Katedra aplikovanej informatiky

Fakulta matematiky, fyziky a informatiky

Univerzita Komenského v Bratislave

Systém pre rozpoznávanie slov a verifikáciu hovoriaceho

Dizertačná práca

RNDr. Vladimír Chudý

Školiteľ: Doc. Ing. Igor Farkaš, PhD.

Bratislava, 2012

Poďakovanie

Chcem sa poďakovať svojím priateľom Luciusovi Chudému, Klaudovi Viceníkovi za ich podporu a priateľstvo. V nie poslednej rade sa chcem poďakovať svojmu školiteľovi Doc. Ing. I. Farkašovi, PhD. za jeho podporu a cenné pripomienky.

Obsah

Zoznam použitých skratiek

ANN- Artificial Neural Network, umelé neurónové siete
BP - Back Propagation, siete so spätným šírením chyby
CD - Context dependent, kontextovo závislé fonémy
CI - Context independent, kontextovo nezávislé fonémy
CMP - Cochlear Microphone Potential, kochleárny mikrofónový potenciál
DCT - Direct Cosine Transformation, priama kosínová transformácia
DTW - Dynamic Time Warping, dynamické časové skreslenie
FFT - Fast Fourier Transformation, rýchla Fourier transformácia
FIR - Finite Impulse Response, konečná impulzná odozva
GLM - General Linear Method, obecná lineárna metóda
GMM - Gauss Mixed Model, gaussovské zmiešané modely
HMM - Hidden Markov Model, skryté markovove modely
ICA - Independent Component Analysis, analýza nezávislých zložiek
IIR - Infinite Impulse Response,nekonečná impulzná odozva
KNN - K - Nearest Neighbor, metóda zhlukovania k najbližších susedov
K-MEANS - metóda zhlukovania s dopredu zadaným počtom zhlukov k
LDA - Linear Discriminant Analysis, lineárna diskriminačná analýza
LMS - Least Mean Squares, metóda najmenších štvorcov
LPC - Linear Prediction Coefficients, koeficienty lineárnej predikcie
PDF - Probability Density Function, funkcia hustoty pravdepodobnosti
PLP - Perceptual Prediction Prediction, percepčná lineárna predikcia
MFC - Mel Frequency Cepstrum, Mel frekvenčné kepstrum
MLP - Multi - Layer Perceptron, viac vrstvový perceptrón
NP4 - Neural Phonetic Preprocessing Parrot, neuronálne fonetické predspracovanie podobné papagájovi
NP5 - Neural Phonetic Preprocessing Parrot PNN, ako predošlé plus PNN
PNN - Probanilistic Neural Network, pravdepodobnostná neurónová sieť
RBF - Radial Basis Functions, radiálne bázové funkcie
SRS - Subject Recognition System, systém rozpoznávania subjektu
SVM - Support Vector Machine, metóda podporných vektorov
TDNN - Typological Expert Neural Network, Typologické expertné neurónové siete
TIM - Topological Invariants Method, metóda topologických invariantov
VAD - Voice Activity Detection, detekcia rečovej aktivity
VRS - Voice Recognition System, systém rozpoznávania hlasu
WER - Word Error Rate, chybovosť prepočítaná na slovo

Zoznam obrázkov

Zoznam tabuliek

Abstract

The thesis presents topological invariants as speech features for automatic speech recognition and speaker verification. We applied method of topological invariants - TIM as eligible speech features for automatic speech recognition systems based on HMM. It provides an introduction to the mathematical concept of topological invariants, space and other symmetries focusing on their application to the speech recognition and speaker verification. This involves a basic overview of the relevant auditory characteristic and its modeling in order to identify possible symmetries and invariants. Once the whole concept is derived, few of its modifications vital for HMM systems like: reduction of features dimension, within class feature decorrelation, or rotation of the Nyquist plane are presented and evaluated on a real system using different scenarios.

The final system is trained according to the MASPER training procedure applied to both context dependent and context independent HMM models of different complexities. All the training and tests are performed on the professional speech database MOBILDAT-SK, where the achieved accuracy figures reached up to 97.7%, 98.7% and 98.9% for string of digits, application words, and isolated digits tests respectively.

The thesis also explores to some extent speaker verification in real-life conditions and provides a brief overview of the speaker recognition technology. We present and examine topological invariants (TIM), together with bad-pass filtering according to MEL-frequency scale, as standalone speech features in connection with probabilistic neural networks (PNN) that act as classifiers of speaker voices. We observed a performance, expressed by overall performance in %, is about 96 % for 1-word verification task and above 98 % for 3-word verification task, similar to MFC and PLP like features in tasks which have the similar setup (number of speakers and utterances type).

In our results we focused on the noise performance, dimensionality of feature vectors, and the problem of number of omitted speakers. Further, our data and results strongly support the idea of a general independence of speaker verification on speaker's native language and a family relationship of speakers. The implications of these results for further development of automatic speech recognition and speaker verification/identification are discussed from point of view of phoneme symmetries.

We discuss to some extent in conceptual manner a model of perception which involves these inner symmetries of phonological system of speech. Finally we discuss some interesting analogies between phonemic system and system of elementary particles.

Keywords: Speech recognition; speaker verification; topological invariants; speech features; Nyquist plane; MFC; PLP; HMM; LDA; MASPER; decorrelation transform; auditory system; TIM; MEL-frequency; PNN; speech perception; phoneme symmetries; phonological system; elementary particles.

Abstrakt

V dizertačnej práci prezentujeme metódu topologických invariantov ako rečových čŕt pre automatické rozpoznávanie reči a verifikáciu hovoriaceho. Aplikovali sme metódy TIM ako užitočné a spôsobilé rečové črty pre systémy automatického rozpoznávania reči na základe HMM. Poskytli sme úvod do matematického pojmu topologických invariantov, priestorových a iných symetrií, so zameraním na ich aplikáciu v konkrétnych systémoch rozpoznávania reči a verifikácie hovoriaceho. Zahŕňa to základný prehľad relevantných sluchových charakteristík a ich modelovanie tak, aby sme identifikovali možné symetrie a invarianty. Po definovaní a odvodení všetkých pojmov sme aplikovali niekoľko veľmi užitočných modifikácií, vhodných špeciálne pre HMM systémy ako: redukcia dimenzionality čŕt, v rámci tried čŕt dekorelácia, alebo rotácia Nyquist roviny. Vyhodnotili sme ich prínos na reálnom systéme použijúc rôzne scenáre.

Konečný systém sa učí na základe učiacej procedúry MASPER, ktorá sa aplikuje aj na kontextovo závislé aj na kontextovo nezávislé HMM modely rozličnej zložitosti. Celé učenie a testovanie sa prevádzalo na profesionálnej rečovej databáze MOBILDAT-SK, kde sme dosiahli úspešnosti až do 97.7%, 98.7% a 98.9% pre testy na reťazce číslic, aplikačné slová a izolované číslice, respektíve.

V dizertačnej práci skúmame tiež do určitého rozsahu verifikáciu hovoriaceho v reálnych podmienkach a poskytli sme stručný úvod do technológie rozpoznávania hovoriaceho. Prezentovali a otestovali sme topologické invarianty (TIM), spolu s filtrovaním vo frekvenčných priepustiach podľa MEL-frekvenčnej škály ako samostatné rečové črty v spojení s pravdepodobnostnou neurónovou sieťou (PNN), ktorá slúži ako klasifikátor hlasov hovoriacich. Zistili sme, že úspešnosť, vyjadrená v % , okolo 96 % pre jednoslovnú verifikáciu a 98 % pre trojslovnú verifikáciu, je obecne porovnateľná so systémami založenými na MFC alebo PLP črtách v úlohach, ktoré majú podobné nastavenie(počet hovoriacich a typ prehovorenej reči).

Pri našom testovaní sme sa sústredili na výkonnosť vzhľadom k šumu, k dimenzionalite vektorov čŕt a problém počtu vynachaných hovoriacich. Ďalej, naše dáta a výsledky podporujú myšlienku obecnej nezávislosti nášho systému verifikácie hovoriaceho na jeho natívnom jazyku a na rodinných vzťahoch medzi hovoriacimi (hlasová príbuznosť).

V závere sme prediskutovali implikácie našich výsledkov pre ďalší rozvoj automatického rozpoznávania reči a verifikáciu hovoriaceho z hľadiska symetrie fonémického systému. Do určitého rozsahu konceptuálne diskutujeme nami navrhnutý model percepcie, ktorý zahrňuje práve spomínané symetrie foném fonologického systému reči. Nakoniec prediskutujeme niektoré zaujímavé analógie medzi systémom symetriií foném a symetriami elementárnych častíc.

Kľúčové slová: Rozpoznávanie reči; verifikácia hovoriaceho; topologické invarianty; rečové črty; Nyquistova rovina, MFC; PLP; HMM; LDA; MASPER; dekorelačná transformácia; sluchový systém; TIM; MEL-frekvencia; PNN; percepcia reči; symetrie foném; fonologický systém; elementárne častice.

1 Úvod

Dnešné rozhrania stroj-človek využívajú zložité optické a manuálne nástroje, ktoré ale zaneprázdňujú ruky a oči človeka. Riešenie tohto problému je rečová komunikácia, ktorá môže uvolniť užívateľa od zložitých a neprirodzených manipulácií. Zároveň s rastom moderných telekomunikačných technológií a multimediálnych aplikácií rastie potreba technológií s explicitnými rečovými schopnosťami. Rečová komunikácia sa tak stáva podstaným znakom v rozvoji budúcich technológií z hľadiska viacerých aspektov:

1. Vedecké aspekty - dosiahnuté výsledky v rečovej komunikácii ovplyvnia aj príbuzné oblasti a dôležité oblasti ako sú Umelá inteligencia a Kognitívne vedy, *Mařík, Štěpánková, Lažanský a kol.* (1993-2003), *Rittter, Kohonen* (1989), *Hinton, Anderson* (1981), *Hogg, Huberman* (1987), *Minsky, Papert* (1969).

2. Ekonomické aspekty - spracovanie reči sa nachádza na priesečníku informatiky a telekomunikácie, čo sú dve technológie, ktoré budú určite modelovať našu budúcnosť, *Martin* (1975).

3. Kultúrne aspekty - v dnešných dňoch sa technológie komunikácie stávajú stále väčšou súčasťou našej spoločnosti a preto je veľmi dôležité, aby sa stroje adaptovali smerom k našim prirodzeným jazykom a nie naopak, a nie iba k jednému jazyku.

V dizertačnej práci sa zaoberáme invariantnými prístupmi k spracovaniu rečového signálu na báze neurónových sietí a ich softwarovými a hardwarovými implementáciami. Na základe biologických a kybernetických predpokladov sme vytvorili ucelený systém, ktorý modeluje primárne rečové úlohy – rozpoznávanie slov a verifikáciu hovoriaceho.

Model opisuje proces percepcie cez invariantnú extrakciu čŕt, učenie a klasifikačné úlohy pomocou pravdepodobnostných neurónových sietí. Cieľom modelovania je rozpoznávanie, ktoré je nezávislé – invariantné od hovoriaceho, a redundantných fyzikálnych a fonetických parametrov, resp. identifikácia hovoriaceho invariantná na redundantných fyzikálnych a fonetických parametroch.

K tomu prislúchajúci prístup cez grupy symetrií je explicitne zahrnutý v modeli pomocou homotopických a kalibračných grúp symetrie.

Štruktúra dizertácie

V druhej kapitole sa zaoberáme prehľadom metód a prístupov k spracovaniu reči cez počítač, špeciálne k rozpoznávaniu izolovaných slov a verifikácii hovoriaceho.

V tretej časti diskutujemea prevádzame analýzu reči z pohľadu fonetického predspracovania, ktoré je založené na neurónovom modeli. Rozoberáme biologickú motiváciu, ktorá je pri vstupe rečového signálu do sluchového traktu. Hlavný dôraz je na preskúmaní a možnom vylepšení štandardných metód. Posledné používajú nasledujúcu stratégiu:

1. aplikuje sa štandardné akustické predspracovanie, ktoré transformuje rečový signál na koeficienty, parametre typu – frekvenčných spektier, kepstrálnych, LPC (Linear Prediction Coefficients) koeficientov, PLP (Perceptual Linear Predictive), MFC (Mel-Frequency Cepstrum).

2. použije sa učenie a metódy klasifikácie získaných reprezentácií rečového signálu.

Predošlé stratégie sa dajú zhrnúť ako intenzitné, energetické prístupy, avšak je známe, že biologický sluchový systém je citlivý nielen na spektrálne reprezentácie ale aj rôzne prechodové javy v časovej oblasti. Cieľom práce je aj vyšetriť do akého rozsahu sa dá aplikovať fázový (nie energetický) prístup na rozpoznávanie reči a verifikáciu hovoriaceho.

Vo štvrtej kapitole tejto práci sme navrhli biologicky plauzibilný, nelineárny systém, systém topologických invariantov, ktorý je závislý na extrakcii invariantných čŕt, čo následne implikuje prístupy grúp symetrie. Navrhnutá trieda modelov sa dá chápať ako model umelých neurónových sietí v rečovej percepcii a klasifikácii.

V druhej časti štvrtej kapitoly sa sústredime na metódy učenia a klasifikácie pomocou umelých neurónových sietí. Sústredime sa na triedu pravdepodobnostných neurónových sietí, Probabilistic Neural Network, ďalej PNN. V klasických modeloch sa procesy učenia a klasifikácie odlišujú, prevádzajú sa v dvoch oddelených fázach - učenia a odozvy alebo reakcie neurónovej siete.

Avšak, u ľudí oba mechanizmy fungujú súčasne, hoci hlavný učiaci proces nastáva v prvých rokoch života. Takže dospelí používajú už efektívne zostrojený súbor učiacich prototypov, ktorý sa len jemne adjustuje v čase dospelosti. V tejto práci modelujeme oba mechanizmy v rámci jednej neurónovej štruktúry - PNN. Namiesto toho, aby tieto siete pracovali s triedami vzorov pre dané slovo alebo hovoriaceho, pracujú implicitne s úplnými pravdepodobnostnými rozdeleniami vzorových paternov slov alebo hovoriacich.

V časti 5 venovanej výsledkom zhrnieme hlavné výsledky pre rozpoznávanie slov a porovnáme ich so štandardnými prístupmi. V časti 6 uvedieme výsledky, ktoré sme dosiahli pre úlohu verifikácie hovoriaceho. V siedmej časti, venovanej diskusii, diskutujeme niektoré zaujímavé, neštandardné črty nášho prístupu z hľadiska ľudskej rečovej percepcie a symetrií prirodzených jazykových systémov, špeciálne slovenčiny. a dôsledky nášho modelu z hľadiska informačného chápania symetrií a fundamentálne koncepcie k prístupu obecnej teórie poľa na generovanie a spracovanie symetrií v percepcii a spracovaní rečového signálu nervovým systémom človeka

V poslednej časti diskutujeme niektoré možné hypotézy vyplývajúce z nášho chápania percepcie.

V dodatkoch rozoberáme technické, programové a niektoré matematické aspekty nášho prístupu. Zároveň pre ilustráciu uvádzame systémy kvalifikácie elementárnych častíc a anglických foném.

2 Prehľad používaných metód a prístupov

Súčasné modely rozpoznávania nie sú úplne dokonalé pre úlohy nezávislé od hovoriaceho a zároveň pre súvislú reč. Podobne, úlohy verifikácie sa uvažujú iba pre ohraničený počet hovoriacich, väčšinou veľmi nízky (pod 10) a pre úlohy závislé na slove, slovách alebo pre jednotlivé slová, chápané ako heslo a pre kvalitné akustické podmienky rečového vstupu. V nasledujúcom prehľade sa sústredime na prístupy k takto ohraničeným úlohám rozpoznávania a verifikácie.

2.1 Rozpoznávanie slov

Obecne povedané rozpoznávanie reči spočíva v určení jazykového obsahu slova alebo vety, ktorú vysloví hovoriaci. Proces ľudského rozpoznávania reči často používa viacero senzorických zdrojov, ako sú mimika, gestikulácia, sluchový vstup samotný a spätnú väzbu z porozumenia reči a jazyka, na presné určenie jazykového obsahu. Za prvý pokus o elektronické spracovanie reči z pohľadu úlohy rozpoznávania alebo dokonca porozumenia sa môže pokladať práca *Pottera, Koppa* a *Koppa* „Visible Speech“ z roku 1947, (1975). Bez ohľadu na fakt, že od roku 1960 je jasné, že konečný cieľ rozpoznávanie súvislej reči nezávisle na hovoriacom je jedným z najťažších problémov rozpoznávania, práca v tejto oblasti pokračuje vyšetrovaním, skúmaním, čiastočných úloh rozpoznávania, pozri napríklad obsiahlu správu výskumného tímu projektu ARPA, *Newell, Barnett* (1973). Táto správa a následne výskum ňou ovplyvnený viedol k takzvanej dogme SUS - Speech Understanding System - výskumu porozumenia reči. Skladá sa z troch tvrdení, záverov:

1. Kritérium úspešnosti je porozumenie správy, komunikácie a nie len rozpoznanie foném, slov alebo viet.

2. Musia byť použité všetky zdroje znalostí.

3. V samotnom rečovom signále nie je informácia dostatočná na určenie obsahu správy

1. dogma (pojem prevzatý z *Newell, Barnett* (1973)) sa má rozumieť ako vyhlásenie o základnom princípe každého systému a síce, že systémové, celkové správanie je dôležitejšie a nie chovanie jednotlivých zložiek. Zaujímavé na tejto dogme je, že doteraz nikto pragmaticky neurčil, t.j. nezmeral úspešnosti jednotlivých úrovní, zložiek a neporovnal ich. Z tohto hľadiska sa nám zdá táto dogma skôr filozofická a nie pragmatická, vedecká.

2. dogma bola z hľadiska histórie a zamerania výskumu, ktorý prebehol odvtedy až do dnešných dní, oveľa významnejšia. Usmernila úsilie výskumníkov na dlhé roky k štruktúre celého systému a nie k analýze reči, pretože úloha spojenia všetkých zdrojov znalostí dohromady je centrálnou a nie periférnou úlohou. Dá sa chápať ako operačný princíp, ktorý je základným kameňom SUS.

3. dogma je oveľa viac hodnovernejší princíp, ku ktorému sa hlásia alebo hlásili viacerí bádatelia, *Newell, Barnett* (1973), *Reddy, Erman* (1975). Táto dogma sa dá vysvetliť z pohľadu ľudského adaptívneho správania sa, ktoré predpokladá, že ľudia nevkladajú ro rečového signálu viac ako je nutné pre prijímateľa (akýsi princíp minimálnej akcie pri komunikácii, pozri aj diskusiu na záver, časť 7). Takto, ak obyčajný poslucháč používa určitý kontext, hovoriaci bude degradovať - minimalizovať vstup - rečový signál odpovedajúco k danému kontextu. Na druhej strane táto dogma sa dá použiť aj ako podpora pre druhú dogmu. Ak platí tretia dogma, potom je určite pravdivá aj druhá dogma. Ak tretia dogma neplatí, potom ... a toto bolo vlastne motívom našej práce, snaha o čo najlepšie vyriešenie problému na najnižšej úrovni, *Reddy* (1966) .

Na tomto štádiu diskusie je si dobré pripomenúť ako reprezentujeme rečový vstup - začneme vlnami tlaku vo vzduchovom stĺpci až po elektronickú reprezentáciu pomocou postupnosti diskrétnych meraní amplitúdy rečového signálu, ktoré sú namerané s dostatočnou presnoťou 12 - 16 bit a s dostatočnou rýchlosťou konverzie 10 - 20 Khz. Samozrejme tieto parametre závisia od konkrétnej úlohy resp. prenosového kanálu, na tomto mieste ich treba chápať ako najlepšie postačujúce podmienky pre obecne postavenú SUS. Pod výstupom, porozumením chápeme v tejto práci konkrétnu reprezentáciu porozumenia relatívnu k danému informačnému systému. Inými slovami nechápeme ho abstraktne ale pragmaticky. A reprezentáciou porozumenia je potom tá reprezentácia z ktorej vyplýva, je možná nejaká akcia. Niekedy sa musí pojem takejto reprezentácie rozšíriť, aby zahrnul ďalšie reprezentácie, ktoré sa napríklad používajú systémom na uchovanie informácie a jej použitie neskoršie na nejakú akciu. V tejto definícii sme nepovedali nič o tom ako systém konkrétne kóduje znalosť pre následnú akciu. Avšak v praxi - pre počítačové systémy, kde môžeme vyšetriť vnútornú štruktúru programu, nie je žiaden problém s určením toho z čoho vyplýva daná akcia, *Winograd* (1975).

Príkladom štandardného prístupu založeného na analýze rečového signálu môže byť dnes už klasická práca „Computer Recognition of Connected Speech“ ktorej autorom je *Reddy* (1975). V tejto práci sa parametre namerané na časových segmentoch reči spracovávajú s logikou a na základe znalostí získaných z analýzy reči pomocu takých fonetických parametrov ako sú energia, základný tón, prechody nulou, formanty a podobne. Pri rozhodovaní o zaraďovaní alebo klasifikácii segmentov reči k jednotlivým fonetickým triedam sa využívajú znalosti fonetiky, prozeodiky a gramatiky. Veľkým záporom takýchto prístupov je množstvo ad hoc konštánt a parametrov, ktoré neprechádzajú procesom učenia a ktoré sú priveľmi zviazané s danými realizáciami rečových jednotiek a daným jazykovým korpusom a systémom.

Podobne uvažuje aj *Flanagan* (1972), kde prichádza k záveru, že automatické rozpoznávanie reči nebude pravdepodobne možné bez hlbokej analýzy a aplikácie gramatických, kontextuálnych a sémantických väzieb. Tieto väzby ale nie sú až doteraz známe resp. úspešne rozpracované, *Rabiner, Wilpon* (1979) .

V našej práci, pre účely rozpoznávania reči pomocou počítača, sme sa obmedzili iba na konverziu rečového signálu na slová. Odpovedá to napríklad rečovej komunikácii cez telefónnu linku, pri ktorej nie je prístupná iná senzorická informácia.

V 80. a 90. rokoch minulého storočia boli aplikované vo väčšine prác tri prístupy k riešeniu diskutovanej problematiky. Prvé dva, Dynamic Time Warping - DTW**,** *Sakoe, Chiba,* (1978) a Hidden Markov Model - HMM, *Piccone* (1990), predpokladajú štatistické prístupy v klasickom zmysle slova, tretí je, obecne hovoriac, Artificial Neural Network- ANN prístup.

Príkladom DTW prístupu môže byť systém na rozpoznávanie hesiel s obmedzenou syntaxou a lexikou vypracovaný v TESLE autormi *Ptáčkom*, *Dušekom* a *Dvořákom* (1985).

Všetky tieto prístupy predpokladajú, že reč je hierarchicky ohraničený systém, pričom na každej úrovni tejto hierarchie sa aplikujú rôzne zdroje znalostí. Cieľom počítačových systémov rozpoznávania je najoptimálnejšie kombinovať tieto zdroje znalostí do jedného celku. Príkladom hierarchie sú nasledujúce úrovne:

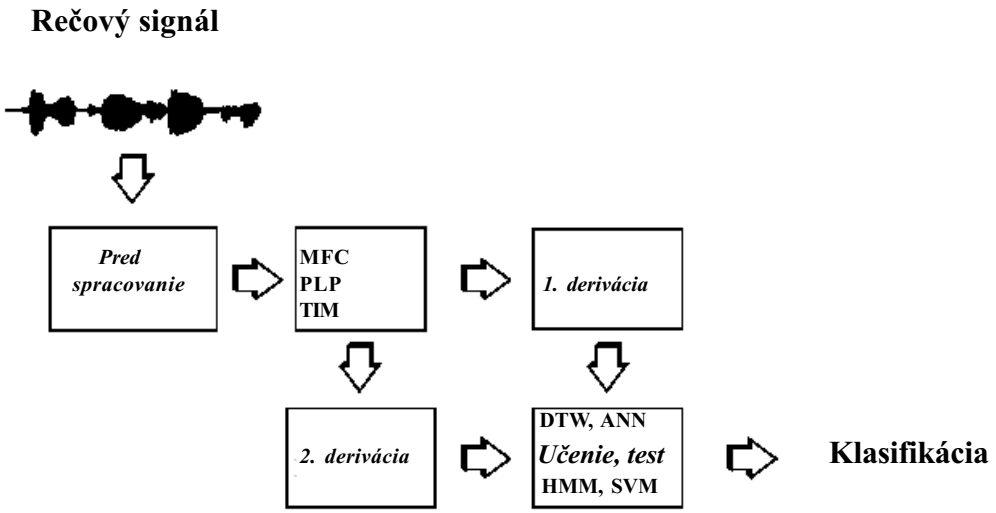
1. Akustická
2. Fonetická
3. Lexikálna
4. Syntaktická
5. Sémantická resp.dialógová
6. Pragmatická

Často sa najnižšie úrovne modelujú pomocou HMM a výsledkom je pozorovaná postupnosť fonetických jednotiek, napríklad hlások.

Zvyčajne sa stratégia rozpoznávania na akustickej a fonetickej úrovni volí nasledovne:

1. aplikuje sa štandardné akustické predspracovanie, pri ktorom sa rečový signál transformuje alebo kóduje na určité koeficienty; spektrálne, kepstrálne, LPC, MFC, PLP, atď.
2. použijú sa zvolené metódy merania vzdialeností, medzi ne sa dá zaradiť aj DTW, učenia medzi ktoré sa dá zaradiť aj HMM a klasifikácie na transformované koeficienty reprezentujúce rečový signál, *Robinson, Fallside* (1988).

Jadrom všetkých spomínaných prístupov je analýza, v časových oknách, digitalizovaného rečového signálu. Ak predpokladáme, že reč je stacionárna, počas krátkeho trvania, rádove 10 ms, namerané kepstrálne koeficienty sú korelované s tvarom vokálneho traktu a polohou hlasiviek, *Maeda, S.* (1982). V skutočnosti nie je tento predpoklad splnený. Poloha vokálneho traktu resp. hlasiviek sa nemení na hraniciach časového okna ani vo vnútri okna nie sú úplne nemenné. Aby sme boli schopní zobrať do úvahy tieto fakty musíme do modelu zahrnúť prvé resp. druhé derivácie relevantných veličín, Obr.1.



Obr.1 Všeobecná schéma klasického systému rozpoznávania

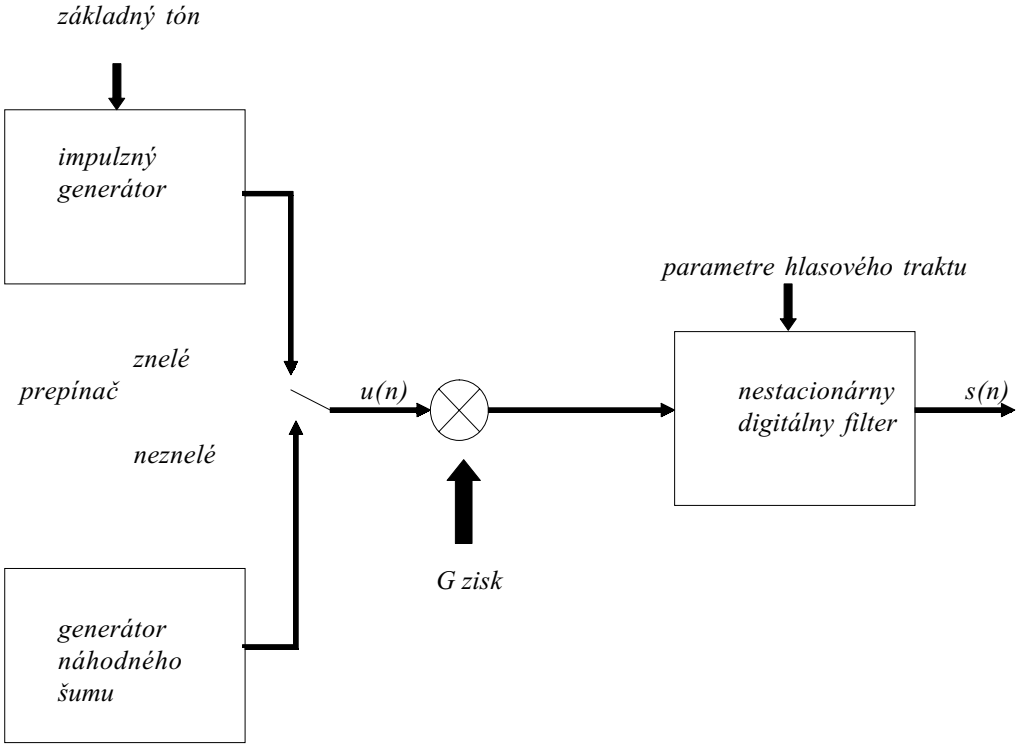
2.1.1 LPC, PLP, MFC parametrizácia rečového signálu

Ako sme už spomenuli na extrakciu vektorov čŕt signálu reči sa používajú funkcie LPC, MFC, PLP, atď. V tejto práci bližšie opíšeme tri metódy parametrizácie rečového signálu: LPC ako najstaršiu metódu, PLP ktorá používa predošlú techniku (LPC) a nakoniec metódu MFC založenú na diskrétnej kosínusovej transformácie (DCT). Tieto tri rôzne spôsoby majú niektoré koncepčné podobnosti z hľadiska spracovania rečových signálov.

2.1.1.1 Lineárna predikcia rečového signálu

LPC techniky sa považujú za výkonné nástroje pre analýzu reči, či ako samostatné alebo ako pomocné nástroje pre sofistikovanejšie prístupy. Používajú sa pre také algoritmy ako je odhad základného tónu, formantov reči, tvaru a parametrov vokálneho traktu, na bitovú kompresiu rečového signálu, alebo ako súčasť napr. PLP, *Markel, Gray* (1976), *Makhoul* (1975), *Makhoul* (1975).

