kde jsou připraveny adresáře s již hotovými dávkami a soubory, které se dají spustit a následně modifikovat.

V adresáři Pokus1 jsou připraveny soubory pro parametrizaci typu FBANK se 16 statickými příznaky na frame. Celý postup je podrobně popsán ve výše uvedeném wordovském dokumentu.

V adresáři Pokus2 jsou připraveny soubory pro parametrizaci typu MFCC s 39 (statické+1.derivace+2.derivace) příznaky na frame.

Úloha pro cvičení

Rozpoznávač číslic založený na HTK

1) Vytvořte rozpoznávač číslic založený na celoslovních modelech
2) Stáhněte si soubor „HTK-example.zip“. Jsou v něm všechny potřebné programy a skripty, včetně řečových dat od 5 mluvčích.
3) Vytvořte si nový adresář Pokus3 a v něm podle návodu v dokumentu „HTK Trénování celoslovních modelů.doc“ realizujte celý proces trénování a testování. V případě nejasnosti se můžete inspirovat skripty v adresáři Pokus1, kde je vše hotovo.
4) Zjistěte a porovnejte úspěšnost (skóre) pro modely v adresářích hmm0 až hmm6 (výsledky iteračních kroků 0 až 6).
5) V adresáři Pokus 2 máte připraven proces trénování s příznaky MFCC39 (39 příznaků typu kepstrálních koef., typ MFCC_0_D_A). Projděte opět celý proces a zjistěte úspěšnost u těchto příznaků.
6) Vyzkoušejte si rozpoznávač naživo.
7) Změňte gramatiku úlohy tak, že nahrávka bude moci obsahovat 1 nebo více slov (číslic) a vyzkoušejte naživo

Úloha pro cvičení (2)

7) Nyní si vytvořte adresář Pokus 5 a proveďte podobné experimenty na datech, které máte k dispozici:
trénovací set: všechny nahrávky osob 30-49
testovací set: nahrávky ze sad 4 a 5 od osob 21xx a 22xx
8) Předchozí experiment je příklad SI testu, přesně stejného, jako jste používali v minulých úlohách. Zjistěte výsledky rozpoznávání pro FBANK a MFCC_0_D_A.
9) Napište mi – opět do pondělí - obdržené výsledky a porovnejte s nejlepšími, které jste dostali předtím s vlastními HMM a DTW.

Příště

Pouze krátké setkání spojené se zadáním závěrečných úloh.

Zkuste si do příště promyslet nějaká vhodná témata, např.:

- hlasové ovládání vlastní nebo existující aplikace, typu kalkulačka, malování, jednoduchá hra, webový prohlížeč, ….. (cca 20 povelů spojených s nějakými akcemi)
- Jednoduchý dialogový systém: počítač se ptá, uživatel odpovídá, např. informace o odjezdu, počasí, stavu konta, ….