Reč sa modeluje ako výstup lineárneho, časovo premenného systému, ktorý sa budí buď periodickým signálom alebo šumom. Je zvykom nazývať tieto dva zdroje signálu znelým (v reálnom trakte odpovedá kmitaniu hlasiviek) a neznelým (odpovedá to neprítomnosti kmitania hlasiviek, šumu) zdrojom, Obr. 2.



Obr. 2 Schéma lineárnej predikcie reči

Pri LPC modelovaní sa uvažuje vzorkovaný, digitalizovaný signál s(n) v čase n, ktorý aproximujeme lineárnou kombináciou predošlých p vzoriek ako

$$s'(n) = a_1 s(n-1) + a_2 s(n-2) + ... + a_p s(n-p), \qquad (1)$$

kde a_i sú konštantné koeficienty, tzv. LPC koeficienty. Pridaním excitačného člena gu(n), podľa modelu na Obr. 2, dostaneme

$$s'(n) = \Sigma a_i\, s(n-i) + gu(n) \qquad \text{pričom v sumácii } i = 1,p \qquad (2)$$

g má význam zisku, a u(n) je normalizovaná (jednotková) excitácia. Ak prevedieme predošlú rovnicu do z oblasti dostaneme

$$S'(z) = \Sigma a_i\, z^{-i} S'(z) + gU(z)$$

a prechodová funkcia, v z obraze, H(z) nášho systému bude potom

$$H(z) = S'(z) / gU(z) = 1 / A(z) = 1 / (1 - \Sigma a_i\, z^{-i}) \qquad (3)$$

priamo vidíme, že H(z) je len pólový model systému. Ak nedávame nejaké špeciálne obmedzenia na koeficienty a_i apriori, a predpokladáme, že rozdelenie chýb modelu od reálneho signálu je gausovské,

potom chyba e(n) pre ľubovolný čas (n) je daná ako

$$e(n) = s(n) - s'(n) = s(n) - \Sigma a_i\, s(n-i)$$

Fitovaciu funkciu (cost function) definujeme na časovom okne konečnej dĺžky N ako

$$E = \sum_{n=0}^{N-1} e^2(n) \qquad (4)$$

Ak nájdeme minimum E podľa parametrov a_i pomocou prvej derivácie E dostaneme sústavu rovníc v tvare

$$\Phi_{i0} = \sum_{k=1}^{p} \Phi_{ik}\ a_k, \qquad (5)$$

kde Φ je kovariančná matica (z definície symetrická) definovaná ako

$$\Phi_{ik} = \sum_{n=0}^{N-1} s(n-i)\, s(n-k)$$

v maticovom zápise potom sústavu rovníc pre minimum zapisujeme ako

$$\mathbf{\Phi_0} = \mathbf{\Phi a}$$

s riešením

$$\mathbf{a} = \mathbf{\Phi^{-1}\Phi_0}. \qquad (6)$$

Metódam riešenia predošlých rovníc sa tu nebudeme venovať (autokorelačná a kovariančná metóda). Pomocou rekurentných vzťahov sa z takto definovaných koeficientov LPC dajú vypočítať spektrálne charakteristiky, kepstrálne koeficienty, prierezy vokálneho traktu a aj formanty reči, *Markel, Gray* (1976).

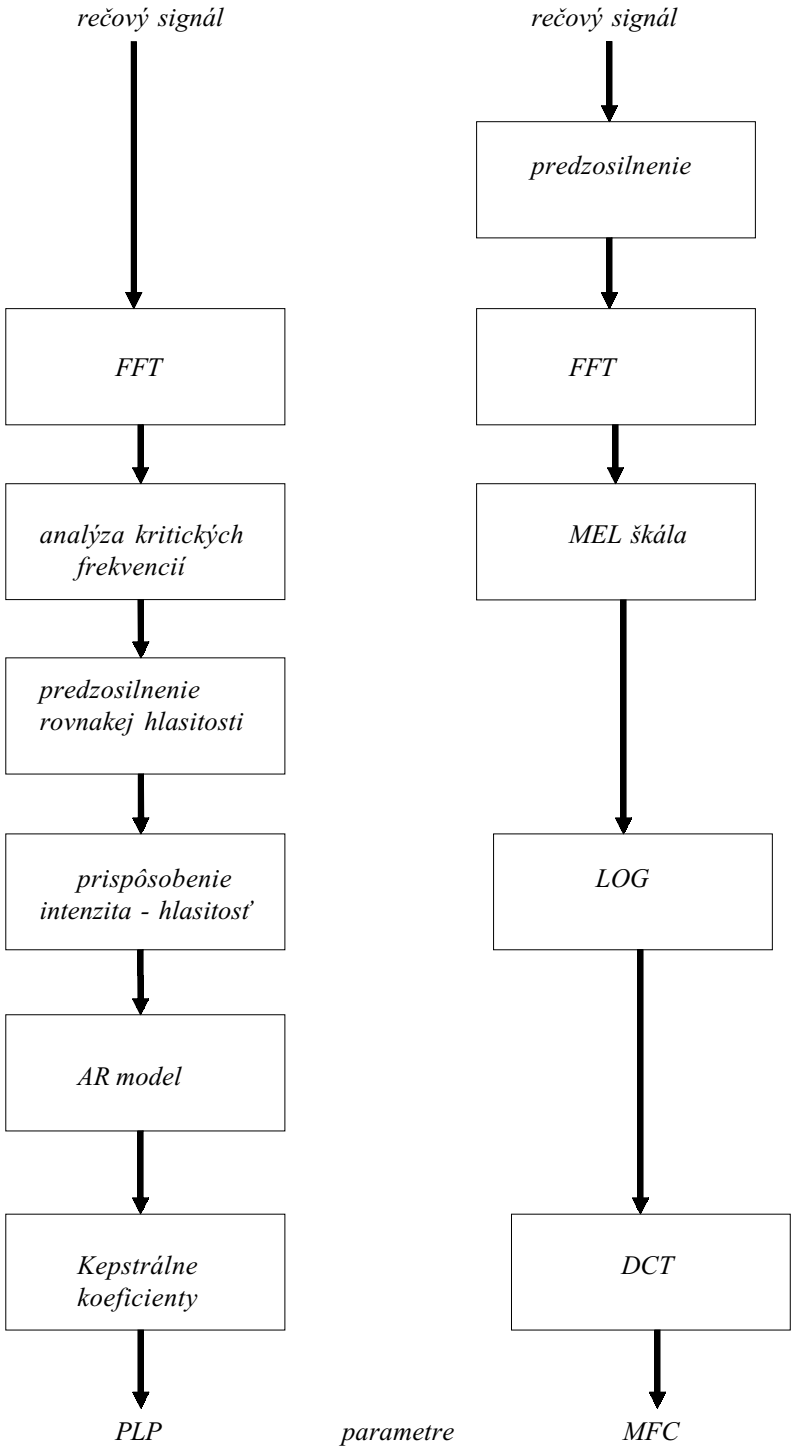
2.1.1.2 PLP parametrizácia rečového signálu

Perceptuálna lineárna predikcia (PLP) je metóda analýzy založená na LPC metóde, v krátkosti prediskutovanej v predošlej časti. PLP zahrňuje nelinearitu precepcie frekvenčnej škály a iné známe vlastnosti psychofyziky počutia, *Hermansky* (1990). Pri analýze PLP sa najskôr použije Fourier transformácia na výpočet krátkodobého výkonového spektra a na takéto parametre sa sa aplikujú perceptuálne vlastnosti, spektrum sa transformuje na Barkovu škálu, a takto upravené spektrum sa predzosilní pomocou funkcie, ktorá aproximuje citlivosť ucha pre rôzne frekvencie.

bark	1	2	3	4	5	6	7	8	9	10
kHz	0,102	0,204	0,309	0,417	0,531	0,652	0,781	0,923	1,079	1,256
bark	11	12	13	14	15	16	17	18	19	20
kHz	1,457	1,692	1,972	2,310	2,727	3,248	3,904	4,729	5,758	7,031

Tab. 1 Barkova škála, ktorá korešponduje percepčnému frekvenčnému rozdeleniu ľudského ucha

Takýto výstup sa komprimuje tak, aby aproximoval nelineárny vzťah medzi intenzitou zvuku a percepovanou hlasitosťou reči. Potom sa použije LPC (iba pólový model), na vyhladenie simulovaného rečového spektra, a nakoniec sa takéto LP parametre transformujú na kepstrálne koeficienty, a tieto sa potom chápu ako parametre pre rozpoznávanie reči alebo hovoriaceho. Celý algoritmus, PLP parametrizácie je uvedený na ľavej strane Obr. 3.



Obr. 3. Schéma PLP (vľavo) a MFC (vpravo) parametrizácie

Podrobnejšie, pre každý segment reči s(k) váhovaný Hammingovým oknom, sa vypočíta krátko-časové energetické spektrum P(ω) zo signálového spektra S(ω) v uvažovanom segmente reči, pomocou vzorca

$$P(\omega) = \left|S(\omega)\right|^2 = \left[\mathrm{Re}\, S(\omega)\right]^2 + \left[\mathrm{Im}\, S(\omega)\right]^2 . \qquad (7)$$

Ďalším krokom je vyjadrenie maskovania zvukov v kritických frekvenčných pásmach s meniacou sa veľkosťou týchto pásiem od frekvencie, pozri Tab.1, používa sa nasledovné nelineárne priradenie

$$\Omega(\omega) = 6 \ln\left(\frac{\omega}{1200\pi} + \sqrt{\left(\frac{\omega}{1200\pi} \right)^2 + 1} \right), \qquad (8)$$

kde $\omega = 2\pi f$ je frekvencia v radiánoch a Ω(ω) je v barkoch. Detaily vlastností takto definovaných priepustí - frekvenčné charakteristiky a aplitúdovo frekvenčné odozvy - sú v *Hermansky* (1990).

Prispôsobenie energetického spektra P(ω) vlastnostiam ľudského ucha potom spočíva v predzosilnení a v aproximácii kriviek rovnakej hlasitosti.

$$\Phi_m\left(\Omega(\omega)\right) = E(\omega)\Psi\left(\Omega(\omega) - \Omega_m\right), \qquad (9)$$

kde Ω_m(bark) je stredná frekvencia m-teho kritického priepusťového filtra, m =1, 2,..., M.

Nasleduje váhovaná spektrálna sumarizácia v energetických spektier. Hodnoty energetického spektra P(ω) po prejdení m-tym kritickým pásmom prispôsobené krivkám rovnakej hlasitosti môžeme vyjadriť ako

$$\Xi\left(\Omega_m\right) = \sum_{\Omega=\Omega_m-2,5}^{\Omega_m+1,3} P(\Omega)\Phi_m(\Omega), \qquad (10)$$

kde hranice sčítania sa dajú vypočítať z inverzného vzťahu (8).

V ďalšom kroku sa prevádza prispôsobenie intenzity na hlasitosť, pomocou vzťahu

$$\xi\left(\Omega_m\right) = \left(\Xi\left(\Omega_m\right)\right)^{0,3}, \qquad (11)$$

čím sa zároveň redukuje amplitúdová variabilita v kritických pásmach. Posledným krokom je iba pólová LPC aproximácia takéhoto spektra, *Hermansky* (1990).

2.1.1.3 MFC parametrizácia rečového signálu

Mel-frekvenčné kepstrálne koeficienty (MFC) je metóda analýzy reči založená na modifikovanej homomorfnej metóde analýzy. Kompenzácia nelineárneho percepčného správania sa ucha je prevedená pásmom trojuholníkových filtrov s lineárnym rozdelením frekvencií podľa melovej škály

$$f_m = 2595 \log_{10}(1 + \frac{f_m}{700}), \qquad (12)$$

kde f je frekvencia v lineárnej škále a f_m je frekvencia v odpovedajúcom nelineárnom mel-frekvenčnom rozsahu.

Vstupný signál s(k) je v prvom kroku predzosilnený jednoduchým FIR filtrom

$$H(z) = 1 - az^{-1}, \tag{13}$$

kde koeficient filtra a je z intervalu [0.9,1]. V ďalšom kroku sa signál oknuje Hammingovým oknom o dĺžke z intervalu približne 10 - 30 ms. Kvôli následnému použitiu rýchlej Fourier transformácie sa dĺžka okna vyberá ako mocnina dvoch. V ďalšom kroku sa pomocou FFT vypočíta amplitúdové spektrum |S(ω)| signálu, *Rabiner, Juan* (1993).

Nasleduje najdôležitejší krok, prevedenie mel-frekvenčnej transformácie do pásma trojuholníkových filtrov rovnomerne rozdelených na frekvenčnej osi v mel-frekvenciách. Potom sa výstup každého takéhoto filtra logaritmuje, čo zohľadňuje dynamický rozsah signálu. V konečnom kroku sa prevedie diskrétna kosínová transformácia (DCT)

$$c_{mf}(n) = \sqrt{\frac{2}{K}} \sum_{j=1}^{K} \log m_j \cos\left[n\left(j - 0.5 \right) \frac{\pi}{K} \right], \tag{14}$$

kde n je počet mel-kepstrálnych koeficientov a K je počet mel-frekvenčných pásiem. Samotná kosínová transformácia generuje obecne nekorelované koeficienty, čo má kladný vplyv v ďalších fázach klasifikácie či pomocou skrytých Markovových modelov (HMM), vektorovej kvantifikácie (SVM), alebo modelov umelých neurónových sietí (ANN).

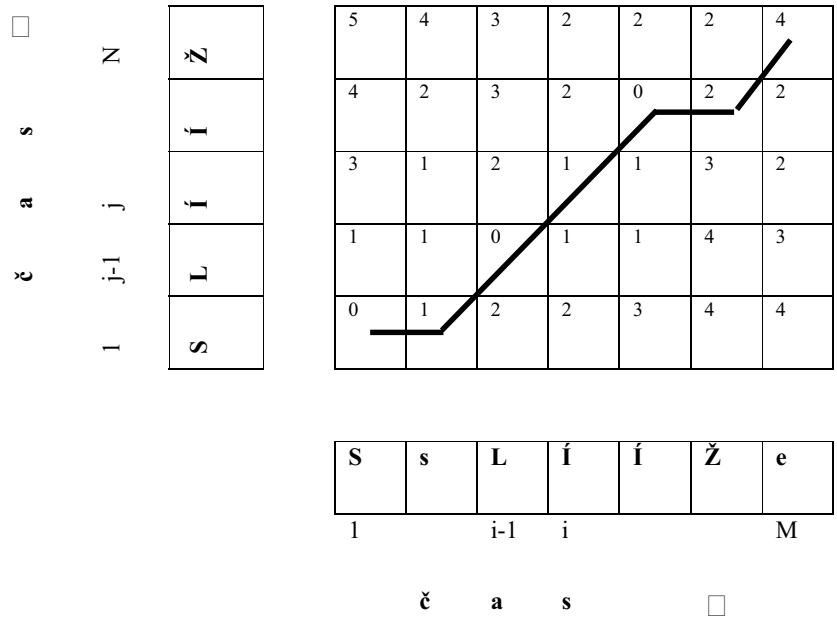
2.1.2. Dynamické programovanie

V tomto type techniky rozpoznávania reči (izolovaných slov) sa testované dáta prevedú na vzory. Proces rozpoznávania potom spočíva v porovnaní vstupného rečového vzoru s uchovanými vzormi - prototypmi. Vzor s najnižšou mierou vzdialenosti od vstupného vzoru sa pokladá za rozpoznané slovo. Najnižšia miera vzdialenosti je založená na dynamickom programovaní. Pre úlohy z rozpoznávania slov sa nazýva Dynamic Time Warping (DTW). Pri definovaní miery vzdialenosti používame dva typy vzdialeností, lokálnu vzdialenosť medzi dvomi vektormi čŕt signálu, pre signál 1 označíme tento vektor x resp. y pre signál 2. Štandartne sa táto vzdialenosť počíta pomocou euklidovskej vzdialenosti. V našom prípade, lokálna vzdialenosť medzi vektorom x (signálu 1) a vektorom y (signálu 2) je daná ako

$$d(x, y) = \sqrt{\sum_i (x_i - y_i)^2} \ .$$

Druhý typ vzdialenosti – globálna – je rozdielom medzi celým signálom (prvé slovo) a druhým signálom (druhé slovo, alebo druhá realizácia toho istého slova) možno s inou časovou dĺžkou. Reč je časovo-závislý proces. Preto realizácie rovnakého slova majú rôzne trvania, a realizácia rovnakého slova s rovnakým trvaním sa budú možno líšiť v strede, pretože rôzne časti slov môžu byť vyslovené s rôznymi rýchlosťami. Aby sme dostali spomínanú globálnu vzdialenosť medzi dvomi rečovými vzormi, musíme previesť časové usporiadanie.

Na nasledujúcom obrázku je toto usporiadanie ilustrované pomocou čas – čas matice, matice blízkosti (proximity matrix). Na vertikálnej osi je znázornený príklad vzoru „SLIIŽ“ a na vodorovnej



Obr. 4 Schéma matice blízkosti pre DTW

je znázornený vstupný vzor. Vstup „SsLIIše" je `zašumená' verzia vzoru „SLIIŽ". Myšlienka za DTW je, že `š' je bližšie k `Ž' v porovnaní s hocičím iným vo vzore. Vstup „SsLIIše" sa porovnáva so všetkými vzormi z daného systému rozpoznávania. Najlepším vzorom pri porovnaní je ten, ktorý má najnižšiu vzdialenosť trajektórie priradenia vstupného paternu k vzoru. Najjednoduchším skóre globálnej vzdialenosti pre trajektóriu je jednoducho súčet lokálnych vzdialeností, ktoré vytvárajú danú trajektóriu.

Aby sa redukovali výpočty, zjednodušujú sa niektoré predpoklady ohľadom smeru priradenia. 1. Porovnávacie trajektórie nesmú byť spätné v čase (podmienka neklesajúcej monotónnosti). 2. V porovnávacej trajektórii sa musí použiť každý časový rámec. 3. Lokálne skóre vzdialeností sa kombinujú do globálnej vzdialenosti jednoducho sčítaním.

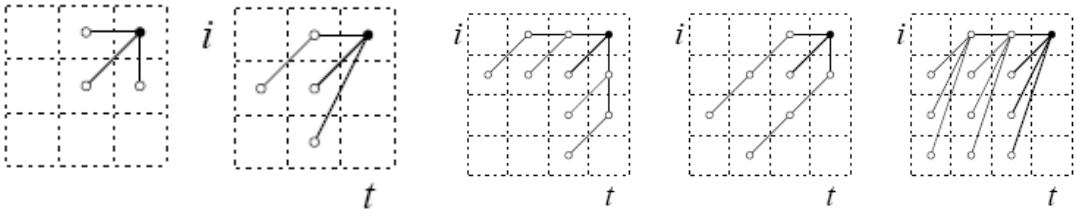
Takýto algoritmus sa nazýva dynamické programovanie (DP). Okrem predošlých predpokladov garantuje nájdenie najmenšej vzdialenosti naprieč maticou blízkosti, pri minimálnom počte výpočtov. DP algoritmus operuje spôsobom, aby sa spracovával po sebe každý stĺpec matice blízkosti (ekvivalentné spracovaniu vstupného paternu časový rámec po rámci, takže pre vzor o dĺžke N je maximálny počet trajektórií, ktoré musíme brať do úvahy, rovný N. Ak sa DP aplikuje na rozpoznávanie reči (založené na vzoroch), nazýva sa Dynamic Time Warping (DTW), *Sakoe, Chiba* (1978).

Ak je D(i,j) globálna vzdialenosť až do bodu (i,j) a lokálna vzdialenosť je v (i,j) daná ako d(i,j) potom

$$D(i,j) = \min (D(i -1, j - 1), D(i - 1, j), D(i, j - 1)) + d(i,j) \qquad (15)$$

Nech platí D(1,1) = d(1,1) (čo je jednoducho počiatočná podmienka), potom predošlý vzťah je rekurzívny algoritmus pre výpočet D(i,j). Konečná globálna vzdialenosť D(M,N) nám udáva celkové skóre porovnania vstupného vzoru s prototypom. Vstupné slovo sa potom pokladá za rozpoznané ako slovo odpovedajúce prototypu s najmenším skóre porovnania. Možné spôsoby iterácie najbližších susedov sú ilustrované na schéme, Obr. 5.

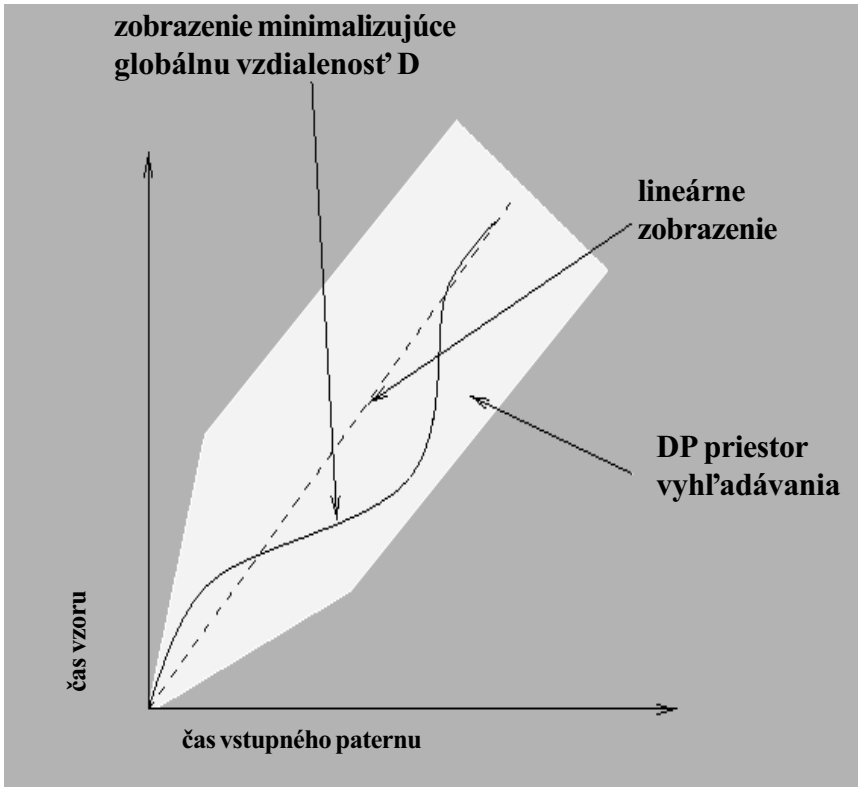
Vidíme, že DTW algoritmus namiesto porovnania hodnôt vstupného vzoru v čase t ku prototypu v čase t, používa priestor zobrazení z časovej postupnosti vstupného vzoru do postupnosti prototypu, aby minimalizoval globálnu vzdialenosť D. Toto zobrazenie nie je vždy lineárne; napríklad, čas t1 vo



Obr. 5 Spôsoby iterácie pre štandardné DTW

vstupnom vzore odpovedá času t1 + 5 v prototype, zatiaľčo čas t2 vo vstupnom vzore odpovedá času t2 - 3 v prototype.

Na nasledujúcom obrázku, Obr. 6 je ilustrované toto zobrazenie, kde vodorovná os reprezentuje vstupný vzor a vertikálna os reprezentuje prototyp. Krivka ukazuje minimálnu vzdialenosť medzi vzorom a prototypom. Priestor vyhladávania je znázornený bielou skosenou časťou. Pomocou DTW sa dá nájsť globálne minimum v polynomiálnom čase $O(N^2V)$, kde N je dĺžka časovej postupnosti prototypu a V je počet prototypov v danej úlohe rozpoznávania, *Nakagawa* (1992).



Obr. 6 Nelineárne zobrazenie v DTW priestore

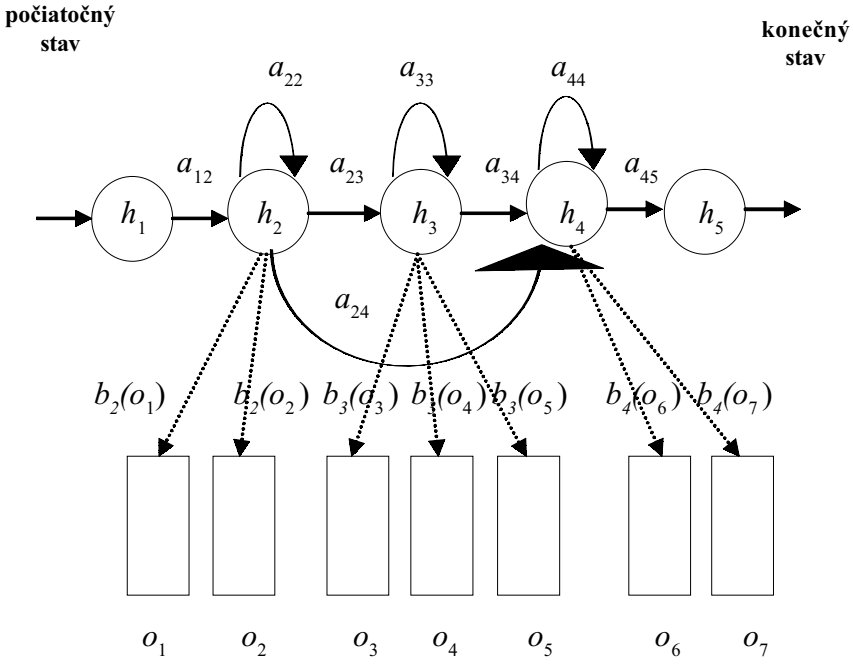
DTW má niekoľko nedostatkov. V prvom rade jeho výpočtová zložitosť je len $O(N^2V)$. Po druhé, ak máme rôzne kanály s rôznymi charakteristikami, potom je veľmi ťažké definovať lokálnu vzdialenosť. Neposkytuje žiadny zmysluplný popis dát. Po štvrté, prototypy sa negenerujú z dát triviálne - a typicky je takéto generovanie sprevádzané párovým porovnávaním trénovacích reprezentácií. Alternatívne, sa ako prototypy uchovávajú všetky pozorované reprezentácie, čo vedie k pomalej odozve systému. To je jeden z hlavných dôvodov prechodu od DTW k HMM (Hidden Markov Model). V našej práci využívame DTW len ako ďalšiu mieru vzdialenosti (k euklidovskej, city-block), resp. pre porovnanie nášho modelu s klasickými prístupmi, medzi ktoré patrí aj DTW.

2.1.3. Markovove modely so skrytou vrstvou

Skryté Markovove modely sa často používajú v kognitívnych vedách, v technike a sociálnych vedách (rozpoznávanie reči, OCR, strojový preklad, bioinformatika, počítačové videnie).

Štandartné DTW je v podstate založené na deterministickom DP (dynamickom programovaní). Avšak, rečové signály sú stochastické procesy. V roku 1988 sa navrhol algoritmus "stochastický DTW". V tejto metóde, sa namiesto lokálnych vzdialeností (ako v štandardnom DTW) používajú podmienené pravdepodobnosti a prechodové pravdepodobnosti namiesto hodnoty trajektórie. Dá sa ukázať, že takýto prístup je ekvivalentný HMM, *Nakagawa* (1992). Podľa *Nakagawa* (1992), úspešnosť rozpoznávania pre izolované slová sa pre stochastické DTW zlepšila z 89.3% na 92.9%.

Skryté Markovove modely je variant pravdepodobnostného stroja s konečným počtom stavov, je to trojica, ($\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{I}$). Kde $\boldsymbol{A}$ sú prechodové pravdepodobnosti, $\boldsymbol{B}$ sú výstupné pravdepodobnosti a $\boldsymbol{I}$ sú počiatočné pravdepodobnosti. To čo nie je pozorovateľné sú skryté stavy $\boldsymbol{H} = \{ h_i \}$, i = 1, . . . , N, a výstupná abeceda (pozorovania) $\boldsymbol{O}$. Stav v danom okamžiku nie je pozorovateľný. Každý stav produkuje výstup s pravdepodobnosťou $\boldsymbol{B}$, *Rabiner* (1989), (1993) pozri nasledujúci obrázok, Obr.7.



Obr. 7 Stavový diagram, konkrétneho, HMM automatu

Prechodové pravdepodobnosti $\boldsymbol{A} = \{a_{ij} = P(h_j \text{ v } t+1 \mid h_i \text{ v } t)\}$, kde P(a | b) je podmienená pravdepodobnosť a pri zadanom b, t = 1, . . . , T je čas, a h_i je z $\boldsymbol{H}$. Inými slovami, $\boldsymbol{A}$ je pravdepodobnosť, že ak v danom čase je stroj v stave h_i pravdepodobnosť, že nasledujúci stav je h_j. Výstupná abeceda je definovaná ako $\boldsymbol{O} = \{ o_k \}$, k = 1, . . . , M . Výstupné pravdepodobnosti sú $\boldsymbol{B} = \{ b_{ik} = b_i(o_k) = P(o_k \mid h_i) \}$, kde o_k je z $\boldsymbol{O}$. Inými slovami, $\boldsymbol{B}$ je pravdepodobnosť, že ak v danom čase je stroj v stave h_i pravdepodobnosť, že výstupný stav je o_k. Pravdepodobnosti v počiatočnom stave sú popísané ako $\boldsymbol{I} = \{p_i = P(h_i \text{ v } t = 1)\}$.

Pre náš konkrétny model platí:

$$\sum_{i=0}^{N-1} p_i = 1 , \quad \sum_{j=0}^{M-1} a_{ij} = 1 \quad \sum_{i=0}^{N-1} p_i = 1 \tag{16}$$

kde N je počet HMM stavov (v našom prípade 5), M je počet výstupných pozorovaní. Matica pre prechodové pravdepodobnosti modelu na Obr. 7 je

$$A = \begin{bmatrix} 0 & a_{12} & 0 & 0 & 0 \\ 0 & a_{22} & a_{23} & a_{24} & 0 \\ 0 & 0 & a_{33} & a_{34} & 0 \\ 0 & 0 & 0 & a_{44} & a_{45} \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} . \qquad (17)$$

Výstupné pravdepodobnosti $b_j(o_t)$ sú funkciami hustoty pravdepodobnosti, ktoré generujú vektor $\boldsymbol{o}_t$ v čase t v stave j. Jednou z možností je použiť superpozíciu gausovských pravdepodobností

$$b_j(\mathrm{o}_t) = \sum_{m=1}^{M_r} C_{jm} N(\mathrm{o}_t, \mu_{jm}, \Sigma_{jm}), \qquad (18)$$

kde M_r je počet normálnych zložiek v zmesi, C_{jm} je váha m-tej zložky v zmesi a $N(o_t, \mu_{jm}, \Sigma_{jm})$ je viac rozmerné gausovské rozdelenie so strednou hodnotou μ a kovariančnou maticou Σ. Platí

$$N(\mathrm{o}, \mu, \Sigma) = \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} e^{\frac{1}{2}(\mathrm{o}-\mu)^T \Sigma^{-1}(\mathrm{o}-\mu)}, \qquad (19)$$

kde n je rozmer vektora o.

Predpokladá sa, že každá pozorovaná postupnosť vektorov $\boldsymbol{O} = \{ o_1, o_2, ..., o_T \}$ sa generovala modelom HMM - $\boldsymbol{M} = \{\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{I}\}$. Pravdepodobnosť, že náš vektor $\boldsymbol{O}$, Obr. 7, bol generovaný modelom $\boldsymbol{M}$ je nasledovná

$$P(O, S|M) = a_{12} b_2(\mathrm{o}_1) a_{22} b_2(\mathrm{o}_2) a_{23} b_3(\mathrm{o}_3) a_{33} ... \qquad (20)$$

V praxi poznáme iba vektor $\boldsymbol{O}$, stavy stroja sú skryté. Pravdepodobnosť generovania postupnosti $\boldsymbol{O}$ z modelu $\boldsymbol{M}$ sa potom vypočíta ako súčet cez všetky možné sekvencie stavov modelu

$$P(O|M) = \sum_X a_{x(0)x(1)} \prod_{t=1}^{T} b_{x(t)}(o_t) a_{x(t)x(t+1)}, \qquad (21)$$

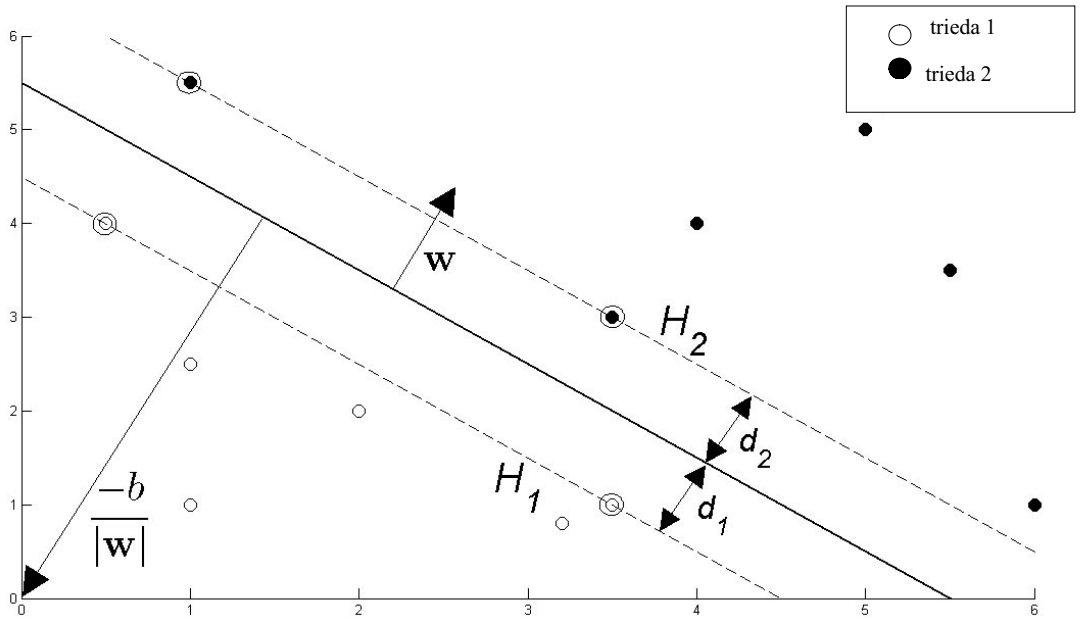
kde $x(0)$ je vstupný stav modelu a $x(T+1)$ je výstupný stav modelu. Tieto stavy negenerujú žiadne pozorovania.

2.1.4. Prístupy klasifikácie pomocou metódy podporných vektorov

Koncom 90. rokov minulého storočia sa znovu objavila metóda klasifikácie lineárne separovateľných dát nazývaná Support vector machines, SVM, *Cristianini, Shawe-Taylor* (2000). Onedlho sa táto metóda zovšeobecnila aj na nelineárne prípady pomocou „kernel triku“. Zároveň sa zistilo, že SVM sa dá použiť nielen na prípady klasifikácie ale aj na, napríklad, lineárnu a nelineárnu regresiu.

V ďalšom majme N učiacich vektorov x_i každý o dimenzii M (môžeme si ich predstaviť ako vektory čŕt - parametrov), pričom každý z nich patri do jednej z dvoch tried, predpokladajme, že vektory sú lineárne separovateľné, to znamená, že existuje hyperrovina v M rozmernom priestore, ktorá oddeľuje uvedené dve triedy, formálne máme N usporiadaných dvojíc

$$\{x_i,y_i\} \qquad i=1,2, ..., N; \; y_i \in \{-1,1\}; \; \mathbf{x} \in R^M$$



*Obr. 8 Support vector Machines v dvoch rozmeroch, H1 a H2 sú úsečky prechádzajúce podpornými vektormi, príkladmi označenými krúžkami. Úsečka SVM optimálna je definovaná parametrami **w** a -b/**w** Pomocou d1= d2 sme označili tzv. SVM okraj, maximálnu medzeru medzi podpornými vektormi*

Pre ilustráciu uvádzame situáciu v dvoch rozmeroch, kde je oddeľujúcou rovinou úsečka, pozri Obr.8. Hyperrovina na Obr.8 sa dá opísať pomocou rovnice

$$\mathbf{w.x} + b = 0 \qquad (22)$$

kde vektor **w** je kolmý na spomínanú rovinu, a - b/|**w**| je kolmá vzdialenosť hyperroviny od počiatku súradnicového systému. Podporné vektory sú takto učiace príklady, ktoré sú najbližšie k hyperrovine oddeľujúcej dané dve triedy a účelom SVM je orientovať túto rovinu tak, aby bola čo najďalej od najbližších členov dvoch tried.

Ak zlúčime podmienky pre naše dve triedy do jednej podmienky, potom platí

$$y_i(\mathbf{w.x} + b) - 1 \geq 0 \qquad \text{pre všetky i.} \qquad (23)$$

Aby sme maximalizovali medzeru medzi podpornými vektormi pre dve triedy, potom musíme minimalizovať |**w**| (veľkosť medzery je 1/|**w**|) za podmienky - väzby (23), čo je ekvivalentné

$$\min \frac{1}{2}|\mathbf{w}|^2 \text{ za podmienky } y_i(\mathbf{w.x} + b) - 1 \geq 0 \qquad \text{pre všetky i.} \qquad (24)$$

Na riešenie takto formulovaného problému použijeme metódu Lagrangeových multiplikátorov

α, kde $\alpha_i \geq 0$ pre všetky i. Potom môžeme prepísať (24) ako

$$L = \frac{1}{2}|w|^2 - \sum_{i=1}^{N} \alpha_i y_i (x_i w + b) + \sum_{i=1}^{N} \alpha_i \qquad (25)$$

Chceme nájsť w a b také, aby (25) bolo minimálne a maximálne z hľadiska α. Kvôli tomu parciálne derivujeme (25) podľa w a b a výslednú deriváciu kladieme za rovnú nule, čo vedie k systémom rovníc

$$w = \sum_{i=1}^{N} \alpha_i y_i x_i$$

a (26)

$$0 = \sum_{i=1}^{N} \alpha_i y_i \quad .$$

Po dosadení (26) do (25) dostaneme tzv. duálnu formu pôvodnej formulácie optimalizačného problému (25)

$$L_D = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i y_i y_j x_i x_j \alpha_j \qquad (27)$$

spolu s väzbovými podmienkami $\alpha_i \geq 0 \forall_i$ a $0 = \sum_{i=1}^{N} \alpha_i y_i$ vedie ku konvexnému maximalizačnému problému z hľadiska α. Pomocou kvadratického programovania sa dajú vypočítať jednotlivé α a pomocou prvej rovnice z (26) vypočítame w. Čo nám zostáva je vypočítať b. Ľubovolný bod, ktorý vyhovuje podmienke podporných vektorov (druhá rovnica v (26)) bude mať tvar

$$y_s (x_s . w + b) = 1$$

čo dosadením do prvej rovnice v (26) dáva

$$y_s (\sum_{m \in S} \alpha_m y_m x_m x_s + b) = 1 \qquad (28)$$

kde S označuje množinu indexov i, ktoré vyhovujú podmienke podporných vektorov, pričom α_i je kladné. Ak vynásobíme (28) y_s dostaneme, uvedomujúc si, že z definície $y_s . y_s$ sa rovná 1,

$$b = y_s - \sum_{m \in S} \alpha_m y_m x_m x_s \quad . \qquad (29)$$

V ďalšom sa zmienime, len z dôvodov úplnosti, o nelineárnom tvare SVM. Mnoho úloh,

ktoré nie sú lineárne separovateľné v priestore vstupných vektorov x sa môže stať separovateľnými vo vyšších rozmeroch, ak sa dá použiť zobrazenie $x \rightarrow \phi(x)$, také, že v novom priestore sú transformované vektory už lineárne separovateľné. Ako môžeme vidieť z (27) Lagrangova funkcia v duálnej forme závisí iba od skalárneho súčinu $x_i x_j$. Trieda funkcií $k(x_i x_j)$, tzv. kernelove funkcie, podobné gaussovským, nazývané tiež Radial Basis Functions, RBF

$$k(x_i, x_j) = e^{-\left(\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)} \tag{30}$$

majú tú základnú vlastnosť, že nemusíme explicitne poznať tvar konkrétnej kernelovej funkcie, aby sme mohli previesť zobrazenie z nižšie rozmerného priestoru vektorov čŕt do vyššie dimenzionálneho priestoru, v ktorom sú už transformované vektory lineárne separovateľné. Matematické požiadavky na to, aby sa daná funkcia, alebo trieda funkcií mohla chápať ako kernelova funkcia presahujú rámec tejto práce.

2.1.5. Prístupy umelých neurónových sietí

Od konca 90. rokov sa hlavný dôraz kladie na skúmanie a zlepšenie metód učenia a rozpoznávania, hlavne pomocou metód ANN, *Lippmann* (1988), (1989). Závery tohto prehľadového článku môžeme zhrnúť do niekoľkých bodov. V prvom rade úspešnosť systémov rozpoznávania reči je nižšia ako u ľudí. Pre malé súbory jednotiek rozpoznávania ako sú dopredu segmentované slova, samohlásky a spoluhlásky sa dá dosiahnuť pomocou viacvrstvových časovo oneskorených sietí ANN úspešnosť porovnateľná s klasickými prístupmi.

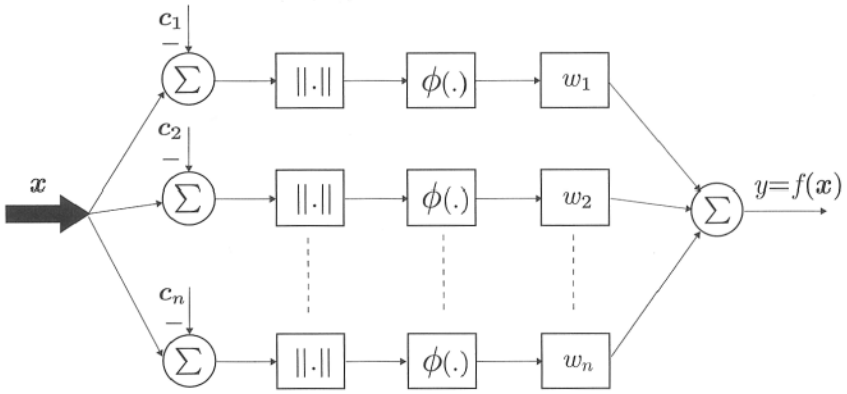
Na prelome 80. a 90. rokov sa vyvinuli metódy učenia viacvrstvových dopredných sietí, spätné šírenie chyby BP - Back Propagation, ktoré sa testovali pri úlohách rozpoznávania číslic a samohlások, kde preukázali úspešnosť rozpoznávania porovnateľnú s klasickými metódami. Vyvinuli sa techniky na spracovanie aj veľkých databáz slov, na redukciu času učenia a na učenie sietí s rekurentnými spojeniami. Jedným z úspechov bola ANN implementácia takých klasických algoritmov ako sú vektorová kvantifikácia a Viterbiho algoritmus, *Ahalt, KrishnaMurthy, Chen, Melton,* (1990), *Rumelhart, Zipser* (1985).

Pre fyziologické predspracovanie reči a kochleárne priepuste filtrov a aj spomínané ANN metódy sa aplikovali metódy mikro analógovej VLSI techniky a tým sa začali vyvíjať kompaktné hardvérové implementácie celých systémov rozpoznávania v jednom integrovanom elektronickom obvode, *Liu, Andreou, Goldstein* (1991), *Lyon, Mead* (1988).

V tomto úvode opíšeme len jeden prístup ANN používaný pri rozpoznávaní reči a hovoriaceho pre úlohy kvalifikácie alebo zhlukovania - RBF - Radial Basis Function. Obecne hovoriac, motiváciou modelovania niektorých funkcií mozgu ako výpočet, je hypotéza, že mozog je vysoko zložitý, nelineárny, adaptívny a paralelný stroj na spracovanie informácie. Ľudský mozog má od narodenia zložitú štruktúru umožňujúcu mu vytvárať „vzťahy, pravidlá“ pomocou „skúseností“. Tento spôsob (dokonca fyziologických zmien) prispôsobovania sa okoliu kulminuje v približne dvoch rokoch, ale pokračuje až do staroby *Hertz, Krogh, Palmer* (1991).

V klasických modeloch ANN sa procesy učenia a klasifikácie odlišujú, prevádzajú sa v dvoch oddelených fázach - učenia a odozvy alebo reakcie neurónovej siete. Avšak, u ľudí oba mechanizmy fungujú súčasne, hoci hlavný učiaci proces nastáva v prvých rokoch života. Takže dospelí používajú už efektívne zostrojený súbor učiacich prototypov, ktorý sa len jemne adjustuje v čase dospelosti.

Medzi množstvom rôznorodých prístupov ANN je prístup RBF jedným z najobvyklejších a zároveň najjednoduchších, *Gupta, Jin, Homma* (2003) . Architektúra tejto siete v svojom základnom tvare pozostáva z troch oddelených vrstiev, pozri Obr.9.



Obr. 9 Základná štruktúra RBF neurónovej siete

Vstupná vrstva opisuje senzorické vstupy, druhá vrstva je skrytá, má vysokú dimenziu. Výstupná vrstva odpovedá odozve siete na daný vstupný vzor. Transformácia zo vstupnej vrstvy na skrytú je nelineárna, naopak transformácia zo skrytej vrstvy na výstupnú je lineárna. Ako vidieť aj z predošlého Obr. 9 výstup y(t) je váhovaným súčtom výstupov zo skrytej vrstvy, a je daný ako

$$y(t) = \sum_{i=1}^{N} w_i \phi(\|u(t) - c_i\|), \tag{30}$$

kde *u(t)* je vstup, *ϕ()* je RBF funkcia, || . || označuje nejakú vzdialenosť, pričom väčšinou sa vyberá ako euklidovská, c_i sú známe centrá RBF funkcií a w_i sú váhy. RBF funkcie sú funkcie so základnými požiadavkami monotónnosti a radiálnej symetrie. Centrum, škála vzdialenosti a presný tvar RBF funkcie sú parametrami model. Najznámejšou a najpoužívanejšou funkciou je gaussovská

$$k(x) = e^{-\left(\frac{\|x-c\|^2}{\sigma^2}\right)} \tag{31}$$

Učenie RBF sa dá rozdeliť do dvoch častí: učenie skrytej vrstvy alebo výber RBF funkcie - je to učenie bez učiteľa, fitovanie výstupov siete na transformované vstupné vektory - to je učenie s učiteľom.

Prvá fáza nepoužíva žiadnu informáciu z výstupov. 1. Vyberie sa typ a počet RBF funkcií. 2.Vyberie sa centrum každej z RBF funkcií. 3. Vyberie sa disperzný parameter σ RBF funkcie (31). V prvom kroku sa väčšinou vyberajú všetky RBF rovnaké. Pre druhý krok, výber centier, sa používa K-means algoritmus: majme n dátových vektorov, úlohou je vybrať k - počet zhlukov - tak, že k < n. Najskôr sa vyberie prvých k vstupných vektorov ako k zhlukov

$$c_i = x_i, \quad i = 1,2,...,k \tag{32}$$

Zostávajúce vektory sa priradia k jednému zhluku

$$c_i = \frac{1}{m_i} \sum_{i \in C_i}^{N} x_i, \qquad 1 \leq i \leq k \tag{33}$$

kde m_i je počet vektorov, ktoré patria do k-teho zhluku. Po ukončení tohto algoritmu zhlukovania sa pristupuje k výberu disperzných parametrov σ. Tieto kontrolujú prekrývanie RBF funkcií a a nepriamo aj zovšeobecniteľnosť tejto siete. Najobecnejšou metódou na ich výber je, že sa vyberajú buď rovnaké pre všetky zhluky, alebo ako stredná vzdialenosť medzi dátami v zhluku a centrom zhluku.

Posledným krokom je učenie s učiteľom, cieľom tohto kroku je fitovať nelineárne transformované dáta v predošlých krokoch pomocou lineárnej funkcie výstupu na žiadané výstupy. Na tento účel sa používajú gradientné metódy, kde je cieľom adaptovať váhy *w* tak, aby sa minimalizoval štvorec chyby

$$e(n) = d(n) - y(n) \tag{34}$$

kde *d(n)* je požadovaný a *y(n)* aktuálny výstup v čase n. Toto vedie ku gradientnej formulácii

$$w(n+1) = w(n) + \eta' \nabla_w E(n) \tag{35}$$

kde *w(n)* sú váhy v čase n, η´ je konštanta rýchlosti učenia, a *E(n)* je štvorec chyby daný ako

$$E(n) = e^2(n) \tag{36}$$

a gradient chyby je daný ako

$$\nabla_w E(n) = 2e(n)\left[\partial y(n)/\partial w_0(n) \ldots \partial y(n)/\partial w_N(n)\right]^T. \tag{37}$$

V prípade ak je aktivačná funkcia lineárna *f(u)=u*, potom

$$y(n) = u(n) = w^T(n)x(n) \tag{38}$$

potom platí

$$\nabla_w E(n) = 2e(n)\left[1\, x_1(n) \ldots x_N(n)\right]^T = 2e(n)x(n) \tag{39}$$

pričom *e(n)* je skalár a *x(n)* je (M+1)x1 vektor, označme η=2η´ a dostaneme

$$w(n+1) = w(n) + \eta e(n).x(n), \tag{40}$$

čo je LMS (Least Mean Square) metóda. Z (40) vidíme, že veľkosť zmeny váh - učenie je úmerné chybe *e(n)* a má smer vstupného vektora *x(n)*, čo je charakteristikou LMS metód učenia resp. hľadania minima chybovej funkcie (36). V nelineárnom prípade aktivačnej funkcie sa gradient chyby

mení na

$$\begin{aligned}\nabla_w E(n) &= -2e(n).\frac{df\,[u(n)]}{du\,(n)}\nabla_w u(n) = -2e(n).\frac{df\,[u(n)]}{du\,(n)}\\ &= -2e(n).f'[u(n)]x(n)\end{aligned} \tag{41}$$

Vidíme, že zmena oproti lineárnemu prípadu je daná $f'[u(n)]$, ak je tento člen relatívne malý voči adaptácii váh - učeniu, potom k učeniu nedochádza. Ináč je táto formula podobná LMS prípadu. V ďalšom predpokladajme lineárny prípad a nezmenené váhy pre časy menšie ako N, pričom N je väčšie ako dimenzia vstupných vektorov. Potom ak definujeme stredný štvorec chyby ako

$$\begin{aligned}E &= \sum_{n=1}^{N} e^2(n) = \sum_{n=1}^{N}\left[d^2(n) - 2d(n)w^T x(n) + w^T x(n)x^T(n)\right.\\ &= \sum_{n=1}^{N} d^2(n) - 2w^T\sum_{n=1}^{N} x(n)d(n) + w^T\left(\sum_{n=1}^{N} x(n)x^T(n)\right)\\ &= \sum_{n=1}^{N} d^2(n) - 2w^T\rho + w^T Rw\end{aligned} \tag{42}$$

kde sme definovali korelačnú maticu ako

$$R = \sum_{n=1}^{N} x(n)x^T(n) \tag{43}$$

a kovariačný vektor

$$\rho = \sum_{n=1}^{N} x(n)d(n)\,. \tag{44}$$

Riešením podmienky minima chybovej funkcie (42) podľa váh *w*, inými slovami

$$\nabla_w E = 0$$

dostaneme LSM riešenie v tvare

$$w_{LS} = R^{-1}\rho \tag{45}$$

Ak môžeme pokladať postupnosť $\{x(n)\}$ za stacionárnu náhodnú postupnosť, potom sa dá ukázať, že v pravdepodobnostnom zmysle konvergencie LMS riešenie w(n) dané (40) konverguje k LS riešeniu (45). Podrobnejšie sa týmto detajlom nebudeme venovať v tomto krátkom prehľade.

Na záver spomenieme len niektoré relatívne výhody RBF - efektívne aproximujú nelineárne zobrazenia, čas učenia je rádove nižší v porovnaní s inými prístupmi ako napríklad MLP s BP, dáva obecne o 5% vyššie presnosti klasifikácie ako napríklad MLP s BP, úspešne identifikujú, kvôli nelineárnej a nemonotónnej funkcii RBF, oblasti dát, ktoré nepatria do nejakého známeho zhluku**,** *Gupta, Jin, Homma* (2003) .

2.1.6. Prechodové javy v časovej oblasti a fáza signálu

Doposiaľ sme sa zaoberali spektrálnymi a energetickými parametrizáciami. Avšak, je známe, že biologický sluchový systém je citlivý nielen na spektrálne a energetické reprezentácie reči, ale aj na rôzne prechodové javy v časovej oblasti, ktoré sú zodpovedné za vysoko nelineárne dynamické procesy sluchového systému. V reálnej konverzácii veľmi zriedkavo percepujeme, izolujeme jedinú hlásku. Naša percepcia hlásky je skôr daná okolím danej hlásky, tento jav koartikulácie alebo citlivosti na kontext musíme uvažovať v každom serióznom modeli, *Torrkola* (1991), *Spitzer, Hochstein* (1985), *Tirakis, Delopoulos, Kollias* (1992).

Podobne sa obecne predpokladá, že krátkodobé fázové spektrum alebo všeobecne fáza rečového signálu hrá malú resp. žiadnu úlohu v percepcii reči alebo pri rozpoznávaní reči počítačom alebo človekom *Helmholtz* (1954), *Yang, Wang, Shamma* (1992). Vo viacerých prácach napr. *Allen, Rabiner* (1977)*, Paliwal* (2004) sa uvádzajú prístupy (aj experimentálne) demonštrujúce, že fázové spektrum s časovým oknom 32 ms a väčším vedie k zrozumiteľnosti reči porovnateľnej s energetickým spektrom alebo kepstrálnymi koeficientami. Dokonca aj výsledky rozpoznávania slov sú porovnateľné s klasickými prístupmi.

2.2. Verifikácia hovoriaceho

Vzhľadom k tomu, že na Slovensku sa úlohám rozpoznávania hovoriaceho venuje málo pracovísk, budeme sa tejto problematike venovať v tejto časti trochu podrobnejšie.

Rečový signál nesie veľa typov informácie o zdroji signálu. Ako v predošlej časti, väčšinou sa pokúšame rozlišovať, klasifikovať jazykovú informáciu, ktorá poskytuje význam zvukom priradeným slovám. Avšak existujú aj iné typy zakódovanej informácie, ktoré nám pomáhajú efektívne komunikovať: identita hovoriaceho, chorobný stav, psychologický stav, vek, pohlavie, jazyk.

Medzi práve uvedenými typmi je informácia o identite hovoriaceho určite najužitočnejšou informáciou, ktorá sa dá prakticky použiť v systémoch umelej inteligencie. Proces dekódovania tejto informácie sa označuje, nazýva viacero spôsobmi: rozpoznávanie, verifikácia, identifikácia hovoriaceho - volajúceho - hlasu; hlasový otlačok, hlasová autentifikácia.

Prehľad systémov rozpoznávania hovoriaceho, ktoré sa realizovali do 80. rokov minulého storočia je uvedený v práci *Flanagana* (1972). Z prehľadu vyplýva jedna zaujímavá skúsenosť, totiž aj pre úlohy rozpoznávania resp. verifikácie sa používajú rovnaké súbory parametrov a techniky analýzy vlastne totožné s úlohami rozoznávania slov. Spolu s úspešnosťou týchto prístupov to znie trochu paradoxne, pretože obe spomínané úlohy sú vzájomne „inverzné“, doplnkové.

Umelé systémy predurčené pre extrakciu informácie závislej na hovoriacom majú všetky jednu spoločnú vlastnosť : bezpečnostný prístup t.j. spoľahlivo verifikovať / identifikovať (nárokovanú) identitu užívateľa. Existuje množstvo už implementovaných aplikácií. Dajú sa rozdeliť na vzdialené (pomocou telefónnych liniek, internetu, atď.) a lokálne (iba mikrofón, a podobne) aplikácie.

Typické vzdialené aplikácie zahrňujú:

- finančné operácie
- platenie kreditnou kartou, telenákup (autorizácia prevodu peňazí)
- telefónne aplikácie overujúce telefónne karty, overujúce medzimestké alebo medzištátne hovory, atď.
- prístup do domu
- vojenské aplikácie
- prístup / modifikácia informácie na vzdialenom serveri (obmedzený prístup pre autorizovaných užívateľov)

- prístup k výpočtovým zdrojom zo vzdialeného terminálu
- prístup na internet, intranet, Interaktívnu káblovú TV

Typické lokálne aplikácie zahrňujú:

- kontrola a prístup k zariadeniu
- prístup k zabezpečenej oblasti (jadrová elektráreň, vojenské zariadenia)
- hlasový klúč (dom, auto)
- mobilné telefóny, osobní asistenti (odozva iba na hlas vlastníka zariadenia)

2.2.1 Taxonómia systémov pre verifikáciu hlasu hovoriaceho

Vlastná extrakcia charakteristík rečového signálu závisiacich na hovoriacom - čŕt sa dá použiť viacerými spôsobmi. Použitie sa môže líšiť viac alebo menej závisiac na dvoch hlavných aspektoch:

- uzavretá vs. otvorená množina hovoriacich
- úroveň nezávislosti od textu

Hlavné úlohy rozpoznávania hovoriaceho sa vzťahujú k uzavretosti množiny hovoriaceho:

- Identifikácia hovoriaceho

Účelom úlohy identifikácie hovoriaceho je klasifikovať neoznačený rečový signál ako patriaci k jednému z množiny *n* odpovedajúcich hovoriacich. Množina odpovedajúcich hovoriacich je uzavretá. Štatistické vyhodnotenie tejto úlohy sa prevádza pomocou množstva zle klasifikovaných hovoriacich.

- Verifikácia hovoriaceho

Účelom úlohy verifikácie hovoriaceho je rozhodnúť, či rečový signál patrí alebo nepatrí požadovanému referenčnému hovoriacemu. Inými slovami má sa rozhodnúť, či akceptovať alebo odmietnuť požadovanú identitu hovoriaceho. Vyžiadanie identity sa dá previesť rôznymi spôsobmi: zadaním osobného hesla, PIN čísla, hlasom (rozpoznaním pomocou rozpoznávania slov) atď.. Množina referenčných hovoriacich sa obyčajne nazýva klienti a je tiež uzavretá (ich počet nie je väčší ako niekoľko stoviek, zvyčajne je ich niekoľko desiatok). Máme tiež k dispozícii množinu hovoriacich, ktorí sa nazývajú provokatéri, ktorá je otvorená. V praxi však aj táto množina je ohraničená, identita hovoriacich je neznáma a preto potenciálne množina provokatérov sa skladá zo zvyšku populácie. Pragmaticky resp. komerčne je táto úloha oveľa dôležitejšia ako identifikácia. Štatistické vyhodnotenie tejto úlohy sa prevádza pomocou dvoch typov chýb:

- *nepravá chyba prijatia* (false positive)- „chyba provokatéra“ : je daná relatívnym počtom prijatých provokatérov.
- *nepravá chyba odmietnutia* (false negative): je daná počtom nesprávne odmietnutých klientov.

Okrem týchto dvoch hlavných úloh sú tiež nasledujúce príbuzné úlohy:

- Detekcia zmeny hovoriaceho

Účelom je určiť či neoznačený rečový signál patrí novému alebo starému, rovnakému hovoriacemu (tomu, ktorý nahovoril predošlý signál).

- Porovnávanie hovoriacich

Účelom je je vybrať hovoriaceho z uzavretej množiny vzorov, ktorý je najbližší ku danému súčasnému hovoriacemu, hoci je dopredu známe, že hovoriaci nie je registrovaný hovoriaci. Zdá sa, že toto je špeciálny prípad výberu zhlukov hovoriaceho, kde sa každý zhluk skladá z jediného hovoriaceho.

- Označovanie hovoriacich

Toto je iba špeciálnym typom identifikácie hovoriaceho počas rozhovoru viacerých ľudí. Účelom je lokalizovať kedy vstupy hovoriacich začínajú a kedy končia.

V závislosti od stupňa závislosti na texte môžeme rozlišovať nasledujúce typy systémov:

· Systémy závislé na texte:

Jazykový materiál, ktorý musí hovoriaci vysloviť je apriórne určený počas registrácie.

· Systémy so spoločným heslom

Jazykový materiál sa nedá vyberať. Zvyčajne, hovoriaci je požiadaný vysloviť niekoľko spoločných hesiel z ohraničeného slovníka.

· Systémy so súkromným heslom

Tieto systémy dovoľujú určitú flexibilitu textu napríklad registrovaný užívateľ má možnosť zmeniť množinu svojich hesiel.

· Systémy nezávislé na texte

Jazykový materiál, ktorý hovoriaci musí vysloviť je viac alebo menej náhodný.

· Systémy s vyzvaním

Tieto systémy predkladajú hovoriacemu jazykový materiál, ktorý sa nedá predpovedať. Postupnosť slov alebo veta je vyžiadaná buď vizuálne (systémy s textovou výzvou) alebo hlasom (systémy s hlasovou výzvou). Zvyčajne zahrňujú určitý druh rozpoznávania reči na overenie toho že sa skutpčne povedalo to čo sa vyžiadalo povedať.

· Systémy s voľným textom

Predpokladajú úplne neobmedzenú reč.

Úspešnosť systému je podstatne ovplyvnená kvalitou rečového signálu. Môžeme rozlíšiť dva hlavné prípady.

· Systémy s čistým širokopásmovým signálom

Využívajú bezšumový, širokopásmový mikrofónny signál bez iných kanálových disturbancií. Najčastejšie sú asociované s lokálnymi aplikáciami.

· Systémy so zašumeným, úzkopásmovým tefónnym signálom

Pracujú so zašumeným, úzkopásmovým tefónnym signálom (t.j. podstatne rušený kanálovými skresleniami). Obyčajne tento typ signálu používajú vzdialené aplikácie.

2.2.2 Obecný pohľad na systémy rozpoznávania hlasu

Systémy Rozpoznávania Hlasu (VRS) reprezentujú inverzný problém k rozpoznávaniu reči (t.j. slov). Pri rozpoznávaní slov je cieľom nájsť také črty, ktoré diskriminujú rôzne jazykové kategórie, jednotky (t.j. hlásky, slová atď.) a tak ako je možné sú invariantné na od hovoriaceho (t.j. nemenia sa signifikantne ak sú vyslovené rôznymi hovoriacimi). Na druhej strane, VRS má účel nájsť také črty, ktoré diskriminujú rôznych hovoriacich a tak ako je možné sú jazykovo invariantné t.j. v najlepšom sa dajú extrahovať s ľubovolného jazykového materiálu.

Intuitívne, črty diskriminujúce hovoriaceho sa dajú začleniť do nasledujúcich dvoch skupín:

• Akustické črty „izolovaných“ zvukov. Odrážajú spektrálne charakteristiky tvaru vokálneho traktu hovoriaceho.

• Časové stopy: rýchlosť reči, dĺžka hovorenia. Odrážajú ”štýl” hovorenia, ktorý je jedinečný pre konkrétneho hovoriaceho.

Systémy rozpoznávania hovoriaceho (t.j. buď identifikácia alebo verifikácia) používajú rovnakú obecnú schému celkového spracovania ako sa používa pri rozpoznávaní reči, Obr 3.

Extrakcia čŕt zvyčajne odpovedá extrakcii čŕt na krátko-dobom okne. Modelovanie prevádza niekoľko funkcií: prevádza separovanie najdôležitejších častí signálu, prinajmenšom implicitne sa zaoberá časovými stopami, prevádza kompresiu dát. Fáza klasifikácie uskutočňuje vlastné štatistické rozhodovanie založené na konečných črtách.

V mnohých systémoch fázy modelovania a klasifikácie sú spojené, vnorené a ťažko sa dajú rozlíšiť. Hlavné prístupy VRS sa dajú rozdeliť podľa nasledujúcich kritérií:

- dlhodobé ustrednenie akustických čŕt (napr. spektrum, základný tón) vs. separácia (segmentácia) dôležitých častí signálu

Dlhodobé ustredňovanie akustických čŕt patrí medzi najstaršie navrhnuté prístupy a je založené na myšlienke „vyhladenia" všetkých iných faktorov (t.j. fonetických variácií) a takto zostanú len zložky závislé na hovoriacom (čo reflektuje určité ustrednenie tvaru vokálneho traktu daného hovoriaceho). Avšak, proces ustrednenia môže tiež znehodnotiť informáciu závislú na hovoriacom a na to, aby sme dosiahli stabilnú štatistiku vyžaduje dlhý signál.

Prístup založený na porovnaní najrelevantnejších častí signálu, ktoré sú náležite separované je založený na myšlienke, že iba konzistené porovnanie podobných fonetických jednotiek môže viesť k najodpovedajúcejšej informácii závislej na hovoriacom.

Existujú dva spôsoby, ktorými sa dá dosiahnuť predošlý prístup:

- explicitná (hlásková) segmentácia vs. implicitná (hlásková) segmentácia.

Pri explicitnej segmentácii sa prevádza segmentácia signálu na určité fonetické jednotky (väčšinou hlásky) pomocou rozpoznávania reči (väčšinou HMM). Značne sa tým ale zvýši výpočtová zložitosť, pretože sa musí najskôr vykonať úloha rozpoznávania reči.

Implicitná segmentácia prevádza istý druh zhlukovania vektorov - čŕt. Táto úloha sa môže previesť rôznymi metódami zhlukovania (t.j. vektorovým kvantovaním). V klasifikačnej časti je to zvyčajne stratégia rozhodovania založená na prístupe K-NN (K- najbližší sused).

Ak máme konečný vzor môžeme použiť dve základné štatistické stratégie ako urobiť konečné rozhodnutie (binárna alebo N-triedna klasifikácia):

- nediskriminatívne modelovanie akustických čŕt vs. diskriminatívne modely

Nediskriminatívny prístup vytvára pre každého hovoriaceho pravdepodobnostný model nezávisle od ostatných hovoriacich. Na druhej strane, diskriminatívny prístup vytvára jeden spoločný model, ktorý sa potom učí ako čo najviac zväčšiť diskriminačnú schopnosť medzi rôznymi hovoriacimi.

Predspracovanie - (krátkodobá) extrakcia čŕt

Hlavnou a obecnou myšlienkou, ktorá je za extrakciou primárnych krátkodobých čŕt je, že spektrálne črty vystihujú štruktúru vokálneho traktu konkrétneho hovoriaceho a preto môžu dobre charakterizovať črty závisiace od hovoriaceho.

Celková schéma je podobná schéme, ktorá sa používa v rozpoznávaní reči, pozri Obr.1.

Prvý blok spracovania reprezentuje všetky základné operácie - oknovanie, priepusťové filtrovanie, ktorých účel je ponechať pre ďalšie spracovanie iba najužitočnejšiu časť signálu.

V druhom bloku sa v minulosti používali hlavne charakteristiky založené na LPC (kepstrálne alebo odrazové koeficienty), ale nedávne štúdie ukázali, že posledne menované sú veľmi citlivé na signálové šumy. Robustnejšie črty sú založené na určitých reprezentáciach blokov filtrov *Reynold, Rose* (1995) a techniky rovnaké ako pri rozpoznávaní slov PLP, MFC**,** *Reynold* (2002).

Nasledovný blok reprezentuje konečnú krátkodobú reprezentáciu, zatiaľ čo kepstrálna reprezentácia sa dá získať buď z LPC alebo z výstupu bloku filtrov. Pre tieto kepstrálne alebo energetické parametre sa používajú rôzne doplnkové transformácie: časové derivácie, rasta filtrovanie *Hermansky, Morgan, Bayya, Kohn* (1991, alebo reprezentácie signálového rezidúa (pozri diskusiu nižšie) alebo reprezentácia pomocou kepstrálneho kódovania cez Mel frekvencie MFC, atď..

Zvyčajne sa predošlé krátkodobé črty extrahujú homogénne, t.j. každých 20 ms, obecne prekrývajúco. Hlavným cieľom rôznych modifikácií je nájsť najreprezentatívnejšiu krátkodobú charakteristiku štruktúry vokálneho traktu. Aj keď došlo v posledných rokoch k evidentnému pokroku v tomto smere, zostávajú nasledovné neriešené obecné problémy a nedostatky súčasných prístupov:

- Nie optimálna extrakcia časových charakteristík: rýchlosti reči, trvanie reči.

Neexistuje jasná koncepcia ako správne vyjadriť tieto časové charakteristiky konkrétneho hovoriaceho. Explicitná segmentácia modeluje oddelené fonetické jednotky (t.j. hlásky resp. fonémy) a ľubovolný typ extrakcie čŕt sa prevádza v oknách, ktoré korešpondujú premenlivým oblastiam týchto jednotiek. Takýmto spôsobom získané reprezentácie sú skôr invariantné k časovým stopám, pretože tu dochádza k určitému druhu časovej normalizácie. Toto je v súlade so štúdiami *Matsui, Furui* (1991), *Kao, Rajasekaran, Baras* (1992), v ktorých rozpoznávanie reči umiestnené do začiatku spracovania neposkytuje významné zlepšenie výkonu SRS - Systémov Rozpoznávania Hovoriaceho.

Pri modeloch implicitnej segmentácie vektory čŕt sú extrahované z okien rovnakej veľkosti. Neexistuje explicitný mechanizmus na extrahovanie "časovej" informácie, ale kvôli rovnakej dĺžke okien "časové " charakteristiky sú implicitlne zahrnuté do modelu, pretože rôzne fonetické javy sú rozdelené vnútri okien s rovnakým podielom pre každého konkrétneho hovoriaceho.

V tomto duchu je tiež otáznou potenciálna užitočnosť nehomogénnych a "lokalizovaných" časovo-frekvenčných reprezentácií signálu (napr. Gáborových, Wienerových, wavelet transformácií), ktoré dávajú optimálnejšie pokrytie časovo - frekvenčnej roviny, ale skôr z hľadiska invariantnosti voči hovoriacemu

- Nejasná „lokalizácia" parametrov, ktoré závisia na hovoriacom.
- Dekompozícia signálu z hľadiska rozpoznávania reči vs. rozpoznávania hovoriaceho.

Je zaujímavým faktom, že primárne črty používané v SRS sú rovnaké alebo veľmi podobné črtám, ktoré sa používajú v rozpoznávaní reči. Ak zoberieme do úvahy vzájomne sa vylučujúcu inverznú povahu oboch úloh, je tento fakt veľmi podivný až zarážajúci.

- Analýza LPC-rezidúí

V roku 1995 bola prevedená analýza LPC-rezidúí *Thevenaz, Hügli* (1995), pre účely SRS. Autori vyšetrovali užitočnosť čŕt, ktoré namiesto koeficientov filtra syntézy používajú, koeficienty LPC rezidúí:

$$\mathbf{u}(n) = \frac{1}{G}\left(\mathbf{s}(n) - \sum_{k=1}^{\min(p,\,n)} \mathbf{a}_k \mathbf{s}(n-k)\right) \tag{46}$$

Amplitúdové spektrum týchto koeficientov

$$\left|\mathbf{U}(k)\right| = \left|\sum_{n=0}^{N-1} \mathbf{u}(n)\, e^{-i2\Pi nk\,/\,N}\right|$$

sa potom použilo na získanie **LPC**-reziiduálneho reálneho kepstra

$$\mathbf{v}(n) = \frac{1}{N}\sum_{k=0}^{N-1} \ln\left|\mathbf{U}(k)\right| e^{i2\Pi nk\,/\,N} \tag{47}$$

Výsledky SRS s takýmto kepstrom, založeným na LPC-rezidúach, boli iba o trochu horšie

než výsledky založené na pôvodných (zvyčajne používaných) koeficientoch filtra syntézy . Toto je jasným prejavom toho, že informácia závislá na hovoriacom nie je dobre lokalizovaná v LPC spektre. Pretože aj druhé obvyklé primárne črty používané pre SRS sú podobné, tento fakt nás smeruje k záveru, že črty závislé od hovoriaceho v rámci LPC dekompozície (a pravdepodobne aj v rámci Fourier spektier) majú **"**silne distribuovanú povahu**".**

Fáza modelovania

Cieľ fázy modelovania v SRS je dvojaký.

1) Uskutočňuje ďalšie transformácie krátkodobých čŕt. Napríklad, v prípade ak sa pre SRS používa slovo alebo menšia jednotka treba skonštruovať z odpovedajúcich krátkodobých čŕt odpovedajúce vektory pre slová alebo menšie jednotky. Takýmto spôsobom sa prevádza priama kompresia dát.. Napríklad, ak použijeme menšie jazykové jednotky ako sú hlásky, dajú sa rôznym hláskam priradiť váhy podľa ich významu z hľadiska SRS, *Liou, Mammone* (1995). Potom, optimalizácia týchto váh môže byť časťou modelu učiaceho sa procesu.

2) Vytvára pravdepodobnostný model pomocou parametrov adaptácie alebo učenia, ktoré potom poskytnú nástroje pre štádium SRS vykonávajúce konečné štatistické rozhodnutia.

Pretože väčšina súčasných prístupov používa krátkodobé črty pre štádium konečného rozhodovania, sústredime sa v ďalšom iba na diskusiu o metódach vytvárania modelov.

Pre SRS sa používajú oba typy pravdepodobnostných metód:

- neparametrické metódy - *model odhadu funkcie hustoty pravdepodobnosti* (PDF)
- parametrické metódy, ktoré sa dajú rozdeliť na ďalšie dve skupiny:
- nediskriminačné metódy - *modely odhadu PDF*
- diskriminačné metódy - *modely odhadov tried diskriminačnej funkcie*

Všetky typy predošlých modelov sa dajú použiť na implicitné aj explicitné modely segmentácie pre SRS, avšak, historicky neparametrické metódy sa aplikujú hlavne na modely implicitnej segmentácie.

Neparametrické metódy

Tieto metódy odhadujú PDF vektorov čŕt , ktoré patria do triedy konkrétneho hovoriaceho. Obecne, (a zvyčajne) krátkodobé vektory čŕt sa zhlukujú bez faktora učiteľa pomocou zhlukovania založenom na nejakých vzoroch ako sú K-MEANS, ISODATA, zhlukovanie s vedením alebo vektorovým kvantovaním. Pre ilustrovanie sa stručne zmienime o učení s vektorovým kvantovaním *Kohonen* (1988), ktoré sa použilo v *Bennanni, Fogelman, Gallinari* (1990), (1995), a poskytuje ANN alternatívu k predošlým metódam. Referenčné - „codebook" vektory pre daného hovoriaceho sú menené (učené) podľa nasledujúceho pravidla:

$$\Delta \mathbf{W}_k(t+1) = h_k(t)\left(\mathbf{x}(t) - \mathbf{W}_k\right) \qquad k \in N\ (k_{win}) \qquad (48)$$

kde k je súbor indexov, ktoré korešpondujú s referenčnými vektormi patriacimi do určitého okolia vyhrávajúceho (najbližšieho) referenčného vektora vzhľadom ku vstupnému vektoru čŕt **x**(t)

$$k_{win} = \arg \min_{1 \le k \le S} \left\| \mathbf{x(t)} - \mathbf{W_i} \right\| \qquad (49)$$

Podstatný rozdiel v porovnaní s inými metódami vektorovej kvantizácie je, že pri každom kroku učenia sa mení viac referenčných vektorov (z okolia, susedstva vyhrávajúceho vektora).

Učiaci parameter $h^i(t)$ kontroluje (zmenšujúcu sa) veľkosť susedstva vyhrávajúceho neurónu a (klesajúcu) učiacu rýchlosť počas učiaceho procesu.

Fáza testovania sa uskutočňuje jednoduchým spôsobom:

1. každý vektor krátkodobých čŕt sa klasifikuje podľa označenia triedy najbližšieho (vyhrávajúceho) referenčného vektora (t.j. klasifikácia podľa najbližšieho suseda).
2. Celkové rozhodnutie pre slovo alebo vetu (y) sa zakladá na väčšinovom hlasovaní cez celý signál. Rozpoznaný hovoriaci (alebo abstraktne trieda "provokatéra") sa určí na základe triedy referenčných vektorov, ktoré cez celý signál najčastejšie vyhrávajú.

Doplnkovým cieľom (k odhadu PDF) neparametrických metód je kompresia dát - t.j. namiesto použitia všetkých učiacich vektorov daného hovoriaceho v testovacej fáze, sa môže použiť iba (oveľa) menší počet referenčných vektorov.

2.2.2.1 Nediskriminačné parametrické metódy

Myšlienka, ktorá je za aplikáciou týchto metód na SRS je veľmi jednoduchá, priamočiara:

1. Učiaca fáza: Odhadne sa parametrická PDF vektorov čŕt pre každého hovoriaceho (z množiny učiacich vektorov). Uskutočňuje sa to adaptovaním parametrov PDF (optimalizáciou).
2. Testovacia fáza: Vypočíta sa pravdepodobnosť toho, či neznámy vektor čŕt patrí do triedy konkrétneho hovoriaceho. Rozpoznaný hovoriaci (alebo provokatér) sa určí podľa triedy s maximálnou pravdepodobnosťou.

Najpoužívanejší a najobecnejčí typ parametrických PDF modelov , ktoré sa používajú v SRS je *Gaussov* zmiešaný model (GMM) *Reynolds, Rose* (1995).

Každy hovoriaci $\boldsymbol{\lambda}$ sa reprezentuje svojím vlastným modelom funkcie hustoty pravdepodobnosti (PDF) z generujúceho vektora čŕt $\mathbf{x}(t)$:

$$p(\mathbf{x}(t)|\boldsymbol{\lambda}) = \sum_{i=1}^{M} p_i b_i(\mathbf{x}(t)), \tag{50}$$

kde p_i sú reálne parametre a $b_i(\mathbf{x})$ sú gaussovské bázové funkcie viacerých premenných:

$$b_i(\mathbf{x}) = \frac{1}{(2\Pi)^{D/2}\left|\Sigma_i\right|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mathbf{m}_i)'\,\Sigma_i^{-1}(\mathbf{x}-\mathbf{m}_i)\right\}$$

V poslednom vzťahu vektory stredných hodnôt $\mathbf{m}_i$ a kovariančné matice Σ_i reprezentujú ďalšie parametre v modeli, ktoré treba optimalizovať. Celková pravdepodobnosť pre postupnosť $\mathbf{X}$ krátkodobých vektorov čŕt (ktoré generujú slovo alebo vetu (y)), a ktoré sú generované hovoriacim $\boldsymbol{\lambda}$ je daná ako

$$p(\mathbf{X}|\boldsymbol{\lambda}) = \prod_{t=1}^{T} p(\mathbf{x}(t)|\boldsymbol{\lambda}) \tag{51}$$

Konečné rozhodnutie o hovoriacom je prevedené nájdením maxima nasledujúceho výrazu cez všetkých hovoriacich

$$k_{ident} = \arg \max_{1\le k\le S} \sum_{t=1}^{T} \log\ p(\mathbf{x(t)}|\boldsymbol{\lambda}_k) \tag{52}$$

Dané parametre môžeme optimalizovať pomocou rôznych optimalizačných techník. Dobre zavedenou je algoritmus maximalizácie pravdepodobnosti, *Reynolds, Rose* (1995).

GMM model je všeobecným rámcom, ktorý poskytuje veľa voľnosti pri návrhu PDF rôznymi spôsobmi:

- Môžeme vybrať obecné alebo špecifické (RBF neurónová sieť) bázové funkcie pre každého hovoriaceho
- Dajú sa vybrať nasledujúce kovariančné matice S
  1) špecifické pre uzol
  2) jednu pre model hovoriaceho
  3) jednu pre všetkých hovoriacich

V posledných rokoch sa v SRS často používali , inšpirované rozpoznávaním reči, HMM modely. Treba poznamenať, že vo väčšine prípadov tieto HMM boli ekvivalentné s GMM ako sme ich opísali v predošlých častiach, pretože tieto sa dajú chápať ako jednostavové HMM s Gaussovskou zmiešanou hustotou alebo ako ergodické Gaussovské pozorovanie HMM s fixnými, rovnakými prechodovými pravdepodobnosťami. HMM modely sa používajú v rozpoznávaní reči, pre ich schopnosť modelovať izolované zvuky reči spolu s časovým usporiadaním medzi týmito zvukmi. Toto môže byť výhodným pri textovo závislom SRS, avšak pre textovo nezávislé SRS usporiadanie zvukov v trénujúcich dátach neodráža nutne zvukové postupnosti aké sa vyskytujú v testovacích dátach. Toto sa podporilo v *Tishby* (1991), *Matsui, Furui* (1992), kde sa zistilo, že výkon textovo nezávislých SRS nie je ovplyvnený zrušením prechodových pravdepodobností v HMM modeloch pre hovoriacich. Na druhej strane sa používajú viacstavové HMM v kontexte takzvaných predikčných modelov, ktoré sú založené na schopnosti predpovedať nejaký časový rámec v čase t z kontextu iných rámcov, *Stern, Lasry* (1987). Podobne, ako v predošlých prístupoch, prechody medzi stavmi boli fixované.

2.2.2.2 Diskriminačné parametrické metódy

Hlavnou myšlienkou, ktorá leží za diskriminačnými parametrickými metódami je, že sa trénuje rozhodovacia (diskriminačná) funkcia, ktorá najlepšie diskriminuje hovoriacich, a netrénujú sa nezávisle individuálne PDF modely hovoriacich. Najznámejšie sú reprezentácie neurónových sietí: Viacvrstvové dopredné neurónové siete, modely neurónových sietí RBF, hoci sa tiež urobilo niekoľko návrhov diskriminačných HMM alebo konekcionistických implementácií HMM, *Forsyth* (1995), *Bridle* (1990), *Naik, Lubenski* (1992). Existujú dva základné typy modelov diskriminačných neurónových sietí:

I. Viacvrstvové dopredné NN - ktoré využívajú globálne nosné bázové funkcie

$$y_k = f\Big(\sum_i w_{ki} f\Big(\sum_j w_{ij} x_j\Big)\Big) \quad f = \frac{e^{kx} - e^{-kx}}{e^{kx} + e^{-kx}} \tag{53}$$

II. Radiálne bázové funkcie - ktoré využívajú lokálne nosné bázové funkcie.

$$y_k = f\Big(\sum_i w_{ki} b_i(x)\Big) \qquad b_i(\mathbf{x}) = \frac{1}{(2\Pi)^{D/2}\left|\Sigma_i\right|^{1/2}} \exp\Big\{-\frac{1}{2}(\mathbf{x}-\mathbf{m}_i)'\,\Sigma_i^{-1}(\mathbf{x}-\mathbf{m}_i)\Big\} \tag{54}$$

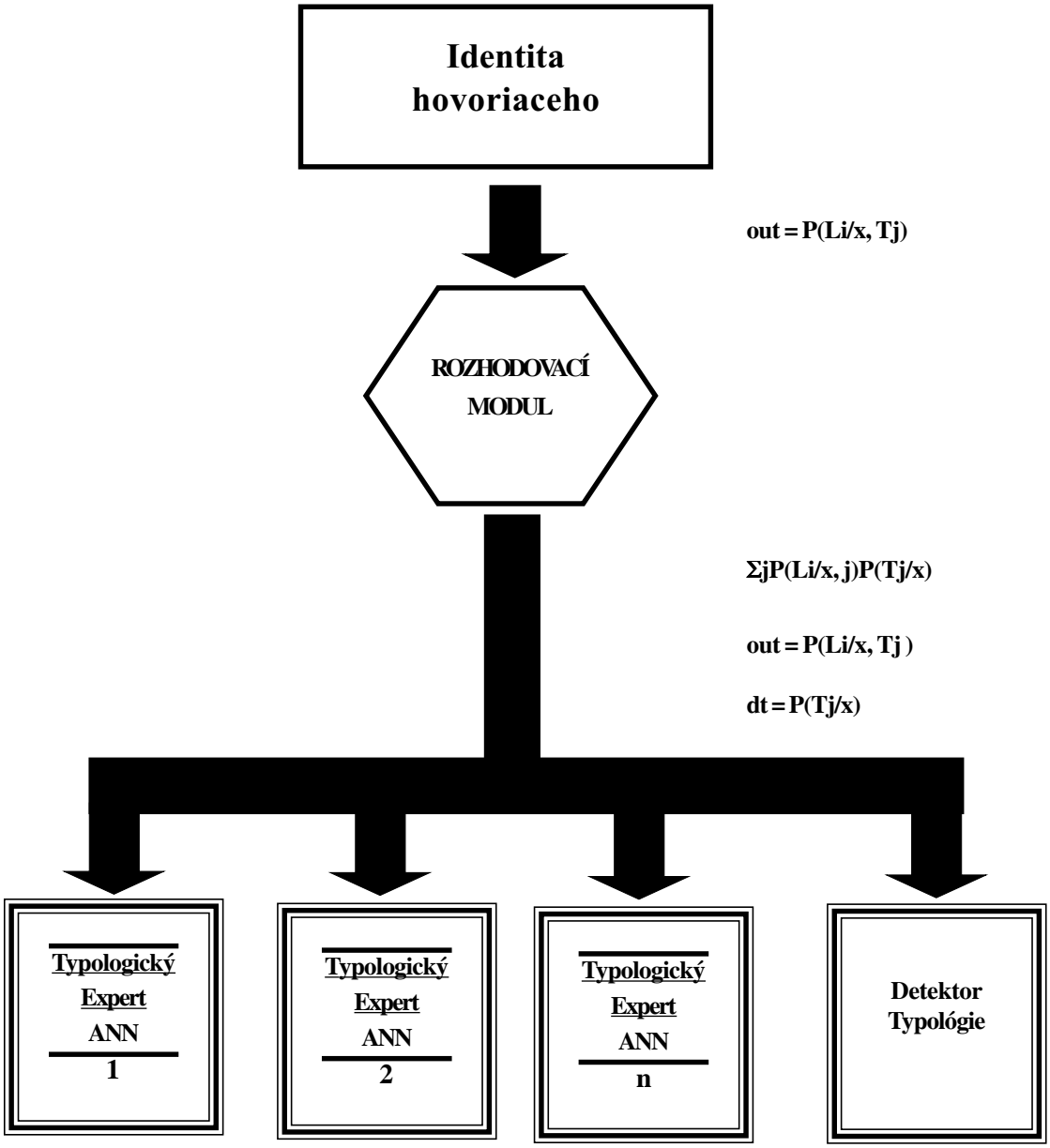
Hlavným obmedzením týchto metód je výpočtová zložitosť diskriminatívneho učenia, inými slovami globálny prístup nutne zahrňuje všetkých hovoriacich a čas trénovania narastá oveľa rýchlejšie ako lineárne s počtom hovoriacich. Spomenieme dve pozoruhodné riešenia, koré boli navrhnuté na zefektívnenie tohto problému, *Girosi, Poggio* (1990), *Park, Sandberg* (1991), *Niranjan, Fallside* (1988).

• Namiesto jednej globálnej diskriminatívnej siete pre všetkých hovoriacich môžeme skonštruovať pre každého hovoriaceho jednu neurónovú sieť, ktorá diskriminuje medzi triedou konkrétneho hovoriaceho a triedou provokatéra modelovanou pomocou zvyšku hovoriacich *Oglesby, Mason* (1991).

• Podobná myšlienka k predošlej je skonštruovať diskriminatívnu neurónovú sieť pre nejakú typológiu hovoriaceho *Bennanni, Gallinari* (1992), (1995).

Vstupný signál sa najskôr vedie do detektoru typológie, v ktorom sa použije zhlukovanie pomocou K-MEANS klusterizácie na rozdelenie populácie do typologických podskupín. Po tejto klasifikácii podľa určitej typológie, sa vektorý čŕt klasifikujú pomocou diskriminujúcich Typologických Expertných ANN založených na trojvrstvovej doprednej časovo oneskorenej siete (TDNN). Softvérová verzia priraďuje každej typológii pravdepodobnostný faktor a signál sa spracováva každým Typologickým Expertom, zatiaľčo výstup je ováhovaný úmerne práve spomínanému pravdepodobnostnému faktoru.

Podobné metódy s typologickou stratégiou boli navrhnuté v *Artieres, Bennanni, Gallinari, Montacie* (1991), *Bennanni, Gallinari* (1995), kde sa kombinuje TDNN v prvej vrstve s HMM v nasledujúcej vrstve.



Obr. 10 Všeobecná schéma pre typologický prístup k SRS

Okrem týchto prístupov boli navrhnuté rôzne hybridné systémy, v ktorých sa kombinujú neuronálne siete s HMM (pozri *Bennanni, Gallinari* (1992) na prehľadnú informáciu).

2.2.2.3 Závery

Naším cieľom v predošlom nebolo prezentovať nejaké čísla, ktoré by ilustrovali výkonnosť / presnosť diskutovaných metód, *Reddy, Ermn, Neely* (1973) . Aj keď mnohé z predošlých prístupov sa testovali na obecných verejne prístupných rečových databázach (TIMIT, NTIMIT, NET-TALK, SWITCHBOARD, YOHO), *Sejnowski, Rosenberg* (1986), *Robinson, Fallside* (1990), je často ťažké ich porovnávať kvôli rôznym štatistickým a technickým podmienkam, napríklad:

- rôzne trénovacie a / alebo testovacie časy
- v rámci úlohy verifikácie existujú dva typy chýb (chybné vylúčenie a chybné prijatie) ktoré sú vzájomne balansované (nárast jedného je na úkor poklesu druhého) zložitým spôsobom. Preto je ťažké priamo porovnávať obidve metódy. S veľkosťou chyby sa môžu zbalansovať aj iné kritéria výkonnosti:
- rozmer reprezentácie čŕt (dôležitý faktor pri prenose takýchto reprezentácií cez sieť)
- učiaca a testovacia doba metódy
- prenositeľnosť systému

Uvedieme teraz približné veľkosti chýb, alebo úspešnosť, ktoré sa podstatne odlišujú pre experimenty s off-line databázami a testovaním v reálnom čase:

- off-line databázy ~ 99%
- testovanie v reálnom čase ~ (90 - 95)% ?

Takisto existuje značný rozdiel, približne 5% a viac, medzi presnosťou čistých, systémov so širokopásmovými mikrofónami a systémami s telefónnou (úzkopásmovou) kvalitou.

Predošlý rozdiel v presnosti sa dá prisúdiť nasledujúcim faktorom, ktoré sú obyčajne inherentné pre komerčné systémy:

- telefónna kvalita reči
- testovanie v reálnom čase vs. výsledky off-line testovania
- dlhodobé variácie v charakteristikách hovoriaceho
- psychologické (zdravotné) variácie v reči
- konkurenčný (zákaznícky) tlak na malé doby trénovania a testovania
- (štatisticky) veľké databázy

Nakoniec, zdá sa, že zostávajú nasledujúce otvorené problémy v SRS, ktoré treba ešte podrobnejšie vyšetriť, skúmať:

- neextrahovanie časových charakteristík: rýchlosti a časové trvanie
- všetky okná (javy) hrajú rovnakú úlohu (už zapracované napríklad pomocou metódy vyhodnocovania fonetického váhovania, *Bennani, Fogelman, Gallinari* (1990))
- správne narábanie s informáciou „čas - frekvencia“ - je potreba alebo nutnosť pre nový typ extrakcie čŕt ?

3 Analýza rečového signálu

V tejto časti sa sústredime na fyziologický popis a analýzu niektorých dôležitých fyzikálnych, akustických a fyziologických faktov a javov z oblasti percepcie ľudskej reči. Definujeme a analyzujeme z pohľadu percepcie spektrálne a časové charakteristiky reči. Sústredíme sa na časť analýzy danú len percepciou a nie formovaním reči. Podobne sa sústredime len na informačne nie energeticky danú analýzu. Aj keď, ako sme videli v úvodných častiach, pri rozpoznávaní slov a verifikácii hovoriaceho hrá dôležitú úlohu aj analýza a závery z formovania reči.

Všetky tieto fakty použijeme neskôr pri modelovaní a počítačovej simulácii procesov rozpoznávania slov a verifikácie hovoriaceho a pochopení percepcie.

3.1 Fyziologické základy percepcie reči

V prvej časti tejto podkapitoly opíšeme niektoré základné fakty o ľudských mechanizmoch rečovej percepcie, *Kaiser* (1957, *Flanagan* (1972), *O´Shaughnessy* (1990). Spomenieme aj niekoľko faktov o percepcii zvukov resp. reči papagájom. Na záver urobíme všeobecné poznámky k abstraktnej problematike prercepcie, ktoré dáme do súvisu s niektorými abstraktnými myšlienkami spracovania informácie, aby sme videli pozadie nášho uvažovania v nasledujúcich kapitolách. Z hľadiska rozpoznávania reči a verifikácie hovoriaceho sa zaujímame hlavne o vzťah medzi zvukovými vlnami vo vzduchu, v okolí vonkajšieho sluchového orgánu a nervovými impulzami v sluchovom nerve. Opíšeme tento vzťah v niekoľkých bodoch.

3.1.1 Percepcia človeka

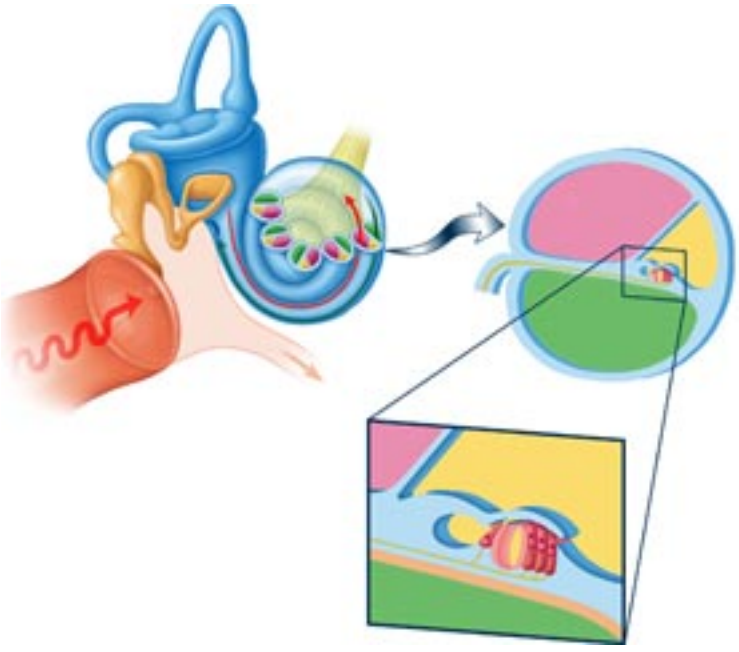
Vonkajšie ucho má funkciu koncentrovať zvukovú energiu čo najefektívnejším spôsobom - zvukovodom s exponenciálnym tvarom, Obr.11.

Obr.11 Systém ľuského ucha - vonkajšie ucho, stredné a vnútorné so slimákom

Najdôležitejšou funkciou stredného ucha je impedančné prispôsobenie zvuku šíriaceho sa vo vzduchu, v prostredí vonkajšieho ucha, k zvuku šíriacemu sa v kvapaline vnútorného ucha. Z množstva meraní prenosových charakteristík stredného ucha, *von Bekesy* (1960), môžeme usúdiť, že prenosová charakteristika stredného ucha je dolná priepusť - vysoko frekvenčný filter so zrezávacou frekvenciou 3000 Hz. Presná efektívna zrezávacia frekvencia a strmosť frekvenčnej charakteristiky práve spomínaného filtra nehrajú v percepcii reči významnejšiu úlohu, Obr. 12.

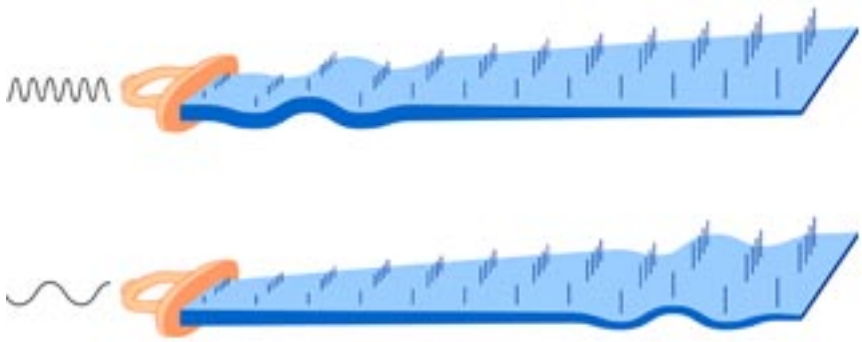
Obr. 12 Stredné a vnútorné ucho

Vnútorné ucho alebo slimák (cochlea) je vyplnený tekutinou a pomocou niekoľkých membrán je rozdelený do niekoľkých dutín Jedna z týchto membrán tzv. bazilárna membrána je časťou orgánu Corti, najdôležitejšieho orgánu počutia, Obr.13. Zvukové vlny prechádzajú pozdĺž bazilárnej membrány, ktorá má rezonančné frekvencie od 20Hz do 20 kHz.



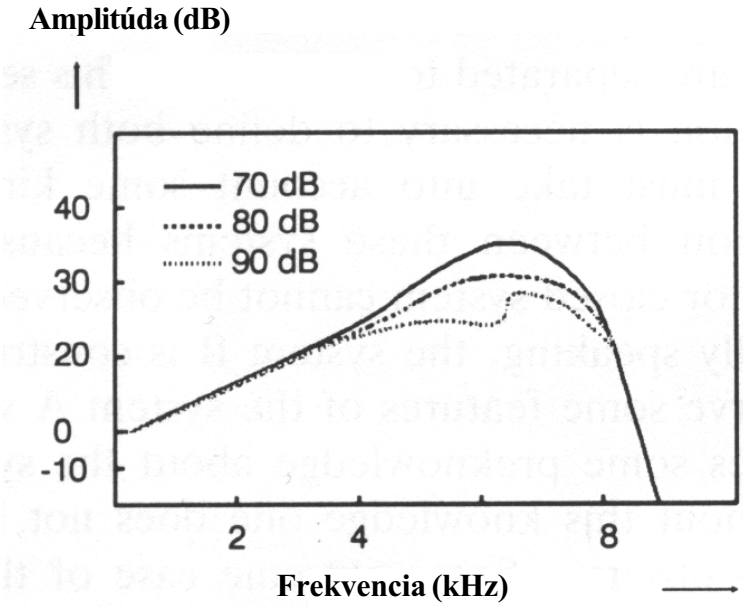
Obr.13 Vnútorné ucho a orgán Cortiho

Približne takýmto spôsobom prevádza bazilárna membrána frekvenčnú analýzu zvukov. Psychoakustické experimenty ukazujú, že frekvenčné rozlíšenie je približne 3 Hz na úrovni 1 KHz, Obr. 14a.



Obr.14a Bazilárna membrána a jej odozva na zvuky rôznej frekvencie

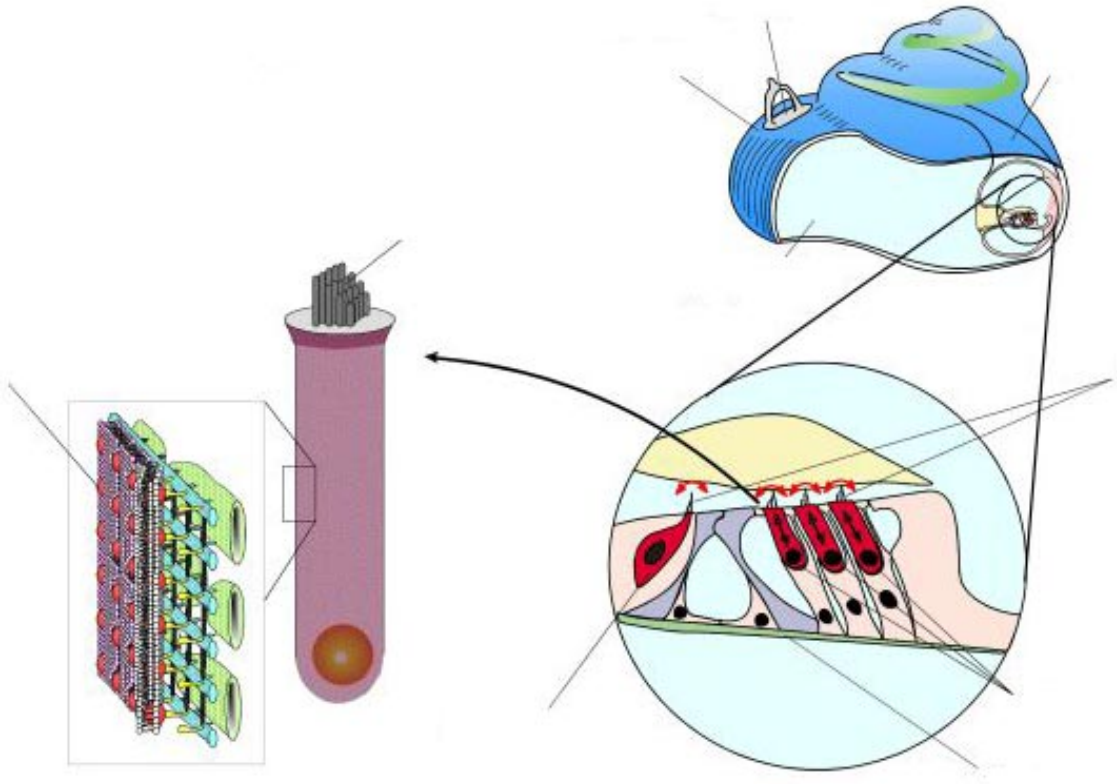
Prechodová charakteristika bazilárnej membrány nameraná pomocou laserovej techniky je na Obr. 14.b pre rôzne úrovne zvukového tlaku *Rhode* (1971). Rôzne relatívne amplitúdy v blízkosti rezonančnej frekvencie implikujú nelinearitu v pohyboch bazilárnej membrány.



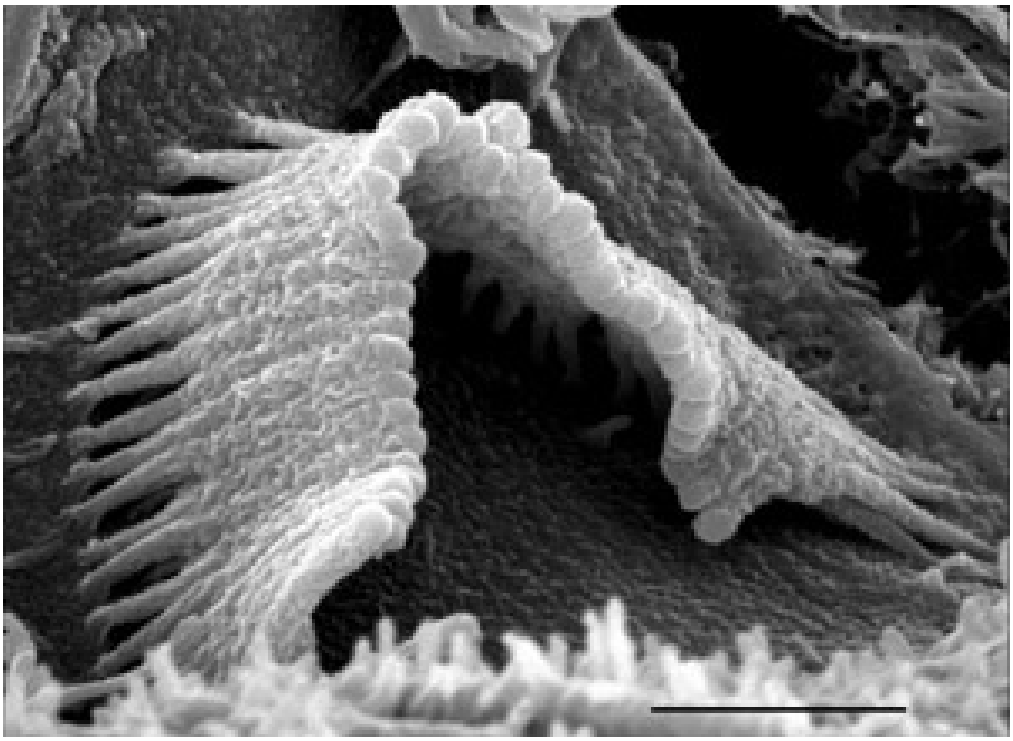
Obr.14b Bazilárna membrána a jej odozva na zvuky rôznej frekvencie, z Rhode (1971)

Najdôležitejším prvkom *Cortiho* orgánu sú vlasové bunky, ktoré sú umiestnené na vibrujúcej časti slimáka, Obr. 15a, b. Transformujú svoje mechanické pohyby na elektrický potenciál - kochleárny mikrofónny potencál, CMP. Existujú dva druhy vlasových buniek, vonkajšie a vnútorné vlasové bunky, každý druh s inou funkciou. Vonkajšie vlasové bunky sú citlivejšie ako vnútorné, je to zapríčinené ich polohou. Počet vlasových buniek v ľudskom uchu je okolo 25 000.

Vonkajšie vlasové bunky slúžia ako adaptívna spätná väzba na regulovanie bazilárnej membrány, jej jemnej frekvenčnej selektívnej citlivosti a ochrane voči vysokým intenzitám signálu a šumu. Ich funkcia nie je v prenose rečovej alebo obecne zvukovej informácie, svedčia o tom experimenty na živých zvieratách, hlavne mačkách a papagájoch s použitím ultra tenkých elektród, *Meddis* (1986), *O´Shaughnessy* (1986). Takto nám zostáva približne 3000 vnútorných vlasových buniek, každá s priemerne 10 synaptickými ukončeniami schématicky zobrazenými na Obr. 16.

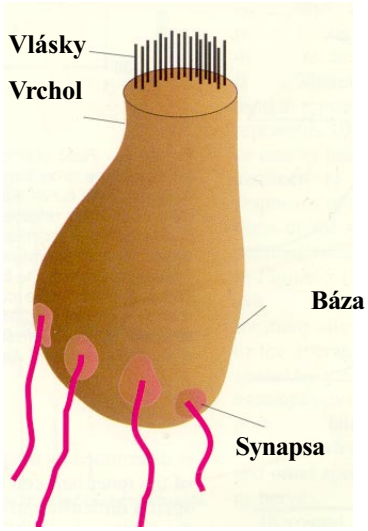


Obr. 15a Orgán Cortiho a vlasové bunky

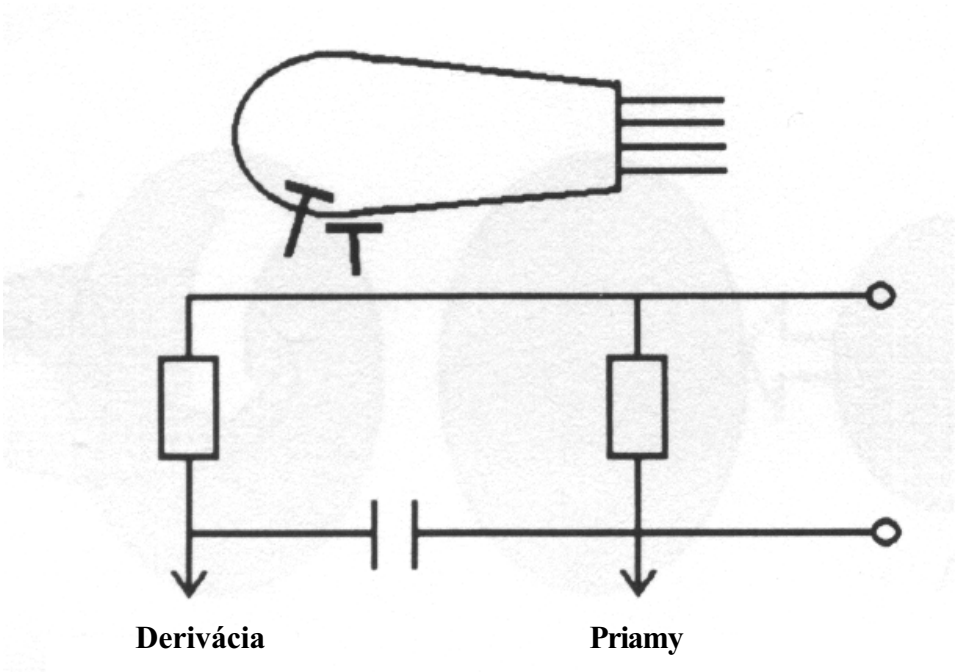


Obr. 15b Vlasové bunky - reálna fotografia, z Strube(1985)

Experimentálne štúdie ukazujú, *Pickles* (1988), *Russell, Nilsen* (1997), že sú dva druhy nervových ukončení. Niektoré nervy končia ako jednoduché tlačítka, vnútri bunky, zatiaľčo iné vytvárajú zhluky na membráne, ale neprenikajú cez membránu bunky. Predpokladá sa, že ukončenia prenikajúce dovnútra cez membránu majú priamy dosah na CMP, zatiaľčo druhý typ ukončení má prístup k potenciálu prostredníctvom kapacitancie, ktorá je vytvorená membránou vlasovej bunky a membránou nervového ukončenia, čiže medzimembránovou štrbinou. Matematicky, technicky hovoriac, môžeme si predstaviť, že výstup prvého typu je približne deriváciou kochleárneho potenciálu a výstup druhého typu je priama kópia tohto potenciálu. Preto existuje opodstatnené presvedčenie, že vlasové bunky účinkujú, v prvom priblížení, ako lineárne mikrofónne zosilňovače, Obr. 17.



Obr. 16 Schématická vlasová bunka so synaptickými ukončeniami



Obr.17 Dva typy nervových ukončení alebo dva typy vlasových buniek

Spôsob akým mikrofónny potenciál vlasovej bunky stimuluje vlákna sluchového nervu je vysoko nelineárny, pretože nervy môžu prenášať impulzy povahy „všetko alebo nič“. Elektrické impulzy, nazývané „spajky“ majú v akustickom nerve dĺžku trvania približne 0,2 ms. Môžu sa aktivovať aj bez akustickej stimulácie, spontánne. Typická rýchlosť emisie týchto spontánnych impulzov je okolo 50 impulzov za sekundu. Pri akustickej stimulácii sa môže táto rýchlosť zvýšiť až na 150 impulzov za sekundu, a počas veľmi silnej stimulácie sa môže táto rýchlosť zvýšiť až na 1000 impulzov za sekundu, *Nobilli, Mammano, Ashomore* (1998).

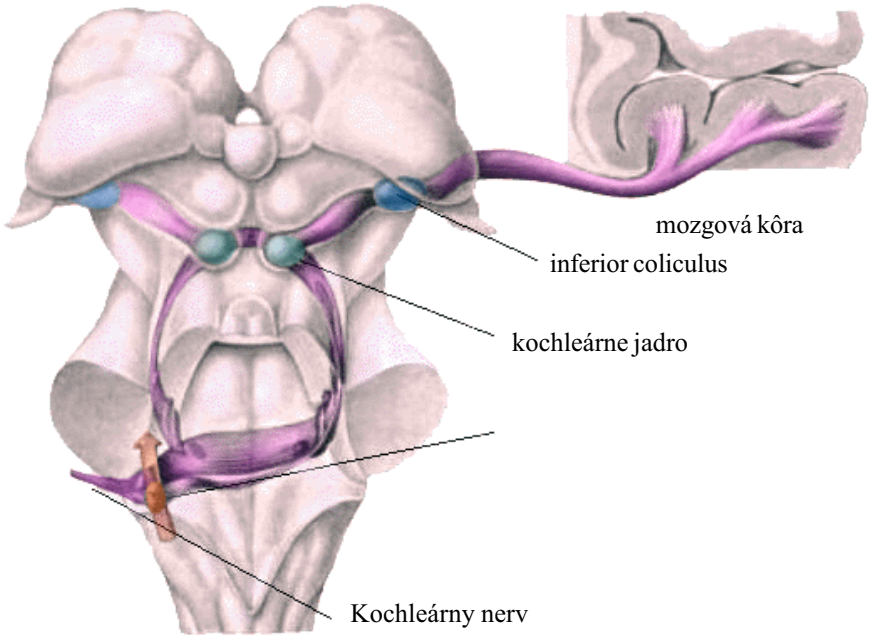
Rýchlosti emisie vyššie ako 1000 impulzov za sekundu nie sú možné kvôli tzv. refrakčnému času alebo refrakčnej perióde nervu po každom impulze, „spajku“. Počas tohto času nervové vlákno znovu obnovuje svoje membránové charakteristiky. V sluchovom nerve je dĺžka tohto času približne 1 - 3 ms.

Po súvislej akustickej stimulácii sa rýchlosti emisie nasýtia na tzv. adaptačné emisné rychlosti. Toto chovanie je vo svojej povahe kolektívne a zdieľajú ho susedné neuróny alebo neuronálne štruktúry. Túto vlastnosť nasýtenia musíme zobrať do úvahy pri každom serióznom modeli rečovej percepcie. Doba adaptácie alebo adjustovania je približne 15 - 20 ms.

Ďalšou dôležitou vlastnosťou aktivity sluchového nervu je potlačenie rýchlosti emisie počas záporných hodnôt stimulačného potenciálu, pod spontánnu rýchlosť emisie.

Pravdepodobnosť intenzity emisie v sluchových nervoch a neurónoch nezávisí v širokom rozsahu intenzity signálu od amplitúdy stimulujúceho signálu. Znovu doba prispôsobenia je v tomto prípade 15 - 20 ms.

Kochleárny nerv vedie ku kochleárnemu jadru a týmto spôsobom priamo do vyšších sluchových centier v mozgu. Detailnejší popis týchto sluchových dráh je na Obr. 18.



Obr. 18 Sluchové dráhy a sluchové centrá

Objavy von Békesyho viedli k rozmachu v modelovaní slimáka. Modely slimáka založené na von Békesyho obraze boli ale v rozpore s psychofyzikálnymi dátami o frekvenčnej selektivite slimáka. Od 70 rokov 20 storočia na základe výsledkov meraní *Rhodeho* (1971), vylepšenými v 80 rokoch autormi *Sellick, Patuzzi, Johnston, B. M.* (1982) bolo jasné, že kmitanie bazilárnej membrány je nelineárne a s oveľa ostrejšou rezonanciou pre nízke hladiny intenzity zvuku, ako sa dovtedy predpokladalo.

Z týchto dôvodov *Davis* (1983) navrhol existenciu aktívneho procesu v kmitaní bazilárnej membrány v rámci orgánu Corti, spätnú väzbu danú vonkajšími vlasovými bunkami. Dôsledkom tejto úvahy bol objav elektromotility vonkajších vlasových buniek, *Kachar, Brownell a kol. (1986), Ashmore* (1990). Pomocou video mikroskopie sa podarilo, na izolovaných vonkajších vlasových bunkách, ukázať závislosť zmeny dĺžky resp. tvaru týchto buniek od medzimembránového potenciálu. Po tomto objave sa zároveň ujasnilo, že ostré vyladenie bazilárnej membrány pre nízke intenzity zvukov, sa dá vysvetliť za predpokladu, že vlasové bunky pôsobia ako aktívne spätné väzby oproti vnútornym viskóznym silám v kvapaline vyplňujúcej slimáka.

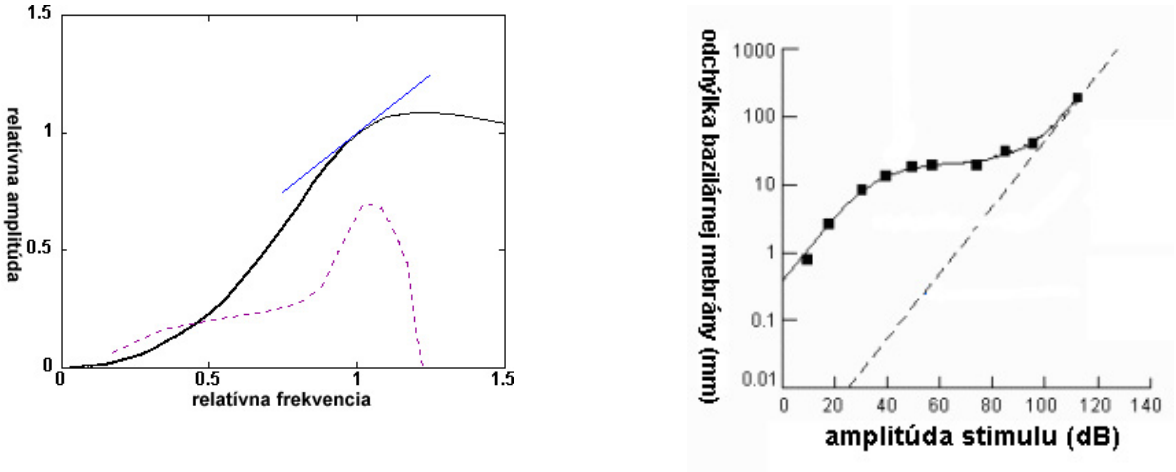
Podrobný výskum, *Nobili, Mammano, Ashomore* (1998), ukázal, že je nutné zahrnúť aj viskózno-elastickú väzbu danú Deiterovými bunkami, naspodu vonkajších vlasových buniek. Ak na orgán Corti pôsobíme súčasne zvukmi evokujúcimi kmitanie bazilárnej membrány a súčasne motorickú spätnú väzbu vlasových buniek, potom pozorujeme dva druhy odchýlok, ktoré sa navzájom kombinujú tak, že vonkajšie resp. vnútorné oblasti vzhľadom na vonkajšie vlasové bunky, kmitajú s opačnou fázou, *Russell, Nilsen* (1997).

Výsledky dosiahnuté v posledných rokoch, *Secker-Walke*r, *Nilsen* (1997), ukazujú, že pre nízke intenzity zvuku je frekvenčné vyladenie nervových vlákien rovnocenné s vyladením odchýlok bazilárnej membrány prefiltrovaných horno-priepusťovým filtrom. Takto, aspoň pre cicavce, sa ukázalo, že excitácie nervových vlákien sú úplne vyjadrené vibráciami bazilárnej mebrány. A nakoniec v roku 2000 skupina vedcov, ktorú viedol *Zhen, Shen, He, Long, Madison, Dallos* (2000), vygenerovala proteín, ktorý je zodpovedný za elektromotilitu vonkajších vlasových buniek, bol nazvaný Prestín. O rok neskôr sa ukázalo, že cielené odstránenie génu Pres, ktorý kóduje prestín, vedie k viac ako sto násobnému zníženiu citlivosti na zvuk, čo znamená, že membránový proteín je fundamentálnou zložkou mechanizmu kochleárneho zosilňovača, *Oliver, He, Klocker, Ludwig, Schulte, Waldegger, Ruppersberg, Dallos, Fakler* (2001).

Bazilárna membrána je úzka a tuhá (akusticky) na okennom konci a široká a pružná na apikálnom konci. Tento prirodzený topografický rozdiel v štruktúre vedie k tomu, že rôzne oblasti vybrujú s rôznymi rezonančnými frekvenciami. A síce koniec blízko strmienka (okenný koniec) vybruje s vyššími frekvenciami ako apikálny koniec, ktorý vibruje s nižšími frekvenciami.

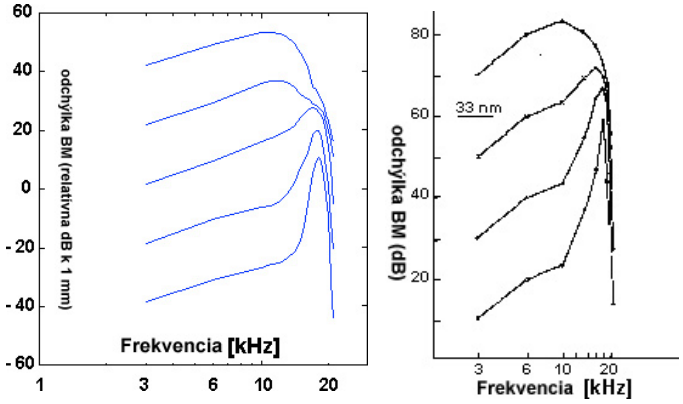
Fázové chovanie slimáka je odlišné od jednoduchej predstavy, ako je súbor frekvenčných priepustí alebo spektrálny analyzátor. Pri vzniku odozvy na obecne časovo-premenný zvuk sa amplitúda na bazilárnej membráne správa podľa rezonančnej frekvencie (charakteristická frekvencia bazilárnej membrány) a táto rezonancia sa vyvíja a pohybuje s oneskorením a šírkou, ktorá závisí na nastavení rezonančných vlastností vlákien, ktoré tvoria bazilárnu membránu. Obecne platí, čím je ostrejšie naladenie tým je vlnový balík spojený so spomínanou rezonaciou úzkejší, ale zároveň tým viac oneskorený na náhlu zmenu frekvencie.Tento vzťah - medzi frekvenčnou selektívnosťou a rýchlosťou odozvy - je analogický vzťahu neurčitosti W.Heisenberga v kvantovej mechanike. Inými slovami, ak sú oscilátory - vlákna bazilárnej membrány - naladené príliš ostro, potom kochlea nemôže sledovať rýchle zmeny frekvencie a jej schopnosť rozpoznávať a diskriminovať životne dôležité zvuky je veľmi slabá. V modeli jednoduchej predstavy - frekvenčných priepustí - fáza odozvy bazilárnej membrány pre každú frekvenciu klesá od 180 stupňov (na báza slimáka) až po 0 stupňov (na apexe - vrchole), na mieste charakteristickej frekvencie je presne -90 stupňov. A pretože je tento model lineárny, potom amplitúdová odozva bazilárnej membrány bude súčtom odoziev pre všetky frekvencie a obecne celková zmena fázy v ľubovolnom mieste závisí nepredvídatelne od štruktúry vstupného signálu. Reálna kochlea sa správa odlišne, *Ashmore* (1990), Obr.19a.

V prvom rade kochlea sa nespráva lineárne, *Lauterborn, Parlitz* (1988). Pre vstupné signály o intenzite 40 - 70 dB sa kochlea správa akoby boli dva rôzne režimy. Odozvy na frekvencie zvuku pod 30 - 40 dB sú zosilnené a pretože sú ostro vyladené sú trochu oneskorené. Odozvy na frekvencie nad 60 - 70 dB sú potlačené, ale naopak pretože sú hrubo naladené, ich odozva je okamžitá. V ľubovolnom prípade, amplitúdy odozvy slimáka na frekvenčné komponenty rôznych amplitúd majú tendenciu byť ekvalizované, pozri Obr.19b.



Obr.19) Správanie odozvy bazilárnej membrány,vľavo b) Pasívna (čiarkovane) odozva slimáka voči aktívnej (reálnejšej) odozve, vpravo odozva, z Davis (1983)

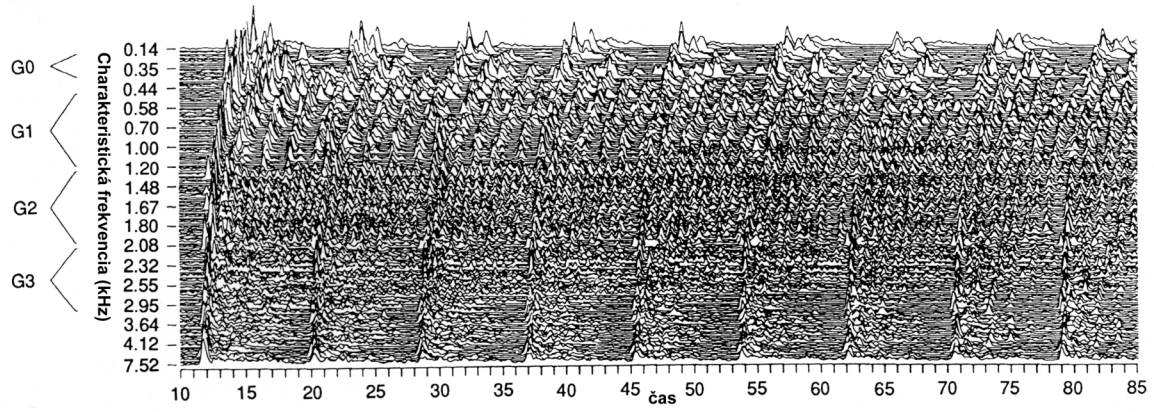
Dôležitým vedľajším efektom takéhoto správania sa je vzájomné potlačenie tónov, efekt, ktorý sa nedá vysvetliť pomocou jednoduchej predstavy akou je súbor frekvenčných priepustí. Tento efekt zapríčiňuje omnoho vyššiu intenzitu odoziev na frekvenčné zložky signálu s veľkou amplitúdou a exponenciálne potlačenie odozvy na susedné frekvenčné komponenty, ale s menšou amplitúdou. Takýmto spôsobom dochádza aj k potlačeniu šumu aj k istému druhu frekvenčnej selektivity pre frekvenčné zložky s dostatočne veľkou amplitúdou, Obr.20, pozri aj modelové a experimentálne odozvy u Johnstona.



Obr.20 Modelová a experimentálna (vpravo) odozvy bazilárnej membrány, Sellick a kol. (1989)

Takto kochlea prevádza analýzu, ktorá sa zdá byť paradoxnou - analýzu kombinujúcu rýchlu odozvu a vysokú frekvenčnú selektivitu, zdanlivo odporujúcu Heisenbergovmu vzťahu. Tieto vzťahy a vlastnosti sa preukázali v analýze časových odoziev z veľkej populácie nervových vlákien na syntetizované zvuky, Obr.21. Na obrázku je kochleogram zobrazený z dát, ktoré boli detekované na sluchovom nerve mačky, *Deng* (1992). Zobrazenie indikuje, že efektívna šírka kochleárnych filtrov je oveľa širšia ako sa preukazuje v mechanických a aj neuronálnych prahových meraniach.

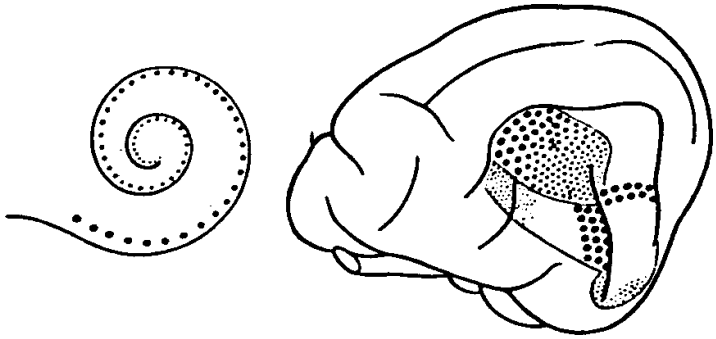
Aby sme mohli vysvetliť tieto odozvy, musíme bazilárnu membránu rozdeliť na segmenty, ktoré kmitajú koherentne s prevažujúcou frekvenciou a s fázovým posunom smerom k vrcholu slimáka. Vidíme, že slimák reaguje prevažne na formanty, čiže extrahuje definujúce vlastnosti zdroja zvuku. Pritom využíva spätno-väzobný saturačný efekt vonkajších vlasových buniek, čo vedie k dvom podstatným efektom - ekvalizácii odozvy a tónovému potlačeniu. Ekvalizácia, efektívne, prevádza percepciu dostatočne vzdialených frekvenčných zložiek nezávisle na ich intenzite. Potlačenie



Obr. 21 Kochleogram zobrazený z dát, ktoré boli detekované na sluchovom nerve mačky, z Deng (1992)

blízkych tónov - potlačenie jedného tónu druhým v jeho blízkosti, je ekvivalentný efektu laterálneho potlačenia, ktorý sa prejavuje v neurónových štruktúrach. Na psycho-fyzikálnej úrovni je toto hlavnou príčinou javu maskovania frekvencií. Ako vedľajší efekt tohto maskovania dostávame ďalšie potlačenie šumu, *Strube* (1985).

Informácia o kmitaní bazilárnej membrány z rôznych polôh je asociovaná na sluchovú kôru pomocou nervových ukončení vlasových buniek. Bazilárna membrána sa mapuje topograficky na zhluk nervových vlákien (nízke frekvencie na jednom konci, vysoké na druhom) v sluchovej kôre. Hovoríme, že tieto stĺpce majú tonotopickú reprezentáciu, *Pickles* (1988), Obr.22.



Obr. 22 Tonotopické usporiadanie sluchovej kôry

Ako sa zvuková informácia šíri zvukovou cestou tak sa uchováva tonotopická organizácia, uzamknutie fázy zvukov vysokých frekvencií sa zosľabuje, a reprezentácia amplitúdovej modulácie sa zosiľňuje. Zvuková informácia sa potom posiela cez talamus do primárnej zvukovej kôry (Brodmanove oblasti 41 a 42), čo je vlastne prvá „relé stanica“ vo vedomej percepcii zvuku. Tieto oblasti kôry pôsobia aj smerom naopak, napríklad pri ochrane stredného ucha, *Pickles* (1988).

Od kochleárneho jadra sa informácia rozdvojuje. Sluchové nervové vlákna vedú do ventrálneho kochleárnehojadra sa správajú, tak že sa uchováva vzájomná časová informácia, takto dovoľujúc časovanie až na mikro sekundy (akčné potenciály sú rádove milisekundy). Ventrálne kochleárne jadrá sa potom projektujú do združenia jadier v medule, ktoré sa nazývajú superior olive. Tam sa porovnávajú okamžité rozdiely v časoch a intenzite zvukov v každom uchu, čím sa určí poloha zdroja zvuku.

Druhý tok informácie začína v dorzálnom kochleárnom jadre. Na rozdiel od časovo citlivej lokalizačnej cesty, tento tok slúži na analýzu kvalít zvuku, ktoré sú dôležité pre komunikáciu.

Zvuky, ktoré sú spracované v primárnej sluchovej kôre sú ďalej spracovávané vo vyšších sluchových centrách a nakoniec v oblasti Wernickeho, ktorá vedie k verbálnemu porozumeniu a asociovaniu objektov so zvukmi a slovami. Brocova oblasť sa naopak zúčastňuje pri verbálnom vyjadrovaní a produkcii zvukov. Je časťou pracovnej pamäti v prednom mozgovom laloku. Táto pamäť je dôležitá napríklad pri usporiadaní slov do viet. Nakoniec je tu tvárová oblasť motorickej kôry, ktorá hýbe konkrétne svaly pri produkcii zvukov.

Sluchová dráha, začínajúc od kochleárneho nervu, sa skladá zo 4 hlavných stupňov : 1) z jadier, ktoré sú aj reflexné aj prechodové; 2) centrálne jadro v colliculus inferior; 3) časť corpus geniculatum mediale a 4) sluchovú kôru. Aferentné vlákna vchádzajúce do kochleárneho jadra pochádzajú hlavne z vnútorných vlasových buniek Cortiho orgánu . Eferentné vlákna vychádzajúce z mediálnych resp. laterálnych neurónov v mozgovom kmeni inervujú vonkajšie resp. vnútorné vlasové bunky. Hlavnou prepínajúcou stanicou - ústredňou pre descendentné sluchové dráhy zo superior olivary komplexu je colliculus inferior, ktorý taktiež prijíma priame afferentné vzruchy z kontralaterálneho kochleárneho jadra. Prakticky všetky sluchové vstupy prechádzajú cez teliesko geniculate, v ktorom sa nachádza špeciálny už spomínaný talamický prepínač sluchového systému a projektuje vzruchy do sluchovej kôry.

Iteratívne kvality týchto ľudských nervových okruhov sa rozširujú až k takým aspektom ľudského chovania ako je syntax. Najnovšie štúdie podporujú tieto tvrdenia a zároveň ukazujú, že klasický model „jazykového orgánu" Broca-Wernickeho nemusí byť správny. Zdá sa, že neurónové okruhy zahrňujúce bazálne gangliá regulujú motorickú kontrolu, syntax, a kognície. Podkôrové bazálne gangliá vytvárajú určitý "sekvenčný stroj", ktorý znovu a znovu iteruje motorické príkazy, uchované ako generátory motorických paternov v mozgu. Ďalej sa ukazuje, že bazálne gangliá môžu iterovať generátory kognitívnych vzorov, odpovedajúce kognitívnej flexibilite a asi sa zúčastňujú aj pri asociatívnom učení. Naznačilo sa to aj evolučnou významnosťou regulačného génu FOXP2, ktorý ovláda embryonický vývoj bazálneho ganglia a iných subkôrových elementov *Lieberman, P.,* (2007).

Stavebné bloky zvukov slov sa nazývajú hlásky (nepresne fonémy, ale v tomto kontexte postačujúce), budeme o nich podrobnejšie diskutovať neskôr. Novorodenci, bez ohľadu kde sú narodení, majú od narodenia schopnosť rozlišovať spoločný súbor foném. A ako sa to zistilo, inými slovami ako sme sa pýtali bábätká, či vedia rozlišovať, napríklad medzi ba a pa ? V *Eimas* (1974) sa zistilo, že bábätká navyknuté na opakovanie jedného zvuku, napríklad ba, ba, atď. začali cucať oveľa rýchlejšie, pri elektronicky meranom monitorovanom cumlíku, pri zmene zvuku (pa, pa atď.). *Kuhl* (1991) testoval nahrané hlásky a podobne monitorované cumlíky po celom svete a v rozličných sociálnych prostrediach. Zistilo sa, že sluchový systém bábätok začína klasifikovať hlásky už od veku 6 mesiacov. Tieto klasifikátory pôsobia ako „magnety", *Kuhl* (1991), ktoré

- *priťahujú fonémy, ktoré sú trochu odlišné, aby zneli podobne ako už známe fonémy,*
- *produkujú jasnú hranicu medzi rôznymi známymi fonémami.*

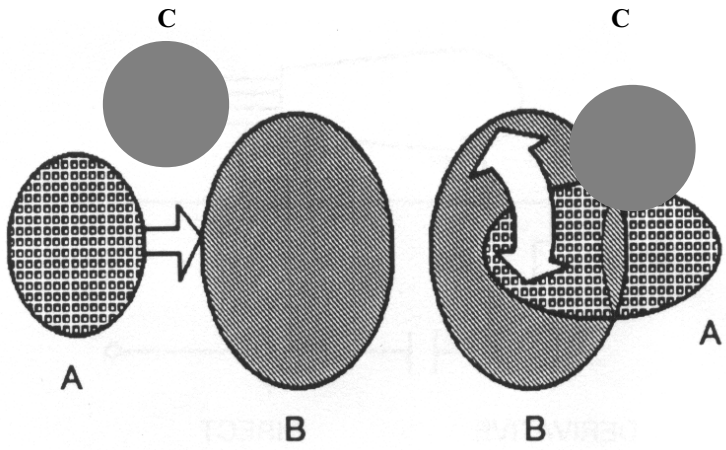
3.1.2 Percepcia papagája

Vieme, že existencia dlhého (35 mm) slimáka v ľudskom uchu je podstatnou pre existenciu kochleárneho spektra. Ale tiež vieme, že vtáky, špeciálne papagáje majú rudimentárny slimák, ktorý je menší než 3 mm. Z tohto dôvodu u papagája nie je možné kochleárne spektrum. Frekvenčná analýza sa prevádza pomocou postupných vĺn, ktoré v tak malom slimáku aký má papagáj nemôžu vzniknúť. Zároveň v uchu papagája nie je jasný rozdiel medzi vnútornými a vonkajšími vlasovými bunkami. Ale aj tak je frekvenčný rozsah ucha papagája (50 - 20 000) Hz a frekvenčné rozlíšenie je podobné ako v ľudskom uchu. Ako vieme papagáj môže uspokojivým spôsobom produkovať všetky ľudské zvuky a s ohraničenou dokonalosťou aj tónovú intonáciu, *Pickles* (1988).

3.2 Obecný pohľad na percepciu

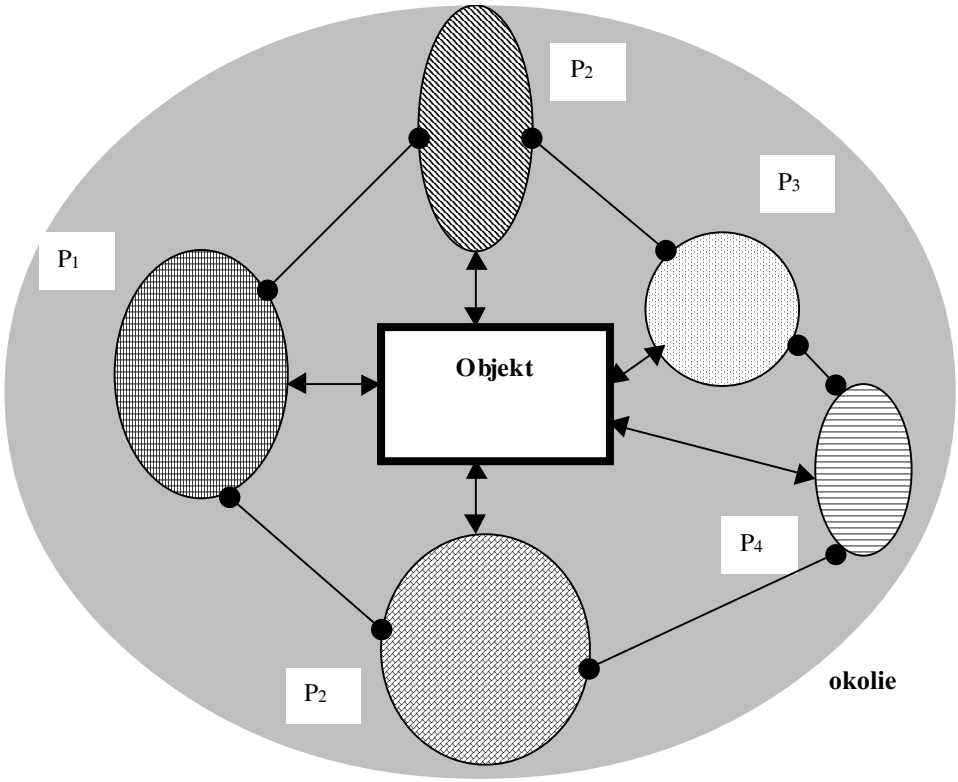
V tejto časti sa pokúsime urobiť niekoľko obecných, konceptuálnych poznámok o percepcii z hľadiska abstraktných myšlienok fyziky, spracovania informácie, komunikácie a úvah o symetriách, aby sme videli aj pozadie našej argumentácie, jej smerovania a navrhnutého modelu.

Predpokladajme, že máme dva systémy so špecifickými a odlišnými funkciami. Prvý systém (A) je pozorovaný a druhý systém (B) je pozorujúci. Z dôvodu úplnosti musíme zahrnúť aj okolie (C). Z tohto dôvodu musíme predpokladať, že oba systémy sú do určitej miery separované, oddelené. Tento predpoklad je nutný, aby sme definovali, odlíšili oba systémy, ale zároveň musíme uvažovať aj istý druh interakcie medzi týmito systémami, pretože izolované a uzavreté systémy sa nedajú pozorovať, *Pitts, McCulloch* (1947), *Shalkoff* (1992). Obecne hovoriac, systém B je konštruovaný tak, aby pozoroval niektoré vlastnosti systému A, čo predpokladá určitú apriórnu znalosť systému A. Konštrukcia systému B môže byť umelá alebo prirodzená - evolučná, Obr. 23.



Obr. 23 Pozorovaný (A), pozorujúci (B) systémy a okolie (C)

Pri evolučnej konštrukcii, počas narodenia ale aj tesne pred ním, zažije jedinec (nutne sa nemusí jednať len o človeka) niekoľko rozlíšení, medzi iným aj okamih kedy rozlíši seba od zvyšku sveta. Takéto rozlíšenia robí potom počas celého zvyšku života, učíme sa že aj keď môžeme rozlíšiť časti , tieto nie sú izolované ale sú v nejakom vzťahu. Inými slovami interakcie opisujú rozdiel medzi celkom a súborom častí. Počas detstva sa učíme rozlišovať medzi týmito vzťahmi , interakciami s okolím - nazývame ich percepciami. Tie ktoré sú zdielané, percepované aj inými ľuďmi sa nazývajú pozorovania, naopak tie ktoré sú jednoznačne osobné nazývame pocity. Tomuto rozdeleniu odpovedajú aj pojmy ako realita, skutočnosť alebo predstava, verejné alebo osobné. Obecne sa táto naša schopnosť rozlišovať časti alebo pozorovať vzťahy medzi nimi nazýva schopnosť klasifikácie. Senzorické vstupy sa spracovávajú - klasifikujú, uchovávajú v pamäti a vyvolávajú z nej. Bez tejto informácie a takto spracovanej informácie nevieme vlastne čo pozorovať. Napríklad v rozpoznávaní reči, určitou extrémnou situáciou je situácia, kde rozpoznávame vlastne to čo už vieme. Takto môžeme definovať percepciu ako určitý druh pozorovania alebo interakcie, ktorá existuje medzi viazanými alebo vzájomne sa vyvýjajúcimi systémami na rozdiel od pozorovania vo vlastnom zmysle kde sú systémy, ktoré uvažujeme, dostatočne oddelené, Obr. 24. Percepciu chceme pochopiť, hlavne, ako zdielanú interakciu s okolím, aby sme sa vyhli alebo aspoň korektne zapracovali problém, apriórnej znalosti, ktorý úzko súvisí s chápaním časti a celku. Tomuto sa budeme venovať neskôr.



Obr. 24 Pozorovaný (Objekt) a pozorujúce, komunikujúce (P_1, P_2,...) systémy - pozorovatelia a okolie objektu a pozorovateľov

3.2.1 Komplexné prístupy k percepcii

Z trochu iného hľadiska mô žeme charakterizovať percepciu ako konečné spojenie v reťazci javov (z vonkajšieho sveta - zvukové vlny, elektromagnetické žiarenie - do vnútorného sveta - akt percepcie) a na jej pochopenie potrebujeme znalosti o každom spojení v reťazci. Tento reťazec aktuálne pretína viacero rôznych vedeckých disciplín, v rozmedzí od fyziky k psychológii (a sociológii, jazykovedy, informatiky). Tieto disciplíny používajú rôzne stupne analýzy, v rozsahu od mikroskopickej (štúdium chovania molekúl) k makroskopickej (štúdium chovania orgánov alebo celých organizmov). Takže, aby sme dostali úplný obraz, potrebujeme analyzovať percepciu na niekoľkých rôznych úrovniach, z ktorých každá ponúka plnohodnotnú perspektívu a rozhľad.

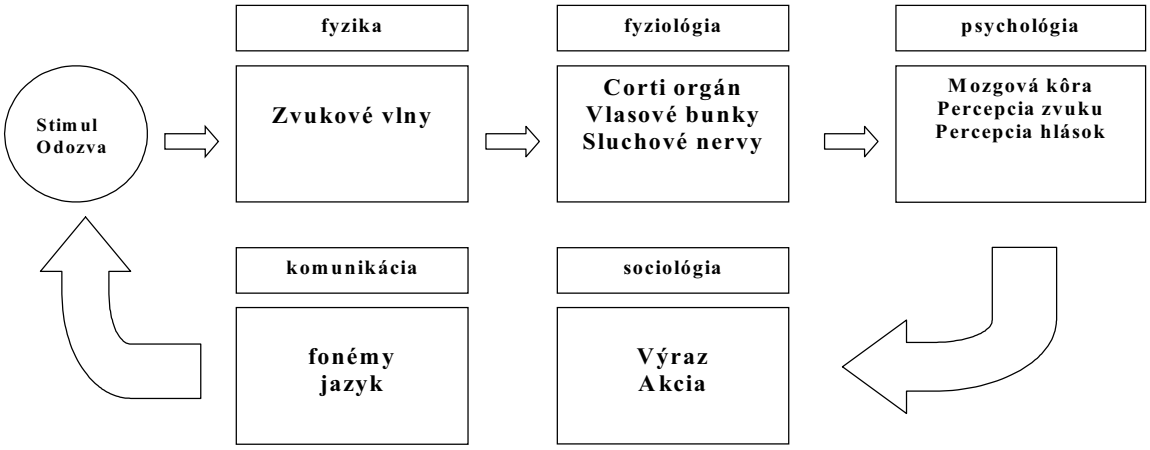
Tri hlavné úrovne, ktoré budeme potrebovať pri našej analýze sú fyzikálna, fyziologická a fonologická. Obecne hovoriac, fyzikálny prístup sa zameriava na popis zvukových vĺn a perceptuálne kapacity rôznych zmyslov a, v prípade ľudí, na skúsenostné aspekty percepcie. Spätne hľadiac k obrazu reťazca, fyzikálny prístup sa dotýka jedného konca sekvencie v reťazci. Fyziologický prístup, sa nachádza v reťazci, relatívne, niekde v strede, a z nášho hľadiska, na konci je fonologický alebo psychologický prístup, *Pisoni, Remez* (2005).

Psycho-biologický prístup sa zameriava napríklad na anatómiu a fyziológiu senzorickej časti nervového systému. Avšak, hranice medzi týmito tromi prístupmi nie sú úplne zreteľné. Tieto tri prístupy pravdepodobne nie sú vzájomne sa vylučujúce ale sú komplementárne; jednoducho nemôžeme sa naučiť všetko čo chceme vedieť o percepcii iba zo štúdia jediného prístupu. Nakoniec, ako vidieť aj na Obr. 25, existujú tiež ďalšie dva prístupy, sociologický a jazykový, ktoré musíme zobrať do úvahy ak chceme porozumieť, napríklad, fonémický systém konkrétneho jazyka.

3.2.2 Metodologický prístup

V prvom rade, neexistuje jediný psycho-biologický prístup k percepcii; namiesto toho ich je niekoľko, odlišujúc sa navzájom mnohými spôsobmi. Pre ľudí (a zvieratá) sa na skúmanie percepcie používa

správanie. Hoci všetky prístupy používajú, pri štúdiu percepcie, nejakú reakciu správania na nejaký podnet, líšia sa v tom aká reakcia sa požíva. Napríklad, nejaký subjekt sa môže inštruovať, aby stlačil jedno tlačítko ak je prítomný zelený objekt a aby stlačil iné tlačítko ak je prítomný červený objekt. Pri študovaní percepcie sa aktuálne používa množstvo špecifických techník.



Obr. 25 Obecná percepčná schéma

Podľa obecnejšej schémy, môžeme vyšetrovať percepciu z pohľadu „formálnosti" alebo ešte lepšie vyjadrené z pohľadu validity. Pod „formálnosťou" myslíme do akého rozsahu sú podnety a ich odozvy štrukturované a kontrolované. Najmenej formálny a zároveň najviac validnejší prístup je fenomenalisticko naturalistický prístup, skrátene fenomenologický. „Fenomenologický" tu znamená, že na evidenciu sa používa vedomá skúsenosť subjektov štúdie. „Naturalistický" tu znamená, že na evidenciu sa používajú odozvy na akékoľvek podnety prirodzene sa vyskytujúce v danom prostredí, okolí, nie je snaha akýmkoľvek spôsobom modifikovať tieto stimuly alebo vytvoriť nejaké umelé.

Tento prístup k štúdiu percepcie má určité výhody. Napríklad, je založený na najľahšie dostupných skúsenostných dátach, ktoré sú evokované prirodzene sa vyskytujúcimi udalosťami a ako už bolo naznačené táto metóda má dostatočnú ekologickú validitu. Skúsenosti o aké sa jedná sú napríklad nárast Mesiaca pri splne, trajektórie padajúcich objektov, rečové vzory u ľudí. Každý z nás zažíva každý deň nespočetné množstvo takýchto skúseností. Takto, aby sme študovali percepciu stačí ak zozbierame a zorganizujeme takéto skúsenosti. Samozrejme, ak urobíme krok trochu napred, môžeme tieto percepčné skúsenosti diskutovať s inými ľuďmi a porovnávať, testovať pomocou jazykového nástroja, *Lieberman* (2007), *Poeppel, Idsardi1, van Wassenhove* (2008).

Vedúcim princípom nášho prístupu na nasledujúcich stranách je Erlangenský program Felixa Kleina, pomocou, ktorého chceme pochopiť našu schopnosť skrz senzoricky rozličné skúsenosti prísť k nejakému spoločnému pochopeniu sveta pomocou komunikácie a našu schopnosť porozumieť svetu skrz jeho reprodukovateľnosť a konzistenciu.

3.2.3 Komunikácia a príncípy symetrie

Prečo môžeme rozumieť niekomu, keď hovorí o svete, hoci nie sme tým druhým ? Môžeme to z dvoch príčin: pretože väčšina vecí vyzerá podobne z rôznych pohľadov, hľadísk, a pretože väčšina z nás už mala podobné skúsenosti. 'Podobné' znamená, že čo my a čo druhí pozorujú nejako korešponduje. Inými slovami, veľa aspektov pozorovaní nezávisí na hľadisku. Napríklad, počet prstov na ruke má rovnakú hodnotu pre všetkých pozorovateľov, kľudová hmotnosť alebo elektrický náboj sú inými takými príkladmi. Preto môžeme povedať, že takéto veličiny majú najvyššiu možnú symetriu. Pozorovateľné veličiny s najvyššie možnou symetriou sa nazývajú vo fyzike skaláry. Iné aspekty sa menia od pozorovateľa k pozorovateľovi; napríklad, zdanlivá veľkosť sa mení so vzdialenosťou pozorovania. Avšak, skutočná veľkosť je nezávislá od pozorovateľa. Obecne,

ľubovolný typ nezávislosti na hľadisku je formou symetrie, a pozorovanie, že dvaja ľudia pozerajúc na rovnakú vec z rôznych hľadísk sa môžu navzájom dorozumieť vlastne dokazuje, že príroda je symetrická (otázkou je samozrejme, či je to úplne nezávisle na pozorovateľoch ako takých).

Existuje ešte iný typ podobností, a síce nielen rovnaký jav vyzerá podobne pre rôznych pozorovateľov, ale rôzne javy vyzerajú podobne pre rovnakého pozorovateľa. Napríklad, od malička sa učíme, že rozžeravený plech v kuchyni opáli prst, urobí to aj mimo domu takisto, a tiež na iných miestach a v iných časoch. Podobne sa učíme, že zajtra zase vyjde slnko. Príroda vykazuje reprodukovateľnosť. Fakticky, naša pamäť a naše myslenie sú možné iba kvôli tejto základnej vlastnosti prírody (hovoríme samozrejme o zachovaní energie a hybnosti).

Bez nezávislosti na hľadisku a reprodukovateľnosti hovorenie s inými alebo so sebou o svete, prírode by nebolo možné. Dokonca ešte dôležitejšie, ukazuje sa, že nezávislosť na hľadisku a reprodukovateľnosť robí viac než len určenie možnosti hovoriť navzájom; fixujú obsah toho čo si môžeme povedať navzájom. Inými slovami, môžeme povedať, že opis prírody vyplýva logicky, skoro bez možnosti výberu, z jednoduchého faktu, že môžeme navzájom komunikovať o našich percepciách, *Fischbach* (1992)*, Kolers, Eden* (1968).

Ak sa stretávame s inými ľuďmi počas detstva, rýchlo objavíme, že niektoré skúsenosti sú zdieľané, zatiaľčo iné, ako sú sny, nie sú zdieľané. Naučiť sa rozlíšiť tento rozdiel je jednou zo základných schopností, ktorým sa učíme počas života. Pri zdieľaných vedeckých pozorovaniach sa sústreďujeme na prvý typ skúseností, fyzikálne pozorovania. Avšak, aj medzi týmito, sa robia rozdiely. V každodennom živote predpokladáme, že váhy, objemy, dĺžky, časové intervaly sú nezávislé na hľadisku pozorovateľa. Môžeme hovoriť o týchto pozorovaných veličinách s kýmkoľvek, a nie sú spory o ich hodnotách, za predpokladu, že sú korektne merané. Avšak, iné veličiny závisia na pozorovateľovi *Lenz* (1990), *Locher, Nodine* (1989).

V prípade pozorovaní závislých na hľadisku, je porozumenie stále možné vynaložením určitého úsilia: každý pozorovateľ si môže predstaviť pozorovanie z pohľadu toho druhého, a skontrolovať či predstavovaný výsledok súhlasí s tvrdením toho druhého. Ak tvrdenie takto predstavované a skutočné tvrdenie druhého pozorovateľa súhlasia, pozorovania sú konzistentné, a rozdiel vo výrokoch je iba kvôli rôznym hľadiskám; v opačnom prípade, rozdiel je fundamentálny, a nemôžu súhlasiť alebo rozprávať sa. Dôležité si je uvedomiť, že pomocou tohto prístupu, môžeme argumentovať dokonca či ľudské vnemy, pocity, úsudky, alebo chuť vedú k fundamentálnym rozdielom alebo nie.

Na pochopenie týchto fundamentálnych rozdielov je nevyhnutný predpoklad externej reality. Z väčšej časti je veda objavovaním a vysvetľovaním externého sveta. Bez tohto predpokladu by boli iba myšlienky a predstavy našej mysle (ktorá by bola jedinou mysľou) a veda alebo niečo iné by neboli vôbec nutné. Okrem predpokladu externej reality, predpokladáme tiež, že táto realita je objektívna. A to sa neustále potvrdzuje našou každodennou skúsenosťou ako aj vedeckými pozorovaniami. Objektivita znamená, že pozorovania, experimenty, alebo merania urobené jednou osobou sa môžu urobiť aj druhou osobou, ktorá dostane rovnaké alebo podobné výsledky. Druhá osoba bude schopná potvrdiť, že výsledky sú rovnaké alebo podobné konzultáciou s prvou osobou. Preto pre objektivitu je podstatnou komunikácia. Fakticky, nejaké pozorovanie, ktoré sa nekomunikuje a s ktorým sa nesúhlasí sa obecne neakceptuje ako platné pozorovanie objektívnej reality. Pretože sa vyžaduje súhlas, objektívna realita sa niekedy nazýva realitou konzensu.

Fakticky, všetky pozorovania takzvanej "externej" reality sú v skutočnosti pozorovania našich vlastných mentálnych dojmov, ktoré sú výsledkom nejakých stimulov, o ktorých sa predpokladá, že sú externé. Tuto "externé" znamená externé voči mysli, nie nutne externé k telu. Napríklad, ak zakúšame bolesť ako odozvu pri pichaní injekcie alebo pri postihnutí chrípkou, nikto by nepochyboval o objektivite našich pozorovaní.

V doterajšej diskusii sme sa zaoberali hľadiskami, ktoré sa líšia v polohe, v orientácii, v čase a najdôležitejšie v pohybe. Pozorovatelia vzájomne môžu byť v pokoji, pohybovať sa s konštantnou rýchlosťou, alebo zrýchlene. Tieto 'konkrétne' zmeny hľadiska budeme rozoberať ako prvé. V tomto prípade požiadavka konzistencie pozorovaní urobených rôznymi pozorovateľmi sa nazýva

princípom relativity. Symetrie asociované s týmto typom invariancie sa nazývajú externé symetrie. Príklady takýchto symetrií sú uvedené v nasledujúcej Tab. 2.

symetria	grupa	priestor pôsobenia	reprezentácie	fyzika	komunikácia
Časový a priestorový posuv	RxR^3	priestor, čas	skaláry, vektory	hybnosť a energia	dovoľuje každodennú komunikáciu
Rotácia	SO(3.)	priestor	tenzory	moment hybnosti	dovoľuje každodennú komunikáciu
Galileiho transformácia	R^3	priestor, čas	skaláry, vektory	ťažisko rýchlosť	dovoľuje každodennú komunikáciu
Lorentzova transformácia	homogénna Lie SO(3,1)	priestoro-čas	tenzory, spinory	energia-hybnosť $T^{\mu\nu}$	dovoľuje technologickú komunikáciu

Tab. 2 Príklady vonkajších - časopriestorových symetrií

V predošlej tabuľke nie sú samozrejme uvedené všetky externé symetrie. Napríklad neuviedli sme Poincarého nehomogénnu grupu, dilatačné grupy, a podobne, *Bhagavantam, Venkataraydu* (1951)*, Moller* (1972),

Druhá trieda fundamentálnych zmien hľadiska uvažuje 'abstraktné' zmeny. Hľadiská sa môžu líšiť použitým matematickým popisom a v tomto prípade sa nazývajú zmenami kalibrácie. Opäť, požaduje sa, aby všetky tvrdenia boli konzistentné v rámci rôznych matematických popisov. Táto požiadavka konzistencie sa nazýva princípom kalibračnej invariancie. Asociované symetrie sa nazývajú vnútorné symetrie, budeme sa im venovať neskôr, pozri nasledujúcu Tab. 3.

symetria	grupa	priestor účinku	fyzika	komunikácia
Elektromagnetická klasická kalibračná invariantnosť	[∞ par]	fázový priestor	electrický náboj	fotóny
Elektromagnetická kvant.mech. kalibračná invariantnosť	ábelovská Lie U(1)	Hilbertov priestor	electrický náboj	fotóny
Slabá kalibračná nie ábelovská	Lie SU(3)	Hilbertov priestor		bozóny
Farebná kalibračná nie ábelovská	Lie SU(3)	Hilbertov priestor	farebný náboj	gluóny

Tab. 3 Príklady vnútorných symetrií

Neuvádzame triedy diskrétnych symetrií, ako sú priestorová parita, časová parita, nábojová konjugácia, ich združenie CPT symetria, chirálna symetria a iné, *Adams* (1969)*, Bhagavantam, Venkataraydu, T.* (1951).

Požiadavky symetrie, konzistencie sa nazývajú 'princípy', pretože tieto základné tvrdenia sú tak silné, že skoro úplne určujú 'zákony' fyziky, ale aj obsah toho, čo si môžeme povedať navzájom pri pozorovaní vonkajšieho sveta - prírody *Pitts, McCulloch* (1947), *Locher, Nodine* (1989).

3.2.4 Symetrie a grupy

Ak hľadáme nejaký úplný popis systému, jeho stavov a jeho pohybu, musíme porozumieť a opísať úplnú množinu symetrií takého systému. Systém, ktorý sa javí totožný ak sa pozoroval z rôznych hľadísk sa nazýva symetrickým alebo hovoríme že má symetriu, je symetrický. Tiež hovoríme, že systém má invarianciu vzhľadom k špecifickým zmenám od jedného hľadiska k druhému. Zmeny hľadiska sa nazývajú symetrickými operáciami alebo transformáciami. Symetria je takto súbor transformácií. Avšak, je aj niečo viac: spojenie dvoch elementov, konkrétne dvoch operácií symetrie, je druhá operácia symetrie. Aby sme boli presnejší, symetria je množina G = {a, b, c,...} prvkov, transformácií, spolu s binárnou operáciou ° nazývanou spojenie alebo násobenie, pre ktorú platia nasledovné vlastnosti pre všetky prvky a, b a c:

asociatívnosť, t.j. $(a \circ b) \circ c = a \circ (b \circ c)$
existuje neutrálny prvok e, taký že $e \circ a = a \circ e = a$
existuje inverzný element a^{-1} taký že $a^{-1} \circ a = a \circ a^{-1} = e$

Ľubovolná množina, ktorá spĺňa tieto definujúce vlastnosti alebo axiómy sa nazýva (matematickou) grupou. Grupy sa vo vede obecne, a špeciálne vo fyzike a matematike objavujú často, pretože symetrie sú skoro všade, ako vidíme. Počet prvkov grupy sa nazýva rád grupy. Ak je tento rád konečný tak aj grupa sa nazýva konečná, inak sa nazýva nekonečná. Nekonečné grupy sa delia na diskrétne alebo spojité. Pokiaľ sa každému prvku grupy dá priradiť prirodzené číslo, potom sa grupa G nazýva diskrétna, naopak spojitá. Prvky hocijakej spojitej grupy sa dajú parametrizovať pomocou reálnych spojitých parametrov

$$g = g(\lambda) \; ; \; \lambda = \lambda_1, \lambda_2, ..., \lambda_n$$

Ak $n \in N$, kde N je množina prirodzených čísiel, potom sa spojitá grupa G nazýva konečno rozmerná, a n sa nazýva dimenzia grupy. V opačnom prípade sa spojitá grupa nazýva nekonečne rozmerná. Majme parametre súčinu dvoch prvkov spojitej grupy $p^l(\lambda, \kappa)$; $p = (p_1, p_2, ...)$, potom ak

$$g(\lambda).g(\kappa) = g(\lambda,k); \; \lambda = \lambda_1, \lambda_2, ..., {}_n \text{ a } \kappa = \kappa_1, \kappa_2, ...,$$

sú tieto parametre analytickými funkciami, svojich parametrov (skrátene povedané funkcie f majú derivácie všetkých rádov vo všetkých svojich argumentoch) a podobne parametre inverzného prvku λ^{-1}; $g(\lambda^{-1}) = g^{-1}(\lambda)$, potom sa takáto grupa nazýva Lie grupa, *Adams* (1969).

3.2.5 Reprezentácie

Pozorujúc symetrický a zložený systém, aký je napríklad nejaký fyzikálny dynamický systém, alebo farebný geometrický ornament (symetrický), alebo systém percepcie a s ním asociovanú nejakú sústavu jazykových jednotiek (hlásky alebo fonémy) prirodzeného jazyka, si všimneme, že každá z jeho častí patrí k podobným objektom, obyčajne sa nazýva multiplet. Uvažovaný ako celok, má multiplet (prinajmenšom) vlastnosti symetrie celého systému. Fakticky, v ľubovolnom symetrickom systéme každá časť sa dá klasifikovať podľa toho do akého multipletu patrí. Multiplet je množina častí, ktoré sa transformujú do seba cez všetky transformácie symetrie. Matematici často nazývajú abstraktné multiplety reprezentáciami. Špecifikovaním do ktorého multipletu komponent patrí, vlastne opíšeme akým spôsobom je komponenta časťou celého systému, alebo akým spôsobom jednotlivé časti interagujú, alebo akým spôsobom komunikujeme, aby sme sa zhodli na obsahu. Pozrime sa ako sa dosiahne takáto klasifikácia. V matematickom jazyku, transformácie symetrie sa obyčajne

opisujú maticami. Napríklad, na rovine, odraz pozdĺž prvej diagonálnej osi sa reprezentuje pomocou matice

$$M(\text{odraz}) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

pretože každý bod (x,y) sa transformuje na (y,x) pri násobení maticou M(odraz). Podobne pre maticu otočenia v rovine

$$M(\text{otočenie}) = \begin{bmatrix} \cos(\alpha) & \sin(\alpha) \\ \sin(\alpha) & -\cos(\alpha) \end{bmatrix}$$

Preto, pre matematikov reprezentácia grupy symetrie G je nejaké zobrazenie, priradenie matice M(a) každému prvku grupy takým spôsobom, že reprezentácia spojenia dvoch elementov a a b nie je nič iné ako súčin reprezentácií M každého elementu:

$$M(a \circ b) = M(a)\, M(b)$$

Napríklad, matica predošlej rovnice odrazu, spolu s korešpondujúcimi maticami pre všetky iné operácie symetrie, majú takúto vlastnosť.

Pre každú grupu symetrie je dôležitou úlohou konštrukcia a klasifikácia všetkých možných reprezentácií. Korešponduje klasifikácii všetkých možných multipletov symetrického systému, ktoré sa dajú nájsť alebo zostrojiť. Takýmto spôsobom, pochopenie klasifikácie všetkých multipletov a častí, nás naučí ako klasifikovať všetky možné časti, z ktorých môže byť zložený nejaký objekt, systém alebo pohyb.

Reprezentácia sa nazýva unitárna ak sú všetky matice M unitárne. Skoro všetky reprezentácie vo fyzike, okrem hŕstky výnimiek, sú unitárne: tento výraz je najobmedzujúcejší, pretože určuje, že odpovedajúce transformácie sú jedno-jedno-značné a invertovateľné, čo v našom prístupe znamená, že jeden pozorovateľ nevidí viac než nejaký druhý. Čiže obyčajne, ak nejaký pozorovateľ môže hovoriť s druhým, druhý môže tiež hovoriť s prvým. Konečná dôležitá vlastnosť multipletu alebo reprezentácie sa dotýka jej štruktúry. Ak sa dá multiplet vidieť ako zložený z pod-multipletov, nazýva sa reducibilný, ináč je ireducibilný ; rovnaké sa dá povedať aj o reprezentáciach. Ireducibilné reprezentácie sa obyčajne nedajú rozložiť ďalej (podobne ale nepresne ako s prvočíslami).

Grupa sa nazýva abelovská ak je grupová operácia komutatívna, t.j. ak platí $a \circ b = b \circ a$ pre všetky páry prvkov v danej grupe. V tomto prípade násobenie sa niekedy nazýva sčítanie. Podmnožina $G_1 \subset G$ grupy G môže byť samotná grupou; potom ju nazývame podgrupou a často nedbanlivo hovoríme, že G je väčšia ako G_1 alebo že G je vyššia grupa symetrie ako G_1 .

Pre reprezentáciu existujú niektoré obyčajné, ale dôležité podstatné podmienky: matice M(a) musia byť invertovatelné, alebo nesingulárne, a operácia identity na G sa musí mapovať do jednotkovej matice. V kompaktnejšom jazyku hovoríme, že reprezentácia je homomorfizmus z G do grupy nesingulárnych alebo invertovateľných matíc. Matica M je invertovateľná ak jej determinant det M nie je nulový. Obecne, ak existuje zobrazenie f z grupy G do inej G´ také, že

$$f(a \circ_G b) = f(a) \circ_{G'} f(b)$$

dve grupy sa nazývajú homomorfné , a zobrazenie f homomorfizmus. Zobrazenie, ktoré je tiež jedno-jednoznačné sa nazýva izomorfizmus. Ak reprezentácia je aj izomorfizmus, potom sa nazýva vlastná. (Pozn. rovnako ako grupy, aj zložitejšie matematické štruktúry také ako okruhy, polia a algebry sa môžu reprezentovať pomocou vhodných tried matíc).

Množina M(n,R) všetkých reálnych štvorcových matíc n x n tvorí komutatívnu Lie grupu relatívne k súčtu matíc. Vzhľadom k súčinu matíc to obecne neplatí, pretože existujú singulárne matice, ktoré nemajú inverznú maticu (regulárne matice sú potom tie, ktoré majú inverznú maticu). Dimenzia Lie grupy je rovná počtu prvkov matice, čiže n^2. Obecnou lineárnou Lie grupou vzhľadom k súčinu nazývame množinu všetkých reálnych regulárnych matíc, označujeme ich

$$GL(n, R) = \{A \in M(n, R); \det A \neq 0\}; \text{ s dim } GL(n, R) = n^2$$

Táto grupa má z hľadiska vedy, špeciálne percepcie, niektoré dôležité podgrupy ako sú špeciálna lineárna grupa:

$$SL(n, R) = \{A \in GL(n, R); \det A = 1\}; \text{ s dim } SL(n, R) = n^2 - 1$$

reálna ortogonálna grupa:

$$O(n, R) = \{A \in GL(n, R); A^T A = I\}; \text{ s dim } O(n, R) = n(n-1)/2$$

špeciálna ortogonálna grupa:

$$SO(n, R) = \{A \in O(n, R); \det A = 1\}; \text{ s dim } SO(n, R) = n(n-1)/2$$

Ako sme už spomenuli pre percepciu (podľa našej definície) sú dôležité komplexné grupy, špeciálne, komplexné unitárne grupy:

$$U(n, C) = \{A \in GL(n, C); A^{\dagger} A = I\}; \text{ s dim } U(n, C) = n^2$$

pričom $A^{\dagger} = (A^T)^*$, je komplexne združená k transponovanej matici. Niekedy sa používa skrátený zápis U(n).

A nakoniec grupa (ktorá, ako uvidíme neskôr, je priamo vyjadrením vnútorných symetrií nášho systému percepcií, rozpoznávania a identifikácie) - špeciálna unitárna grupa

$$SU(n) = \{A \in U(n); \det A = 1\}; \text{ s dim } SU(n) = n^2 - 1$$

V predošlom rozoberaný rozdiel medzi hľadiskovou-invarianciou a veličinami závislými od hľadiska je podstatný. Invariantné veličiny, ako je počet, hmotnosť alebo tvar, opisujú vnútorné vlastnosti pozorovaného systému, a veličiny závisiace na pozorovateľovi vytvárajú stavy systému. Preto, aby sme našli popis stavov nejakého systému, musíme odpovedať na nasledujúce otázky:

- *Ktoré hľadiská sú možné pri popise nášho systému ?*
- *Ako sa popisy nášho systému transformujú z jedného hľadiska na druhé ?*
- *Ktoré veličiny dovoľujú tieto transformácie - symetrie ?*

V tejto práci sa pokúsime skonštruovať modely rozpoznávania a verifikácie ako prvé priblíženie k ľudskej percepcii chápanej, ako sme sa tu pokúsili uviesť, *Lenz* (1990).

Ale predtým ešte prediskutujeme, znovu z obecného hľadiska, dôležité otázky vzťahu časti a celku z hľadiska pozorovania, špeciálne percepcie. Inšpiráciou nám boli knihy *Heisenberga*, Časť a celok (Der Teile und das Ganze), (1969) a Fyzika a filozofia, (1966).

3.2.6 Časť, celok a percepcia

Ako sme už spomenuli v úvodnom zamyslení tejto kapitoly, percepciu chceme pochopiť ako zdielanú interakciu s okolím. V tejto časti sa pokúsime rozobrať problém apriórnej znalosti pri pozorovaní a špeciálne pri percepcii. Fyzikálna teória, aby sa dala pokladať za úplnú, musí špecifikovať ako sú elementy teórie, model, matematické pojmy spojené s ľudskou skúsenosťou. V klasickej fyzike je toto spojenie metafyzické, nie je časťou dynamického procesu ani nie je časťou klasickej fyzikálnej teórie.

Na druhej strane v kvantovej mechanike priradenie matematicky opísaného fyzikálneho stavu k ľudskej skúsenosti je obsiahnuté v matematicky špecifikovanej dynamike. A toto spojenie nie je pasívne, ako vieme, nevypočítava iba nejaké fyzikálne črty prírody, ale vkladá do fyzikálneho systému, na ktoré pôsobí, špecifické vlastnosti, ktoré závisia od výberov uskutočnených pozorovateľom. Pozorovateľom ktorý nie je pasívny ale aktívny pozorovateľ - agent. Aktívne priradenia podobného druhu prevádzame už počas nášho detstva skrz vôľové akty činnosť - odozva, pri ktorých konštruujeme očakávania o tom, ktoré odozvy sú pravdepodobnejšie a ktoré nie pri našom úsilí, našej činnosti. Agenti prevádzajú zvolené činnosti, na ich základe sa očakáva nejaká odozva, spätná väzba.

Príkladom môže slúžiť už klasický Geigerov čítač, ktorý jadrový fyzik umiestni vedľa rádioaktívneho zdroja za účelom, aby zistil či dôjde počas daného časového intervalu ku kliknutiu čítača alebo nie. Potom očakávaná odpoveď „áno“, alebo „nie“ na otázku či „dôjde ku kliknutiu čítača za daný časový interval ?“ špecifikuje 1 bit informácie. Takto aj empirická veda aj obyčajná ľudská činnosť sú založené na takýchto spárovaných realitách akcia - odozva. Aj psychológia aj fyzici sa snažia pochopiť tieto reality v rámci nejakého racionálneho kontextu. V klasickej fyzike (Newton, Maxwell, Einštein) sa objekty modelujú ako maličké kópie planetárnych telies, abstrahované na bodové - hmotné body, ktoré sa pohybujú podľa relatívne jednoduchých pravidiel. Tieto pohyby sú dané iba relatívnymi vzťahmi - súradnicami jednotlivých bodov, a tieto pohyby sa javia ako nezávislé od toho či ich pozorujeme alebo nie.

Podľa kodaňskej interpretácie musí štruktúra fyzikálnej teórie zahrňovať nielen časť, ktorá popisuje nie priamo percepované teoreticky postulované entity (vyjadrené matematickými symbolmi) ale aj časť popisujúcu ľudské skúsenosti, ktoré sa dotýkajú testov a aplikácií ohľadom spomínaných entít v jazyku, ktorý normálne používame ak komunikujeme samy so sebou alebo s druhými ľuďmi. A úplná teória by mala vedieť špecifikovať prepojenie medzi týmito časťami. V klasickej fyzike sú odpovedajúce skúsenosti priamymi záznamami hrubých vlastností objektov. Tieto záznamy (padajúce jablko, výchylka ručičky na meracom prístroji) sa chápu pasívne, nemajú žiadny vplyv na študovaný systém.

V schéme kvantovej mechaniky sa ľudskí pozorovatelia chápu ako aktívny, participujúci. Participácia agentov nezávisí na malosti systému, ktorý sa študuje. Ak agent pozoruje nejaké črty fyzikálne popísaného sveta skrz makroskopické vlastnosti meracieho prístroja, aj vtedy participuje. Citlivosť chovania prístrojov k mikroskopickým súčastiam systému sa odráža, prenáša na samotný prístroj, a z neho na pozorujúceho agenta takým spôsobom, že ak urobí výber dotýkajúci sa toho čo skúmať, aký druh znalosti treba hľadať, potom tento výber môže ovplyvniť podstatným spôsobom informáciu, ktorú môže tento agent prijať alebo iný agent, ktorý s ním môže komunikovať o tomto výbere.

Takto výber aký urobí agent na makroskopickej, praktickej úrovni hlbokým spôsobom ovplyvňuje skúmaný fyzikálny systém. Spôsob akým sa robí tento výber nie je ale súčasťou teórie, v konkrétnom prípade - kvantovej mechaniky. Predtým ako sa pokúsime pokročiť ďalej prediskutujeme už spomenuté vzťahy elementárnych, abstraktných častí a percepovaných vlastností.

3.2.6.1 Časť a celok v klasickej fyzike

Aby sme mohli jednoznačne prediskutovať vzťah časť - celok v klasickej fyzike, vzťah medzi zloženým systémom a jeho časťami, podsystémami, vzhľadom k prepojeniu ich vlastností - musíme formulovať, čo je to zložený systém v stavovom priestore klasickej mechaniky. V klasickej mechanike stav zloženého systému z N bodových častíc je definovaný všetkými pármi zovšeobecnených súradníc. Nech $\omega_i = \{q_i, p_i\}$ sú stavy zložiek nášho systému, potom $\omega = (\omega_1, \omega_2, ..., \omega_N)$ usporiadaná N-tica pre každý čas t je stavom pre N častíc. Potom každá vlastnosť zloženého systému v čase t, ak je zadaná pomocou ω, je určená pomocou ω_i.

Napríklad veličiny ako celková hybnosť, celkový moment hybnosti, celková kinetická energia, celkové ťažisko, atď. pre zložený systém sú určené pomocou odpovedajúcich veličín častí, z teoretickej mechaniky, vieme, že jednoducho sú súčtom odpovedajúcich veličín podsystémov. Takto, inými slovami, sú určené stavmi podsystémov.

Všetky tieto vlastnosti vyplývajú zo zákonov zachovania a požiadaviek aditivity na zachovávajúce sa mechanické veličiny. Odvodené veličiny, ako je napríklad gravitačná potenciálna energia, sa dajú obecne vypočítať pomocou základných veličín. Fakticky, každá odvodená veličina v klasickej mechanike sa dá definovať pomocou aditívnych funkcií pre jeden hmotný bod, ktorej hodnoty sú dobre určené v každom bode a čase.

Klasický svet si tak môžeme predstaviť skladajúci sa zo separovateľných, odlíšiteľných častí, ktoré interagujú pomocou síl. Celé toto sa dá popísať pomocou Hamiltonovej funkcie. Ak je táto funkcia známa, potom maximálna znalosť o fyzikálnych veličinách prináležiacich týmto častiam vedie k úplnej, vyčerpávajúcej znalosti o celom, zloženom systéme. Môžeme to formulovať aj trochu formálnejšie ako princíp klasickej separability: Stavy nejakých priestorovo - časovo separovaných podsystémov $S_1, S_2, ... S_N$ nejakého zloženého systému S sú individuálne definované, určené a stavy zloženého systému sú celé, úplne určené pomocou nich a ich fyzikálnych interakcií, ktoré zahrňujú ich priestoro - časové vzťahy.

V klasickej štatistickej mechanike stav ω nejakého individuálneho klasického systému nemôžeme poznať s absolútnou presnosťou. Preto zavádzame nejakú pravdepodobnostnú mieru $\mu(F)$ na podmnožine fázového priestoru F, ktorú interpretujeme ako pravdepodobnosť, že individuálny stav ω sa nachádza pravdepodobnejšie v podmnožine F fázového priestoru než v iných podmnožinách. Čiže, klasický štatistický stav vyjadruje iba odhad pre stav, a miera μ vyjadruje neurčitosť našich odhadov.

Dá sa ukázať, že ak máme daný stav zloženého systému μ, potom stavy $\{\mu_i\}$, i= 1, 2, ..., N jeho podsystémov sú definované, dajú sa určiť jednoznačne skrz μ. Inými slovami, každý klasický štatistický stav dovoľuje nejaké jednoznačné rozloženie, dekompozíciu na vzájomne oddelené stavy, ktoré sa dajú interpretovať ako naša znalosť o systéme resp. systémoch.

V klasických teóriach poľa, napr. elektromagnetizmus, včítane obecnej teórie relativity, máme podobnú situáciu, *Moller, C.* (1972). Bez ohľadu na fyzikálny obsah a matematický formalizmus, hodnoty polí sú dobre definované v každom bode priestoru danej teórie poľa. Napríklad, úplná znalosť metrického tenzora v každom bode priestoro - času, úplne určuje gravitačné pole v tejto oblasti. Teda existencia poľa v nejakej oblasti je obsiahnutá v jeho častiach t.j. bodoch, inými slovami každému bodu primeraného priestoru (variéty v tomto prípade) je priradený fyzikálny stav a tento určuje lokálne vlastnosti takéhoto bodového systému jednoznačne a určito. Následne tak pre klasické teórie poľa platí tiež náš princíp separability.

3.2.6.2 Časť a celok v kvantovej fyzike

Na rozdiel od klasickej fyziky štandardná kvantová mechanika (a kopenhagenská interpretácia – v jej nie metafyzickom zmysle, t.j. interpretácia ako priradenie elementov fyzikálnej reality modelu – kvantovému formalizmu, nech sa v prvom priblížení chápe pod realitou klasická realita) nie je v súlade s našou koncepciou separability. Zdrojom tohto rozdielu, s ďalekosiahlymi dôsledkami, je štruktúra kvantovej mechaniky daná hilbertovým priestorom a princíp superpozície.

Kvôli jednoduchosti budeme uvažovať iba systém S skladajúci sa z dvoch častí S_1 a S_2. V kvantovej mechanike každému stavu zloženého systému odpovedá matica hustoty ρ, pozitívny operátor so spurom rovným 1. Konkrétne každému čistému stavu $|\Psi>$ odpovedá operátor hustoty taký že $\rho = \rho^2$, a síce projekčný operátor $P = |\Psi><\Psi|$, ktorý prevádza projekcie do podpriestorov $H|_{|\Psi>}$. Ak potom ρ_1 a ρ_2 sú matice hustoty odpovedajúce systémom S_1 a S_2 dvojzložkového systému S, potom stav pre S sa dá reprezentovať maticou hustoty $\rho = \rho_1 \otimes \rho_2$ na priestore tenzorového súčinu $H_1 \otimes H_2$. A podstatnou vecou z hľadiska separability tu je, že $H_1 \otimes H_2$ nie je totožné alebo ohraničené kartézskym súčinom H_1 a H_2 ale ich kartézsky súčin zahrňuje ako vlastnú podmnožinu. Inými slovami, platí, že všetky vektory tvaru $|\Psi_i> \otimes |\Phi_j>$, pričom $\{|\Psi_i>\} \in H_1$, $\{|\Phi_j>\} \in H_2$, sú tvaru tenzorového súčinu, ale naopak, nie všetky vektory z $H_1 \otimes H_2$ sa dajú vyjadriť v tomto tvare. Napríklad, z princípu superpozície do $H_1 \otimes H_2$ patrí aj lineárna kombinácia $|\Psi_i> \otimes |\Phi_j> + |\Xi_i> \otimes |\Gamma_j> + ...$, ktorá sa ale, obecne hovoriac nedá faktorizovať do nejakého jedného súčinu. Platí totiž, že každý vektor $|\Psi> \in H_1 \otimes H_2$ sa dá napísať ako $|\Psi> = \Sigma c_{ij} |\Psi_i> \otimes |\Phi_j>$. Z toho, ale nevyplýva, že existujú také dva súbory komplexných čísiel $\{a_i\}$ a $\{b_j\}$, že $c_{ij} = a_i b_j$.

Stavy, ktoré sa dajú takto rozložiť sa nazývajú súčinové, opačne sú to známe previazané (entangled) stavy Schrödingera, *Wheeler, Zurek* (1983). Von Neumann, *Wheeler, Zurek* (1983) preukázal, že v kvantovej mechanike sa dá zložený systém jednoznačne rozložiť do jeho podsystémov iba vtedy a len vtedy ak má stav zloženého systému tvar tenzorového súčinu. V takom prípade korelácie medzi fyzikálnymi veličinami našich dvoch podsystémov neexistujú a stav systému sa dá vyjadriť ako súčet stavov podsystémov, inými slovami každý podsystém má svoju nezávislosť – separovaný a dobre definovaný stav. V tomto prípade, a jedine v tomto prípade je stav celku redukovateľný na stav častí, v súlade s naším princípom separability.

Takéto stavy sú ale veľmi zriedkavé, skôr výnimkou v kvantovej mechanike. Aj keď v kvantovej mechanike platí pre nejaký čas t, že zložený stav sa dá opísať pomocou súčinového stavu, časová evolúcia takéhoto stavu, skrz Schrödingerovu rovnicu (aby sa zachoval tento zriedkavý tvar) je možná iba pre Hamiltonián špeciálneho tvaru. A síce súčet hamiltoniánov daných podsystémov, $H = H_1 + H_2$, čo je vlastne podmienka neinteragovania dvoch systémov S_1 a S_2. Časová evolúcia zloženého systému je celkom určená unitárnou Schrödingerovou dynamikou, ktorá transformuje čisté stavy na čisté stavy, ale nie súčinové stavy na súčinové. Najmenšia možná interakcia hociktorého z podsystémov systému s jeho okolím, povedie k vzniku previazaného stavu celkového systému.

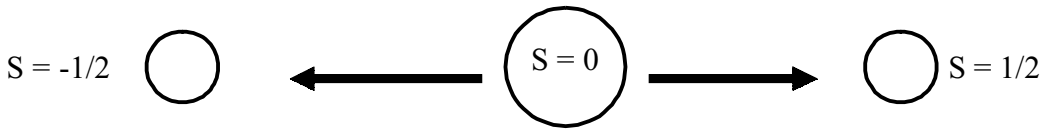
Naopak, majme náš systém, kde v nejakom čase t_0 systémy S_1 a S_2 interagovali, potom predpokladajme že v čase $t > t_0$ sú priestorovo separované. Podľa Schrödingera, *Wheeler, Zurek* (1983) potom ľubovolný čistý stav zloženého systému S sa dá vyjadriť pomocou súčinových stavov ako

$$|\Psi> = \Sigma c_j |\Psi_i> \otimes |\Phi_j> \text{ , kde } ||\Psi||^2 = \Sigma |c_j|^2 = 1$$

a $\{|\Psi_i>\}$ a $\{|\Phi_j>\}$ sú ortonormálné bázové vektory v H_1 a H_2. Ak $|c_j| < 1$ potom existuje korelácia medzi podsystémami S_1 a S_2. Maximálna znalosť celého systému, preto nedovoľuje aby sme zistili maximálnu znalosť o jeho častiach, čo nemá precendens v klasickej fyzike, a čo tak trápilo A. Einšteina, *Wheeler, Zurek* (1983).

Opačne, ak uvažujeme previazaný zložený systém S tak každému podsystému S_1 (S_2) môžeme „priradiť" stav, iba pomocou odvolania sa na partnerský podsystém S_2 (S_1) skrz úplnú informáciu

obsiahnutú v S. Aby sme to vysvetlili bližšie uvažujme prototyp EPR korelovaného systému, spin – 0 pár , alebo singletné páry častíc, Obr. 26.



Obr. 26 Schématické usporiadanie častí a celku v EPR experimente

Takže majme pár častíc so spinom 1/2 v singletnom stave

$$|\Psi> = (1/\sqrt{2})\ \{|\Psi_{+}> \otimes |\Phi_{-}> \ - |\Psi_{-}> \otimes |\Phi_{+}> \}$$

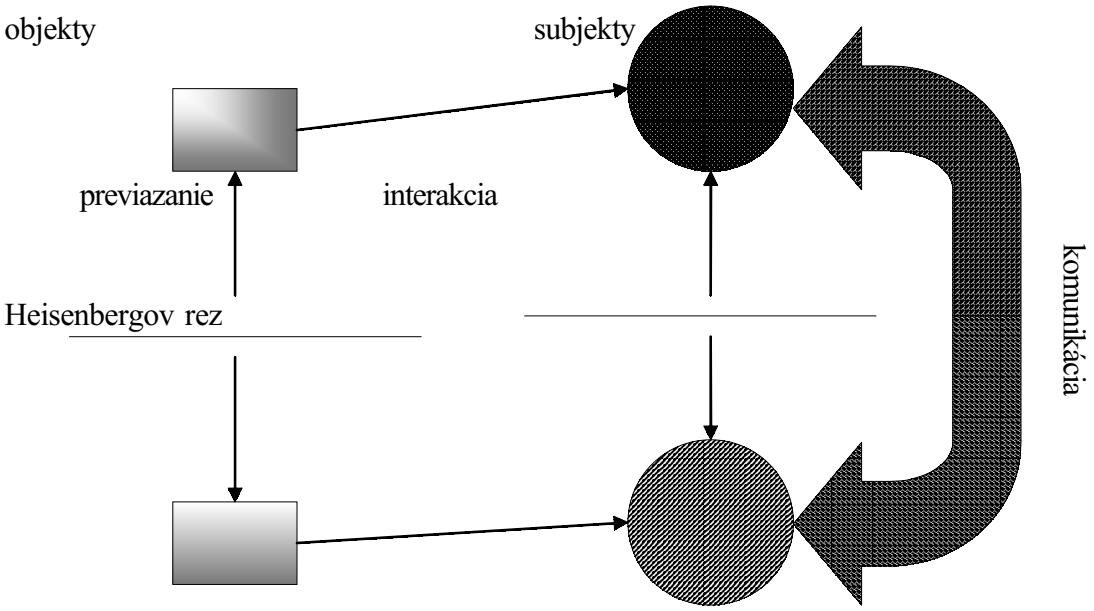
kde $\{|\Psi_{\pm}>\}$ a $\{F_{\pm}>\}$ sú ortonormované bázy dvoj rozmerných hilbertovských priestorov H_1 a H_2 asociovaných s S_1 a S_2. Kvantová mechanika predikovala a experimentálne potvrdila, že S_1 a S_2 sú vždy spinovo opačne orientované, a úplne antikorelované. Ak pre prvý podsystém nameriame v spin +1/2, potrom druhý má -1/2 a naopak. Kvôli rotačnej invariancii singletného stavu to platí pre ľubovolný smer merania. Takto EPR stavy častíc S_1, S_2 jasne vyjadrujú nesaparovateľnosť. Častice S_1 a S_2, niekedy v minulosti interagovali a vytvorili zložený systém S, po oddelení na veľkú vzdialenosť, takto už neiteragujúc, zostali korelované (v našom prípade antikorelované). Podsystémy S_1 a S_2 sú korelované takým spôsobom, ktorý je zakódovaný v stave $|\Psi>$ celého systému.

Pomocou pojmu komplementarity to môžeme vyjadriť nasledovne: časť sa „prejavuje“ skrz celok, a celok sa dá „odhadnúť“ skrz nezávislé chovanie jeho častí. V našom EPR príklade, singletného systému, napríklad žiadny podsystém nemá určitú hodnotu spinu v nejakom smere, a následne iba vlastnosť celkového spinu zloženého systému je tá vlastnosť pomocou ktorej môžeme povedať všetko o spinových vlastnostiach podsystémov (iba v stave celého systému je zakódovaná korelácia medzi rozdeleniami pravdepodobnosti). Previazané korelácie v kvantovej mechanike sa nedajú vysvetliť pomocou nejakých predpísaných vzťahov alebo interakcií medzi odpovedajúcimi časťami. Interakcia je postačujúcou ale nie nutnou podmienkou previazania. Takto existujú jasne špecifikované vlastnosti previazaného kvantového systému, ktoré sa nedajú redukovať, ani nevyplývajú, ani príčinne nezávisia na lokálnych vlastnostiach častí, podsystémov.

Môžeme usúdiť, že zo zadaných (takto aj existujúcich) podsystémov môžeme jednoznačne skonštruovať zložený systém (pomocou stavov tenzorového súčinu). Opačne, jednoznačnosť konštrukcie, alebo čo len existencia, rozloženia celkového systému na zložky nie je garantovaná, v rámci formalizmu štandardnej kvantovej mechaniky. Otázka existencie rozkladu predpokladá separáciu medzi pozorovateľom (primerané meriacie zariadenie spolu s okolím) a pozorovaným, poznávajúcim subjektom a objektom poznania.

Z fundamentálneho hľadiska sa pre kvantovú teóriu javí svet ako nedeliteľný celok. Nie je apriori separovaný. Musí ho separovať pozorovateľ, aby dostal nejaký popis, aby mohol komunikovať o objekte, alebo zaznamenávať experimentálne dostupné fakty. A táto separácia musí byť do vzájomne pôsobiacich ale nepreviazaných podsystémov, meraných objektov a nekorelovaných pozorovateľov ((primerané meriacie zariadenie spolu s okolím), Obr. 27.

Toto rozdelenie sa nazýva *Heisenbergov* rez (alebo *von Neumanov*). V klasickej fyzike sú tieto podmienky splnené automaticky, avšak v neseparovateľných teóriach ako je kvantová mechanika je Heisenbergov rez, principiálnou nutnosťou, pretože každá možnosť prevedenia a opakovania kontrolovaných experimentov predpokladá existenciu rozdelenia subjekt objekt. Ako hovorí *Heisenberg* (1966): „to čo pozorujeme nie je príroda, svet ako taký, ale príroda vystavená našej metóde pýtania sa“. Teda, čo definujeme ako „časť“ alebo ako podsystém, v kvantovo mechanickom formalizme nie je čistý stav, je jednoducho nejaký konkrétny vzor, ktorý vznikne rozdelením, alebo abstrakciou voči zvyšku sveta.



Obr. 27 Heisenbergov rez a rozdelenie objekt a subjekt

A toto sa dá urobiť rôznymi spôsobmi, v závislosti od vybraného kontextu výskumu, pýtania sa, teórie. Naviac, toto rozdelenie „vytvára" možné podsystémy pomocou experimentálneho pozorovania, ktoré zároveň potlačuje previazané korelácie medzi nimi, čím môžeme priradiť takýmto podsystémom nejakú úroveň separovanej reality, ktorej elementy zakúšame obecne ako odlíšiteľné, dobre lokalizované objekty.

Ak previazané stavy interpretujeme ako súbor potencialít, potom tieto kvantovo mechanické potenciality sú fyzikálne reálne, objektívne v zmysle priradenia nejakému objektu schopnosť prejavovať vlastnosti za určitých daných dobre definovaných podmienok, alebo schopnosťou za istých podmienok interferovať navzájom. Samozrejme to neznamená, že sú možné ľubovolné rozloženia.

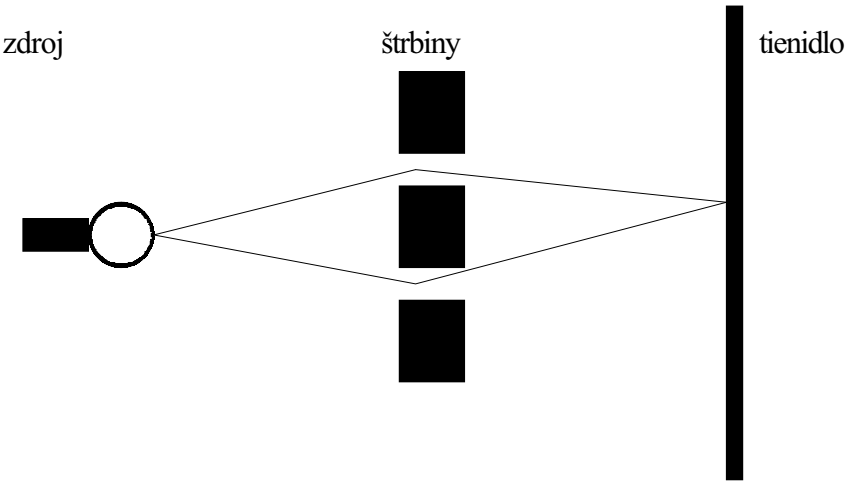
Existujú dynamické obmedzenia, obmedzenia symetrií a obmedzenia dané počiatočnými podmienkami, ktoré kladú podmienky na výber nášho rozloženia (na stavy ako tenzorové súčiny). Napríklad, náš EPR pár pred meraním, sa dá prirodzene rozložiť na redukované stavy dvoch elektrónov. V obecnom prípade, hlavne pre systémy s nekonečne veľkým počtom stupňov voľnosti (kvantové teórie poľa), existuje oveľa väčšia nejednoznačnosť rozloženia do množstva dostupných a fyzikálne nepodobných rozkladov celku na časti.

3.2.6.3 Vzťah neseparovateľnosti k obecnej percepcii

Z predošlých úvah môžeme usúdiť, že v prírode neexistujú žiadne separované systémy, okrem samotného sveta. Z princípu superpozície vyplýva, že môžeme superponovať ľubovolný počet čistých stavov v hilbertovom priestore, aby sme vygenerovali nejaký nový čistý stav, inými slovami, ak máme $|\Psi_1>$, $|\Psi_2>$, ... ľubovolný počet jednotkových vektorov v H, opisujúcich kvantový systém, potom hocijaká lineárna kombinácia

$$|\Psi> = c_1|\Psi_1> + c_2|\Psi_1> + \ldots \text{, pričom } c_i \in C, \Sigma|c_i| = 1$$

tvorí čistý stav v H, ktorý reprezentuje nejakú fyzikálne možný stav systému. Stavy $|\Psi_i>$ sa explicitne interpretujú ako potenciálne realizovateľné, s amplitúdou pravdepodobnosti c_i. Princíp superpozície je priamo spojený s javom interferencie týchto amplitúd, Obr. 27, ktorý odráža povahu vzťahov medzi stavmi kvantového systému, čo sa dá vyjadriť nasledovne.



Obr. 27 Superpozícia stavov na dvojštrbine a interferencia

Fyzikálna premenná A, ktorá je vo vzťahu so superponovaným stavom $|\Psi>$ nemá určitú hodnotu. Inými slovami, pre fyzikálnu premennú A v superponovanom stave $|\Psi>$ vlastných stavov A a hocijaké tvrdenie *T* o veličine A, neplatí že buď *T* alebo 1 – *T*. Hovoríme, že v takých stavoch sú možné hodnoty A objektívne neurčité, a nie len neznáme.

V klasickej fyzike máme tiež superpozície, ale čo je podstatný rozdiel, stavy sú aktuálne a nie potenciálne, všetko to čo je potenciálne možné sa aj realizuje niekedy v čase, a nezávisle od intervencie merania. V kvantovej mechanike potencionality podmieňujú, ale nekontrolujú produkciu aktuálnych javov. Na druhej strane kvantové potencionality sú fyzikálne reálne a objektívne nielen tým, že prepožičiavajú objektu nejaké vlastnosti za určitých podmienok, ale aj za určitých podmienok interferujú navzájom, ako pri koherentných javoch.

Z fundamentálneho, či filozofického alebo fyzikálneho hľadiska, kvantovo mechanická neseparovateľnosť je v rozpore s realitou nezávislou od kontextu, alebo pozorovateľa resp. mysle. Vyplýva to z toho, že Heisenbergov rez je nutne kontextuálny. Ak si tento rez zvolíme, potom aby sme zaznamenali objektívnu informáciu o objekte, je nutné aby meracie zariadenia boli nezávislé (kinematicky), pretože ak by platil opak a stavy meracieho zariadenia by boli previazané s objektom merania, potom by získaná informácia bola relatívnou. Každá definícia pojmov by bola postihnutá takýmto nekonečným regresom závislosti merania od kontextu. Presná formulácia pojmov vyžaduje aby bolo meracie zariadenie – poznávajúci subjekt jasne oddelené od obsahu, ktorý reprezentujú – objekty poznania.

Subjekt sa nedá pokladať za ďalší z objektov. Musí existovať jasné, aj keď možno umelé rozdelenie medzi nimi (objektom a subjektom). Len toto rozdelenie nám dovoľuje akty abstrakcie, diasociácie, stabilizácie a nakoniec registrácie – niekedy sa tieto akty nazývajú obecne klasifikácia, spomínané v predošlej diskusii. Fyzikálny systém sa môže pokladať za meracie zariadenie (predĺženie poznávajúceho subjektu) iba vtedy, ak nie je previazaný s objektom merania. Ak pokladáme kvantový svet za jeden celok, potom pri položení konkrétnej otázky spolu s náležiacim kontextom sa tento celok rozpadá na zdanlivé časti.

Takto realite nezávislej na mysli odpovedá kvantový svet ako celok, a môžeme sa naň odvolávať ako na vnútornú úroveň reality, na druhej strane zavedenie kontextu je relevantné za vonkajšiu úroveň reality, empirickú realitu, ktorá je výsledkom abstrakcie pri ľudskej percepcii. Percepovateľná separovateľnosť a lokalizovateľnosť kontextuálnych objektov empirickej reality sa dosahuje prerušením, abstrakciou (faktuálne) existujúcich previazaných korelácií objektu s jeho okolím.

Formulujúc to ináč, ak v mikrosvete chceme určiť objekt, jeho vlastnosti závislé od stavu, potom to má zmysel iba pre daný kontext, čiže objekty v kvantovej mechanike sú entity závislé od kontextu. Takto objekty v kvantovej mechanike sú konštruované, fenomenálne. Ale na druhej strane

nie sú vynálezmi ľudskej mysle samotnej, ani nie sú noumenálne veličiny v zmysle Kanta. Odrážajú objektívne štrukturálne aspekty fyzikálnej reality z hľadiska nejakého dopredu vybraného kontextu, *Heisenberg* (1966), (1969).

Celý tento spôsob uvažovania sa dá nazvať participujúci realizmus, *Wheeller* (1983). Participujúci - znamená aktívny, pretože myslenie prispieva do skúsenosti, pripúšťa účasť, aktívnu úlohu poznávajúceho subjektu pri tvorení kontextu komunikácie, identifikácie konkrétneho vzoru ako objektu, závisí na našej znalosti a predvýberu experimentálneho kontextu. Realizmus tu znamená, že ak máme daný kontext, konkrétne objekty – štruktúry reality, potom máme dobre definované vlastnosti nezávislé od ich poznania. Je to kategória skôr vo funkcionálnom význame, v zmysle, že poznávajúci subjekt má špecifickú, nezameniteľnú úlohu, pretože vonkajšia realita nie je percepovaná ako niečo dané apriori, hotové, dané nejakým externým hľadiskom, ale ako niečo ovplyvnené pôsobením subjektu a tým aj vlastnosťami subjektu. Participujúci realizmus najlepšie vystihuje Wheellerovská hra, ktorej verziu (upravený preklad) z *Wheeller* (1983) uvádzame na záver: *Vypočujme si historku o hre s dvadsiatimi otázkami. Pravidlá sú jednoduché - jeden účastník večierka sa pošle von z miestnosti, ostatní sa dohovorili na nejakom slove, ten jeden sa vráti do miestnosti a začne sa pýtať: „Je to živé?“ „Nie.“ „Je to tu na Zemi?“ „Áno.“ Tak otázky idú od jedného k druhému okolo, dokiaľ nie je hľadané slovo uhádnuté. Opytovateľ zvíťazí, ak mu stačilo dvadsať otázok alebo menej. Zase vás pošlú von a čakáte tam neuveriteľne dlho. Nakoniec, keď vás znovu vpustia, všetci sa usmievajú. Začnete klásť svoje otázky. Odpovede prichádzajú najskôr rýchlo. Potom sa ale každé čakanie začína predlžovať - divné, keď samotná odpoveď je vždy len jednoduché „áno“ alebo „nie“. Nakoniec sa opýtate, či to slovo je „krava“. „Áno“, znie odpoveď a všetci sa smejú. Vysvetľujú, že keď ste boli vonku, dohodli sa nedohovoriť dopredu žiadne konkrétne slovo. Každý v kruhu mohol na akúkoľvek vašu otázku odpovedať „áno“ i „nie“, ako sa mu chcelo. Ale keď odpovedal, musel mať na mysli nejaké slovo zlúčiteľné so svojou odpoveďou - a tiež zlúčiteľné s odpoveďami na všetky vaše predošlé otázky. Nie je divu, že rozhodnutia medzi „áno“ a „nie“ boli stále ťažšie!*

4 Model percepcie invariantných čŕt

V prvej časti tejto kapitoly sa pokúsime o všeobecný abstraktnejší pohľad na rečovú percepciu z hľadiska niektorých myšlienok spracovania informácie a úlohe invariantných čŕt v modeloch percepcie a rozpoznávania. Matematicky opíšeme konkrétne stupne nášho modelu neurónového fonetického spracovania percepcie. V druhej časti opíšeme učiaci mechanizmus pre vytvorenie prototypov daných fonetických jednotiek, pre konkrétny typ úlohy, a navrhnutie adekvátneho klasifikačného mechanizmu vektorov čŕt na jeden z prototypov (z množiny naučených vektorov čŕt - vzorov).

4.1 Symetrie pri percepcii reči

Z hľadiska spracovania informácie ľudská percepcia a rozpoznávanie sa dajú považovať za špeciálny nástroj, ktorý kompresuje rýchlosť informácie z približne 2^{16} bit/s na 2^8 bit/s pre izolované slovo. Tento odhad vyplýva z nasledujúcich výrazov pre rýchlosť informácie v analógovej alebo digitálnej forme

$$\frac{\Delta I}{\Delta t} \approx BW \log_2\{1 + S/N\} \qquad \text{pre analógovú formu}$$

$$\frac{\Delta I}{\Delta t} \approx 2BW \log_2\left\{2^{A/D}\right\} \qquad \text{pre digitálnu formu} \tag{55}$$

kde $\Delta I / \Delta t$ je rýchlosť informácie alebo tok informácie (bit/s); *BW* je šírka frekvenčného pásma vstupného signálu (Hz); *S/N* je pomer signálu k šumu rečového signálu a *A/D* je počet bitov použitých na konverziu analógového signálu na digitálny signál.

Horná hranica je tok informácie, ktorý je prenesený z vonkajšieho priestoru do ucha prostredníctvom zvukovýh vĺn. Pre vonkajší rečový signál to dáva približne 2^{16} bitov za sekundu.

Dolná hranica hranica vyplýva z nasledujúcich úvah. Ak chceme opísať fonémické - podobné fonémam - jednotky pomocou binárných dištinktívnych čŕt, *Reddy* (1975), potrebujeme na to približne šesť bitov (počet foném je priemerne 40). Ďalej máme okolo 10 fonémických jednotiek pre jednotlivé slovo približne jednu sekundu dlhé. Čo dáva približne tok informácie 2^8 bitov za sekundu *O´Shaugnessy* (1990), *Pickles* (1988), *Flanagan* (1972).

Z tohto môžeme usúdiť, že rečový signál má relatívne vysokú redundanciu. Takto pri návrhu systému percepcie alebo rozpoznávania, musíme zabezpečiť aby kompresia informačného toku nezmenila podstatným spôsobom relevantné fonetické vlastnosti rečového signálu, ktoré sú dôležité pre rozpoznávanie foném, slov, súvislej reči alebo hovoriaceho. V praxi to znamená, že umelý systém rozpoznávania je tak efektívny a optimálny ako je kompresia informačného toku pôvodneho rečového signálu ku konečnej bitovej reprezentácii foném, slov alebo hovoriaceho. Náš prístup je v podstatnej miere založený na extrakcii invariantných čŕt. Preto v ďalšom uvedieme niekoľko faktov o extrakcii invariantých vlastností v percepcii reči.

4.1.1 Požiadavky percepcie invariantných čŕt

Na základe konceptuálnych úvah v predošlej kapitole, podkapitola 3.2 je v percepcii a rozpoznávaní reči možné a užitočné uvažovať tri druhy invariancií. Prvý typ berie do úvahy obyčajný priestor, alebo vonkajší priestor, resp. priestoro-čas. Takto uvažujeme invariancie vzhľadom k

- relatívnej polohe pozorujúceho systému k zdroju rečového signálu

- relatívnej orientácii pozorujúceho systému k zdroju rečového signálu
- relatívnemu pohybu pozorujúceho systému k zdroju rečového signálu

nazveme ich vonkajšie symetrie percepcie alebo rozpoznávania. Druhý typ berie do úvahy rôzne ekvivalentné spôsoby popisu nielen rečového signálu ale aj jeho detekciu-percepciu.Uvažujeme invariancie vzhľadom k

- intenzite rečového signálu
- akustickému a fonetickému šumu
- základnému tónu rečového signálu
- rýchlosti hovorenia
- celkovému trvaniu rečovej jednotky

Všetky tieto invariancie používame v každodennom živote a sú enormne dôležité pre efektívnu a robustnú rečovú komunikáciu. Poznámka: Vyššie uvedené invariancie naopak neposkytujú priamo evolučnú výhodu, samozrejme nepriamo skrz nadobudnutú schopnosť komunikovať ju poskytujú. Vieme, že pri percepcii reči hrajú dôležitú úlohu aj okamžitá energia, dôraz a iné foneticky dôležité parametre, časť 2.1, tu sa pokúsime o návrh modelu bez týchto čŕt a pozrieť sa pokiaľ nás to dovedie.

Na oveľa fundamentálnejšej úrovni tieto invariancie nás nasmerujú k možným vnútorným mechanizmom rečovej percepcie a rozpoznávania, ktoré môžeme potom použiť pre návrh počítačového systému. Je to do istej miery analogické k situácii, pri ktorej z invariancie vzhľadom ku transformáciam z danej grupy symetrií môžeme skonštruovať diferenciálne rovnice - invariantné vzhľadom k týmto transformáciam *Bhagavantam, Venkatarayudu* (1951). Navrhnutý model, ktorý prediskutujeme bližšie v nasledujúcich častiach, sa dá považovať za mechanizmus extrakcie týchto (vymenovaných v predošlom odstavci) invariantných čŕt.

Tretí typ invariancií nazveme vnútorné a odrážajú sa v symetriách fonémického systému, čo prediskutujeme v ďalšej kapitole. V súčasnom štádiu výskumu, sme tieto vnútorné symetrie nezapracovali do navrhovaného modelu. Avšak, považujeme zahrnutie týchto symetrií do modelu za kriticky dôležité a v konečnom dôsledku zodpovedné za ľudskú percepciu a rozpoznávanie reči, pozri predchádzajúcu kapitolu a nasledujúcu, ďalšiu kapitolu a diskusiu.

4.2 Model topologických invariantov

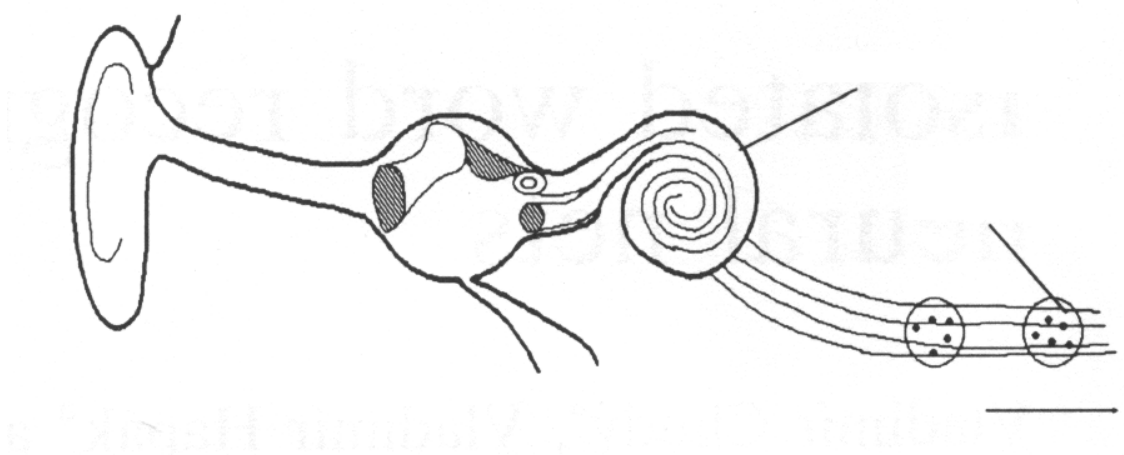
Pri návrhu nášho modelu sledujeme fyziologické mechanizmy percepcie reči. Takže každá fyziologická časť, ktorá prevádza dôležitú operáciu s rečovým signálom (ilustrované na Obr. 29) odpovedá približne následnej časti modelu NP4 - Neural Parrot-like Perception PreProcessor. Preto konštrujeme NP4 schématicky ako na Obr. 30, ktorý sleduje nákres na Obr. 29. Na hornej časti tohto obrázku sme označili fyziologicky dôležité časti, v strednej časti sú umelé analógie spracovania. V spodnej časti máme informačné rýchlosti pre jednotlivé úrovne spracovania

Na tomto mieste musíme samozrejme zdôrazniť, že fyziologická podobnosť s navrhnutým modelom je iba kvalitatívna. Viac informácie sa dá najsť v *Witten* (1982), *Nobili, Mammano, Ashomore* (1998). V tejto časti nebudeme definovať konkrétne hodnoty parametrov, ktoré vystupujú v jednotlivých procesoch, algoritmoch. Numerické výsledky budeme prezentovať v kapitole venovanej výsledkom.

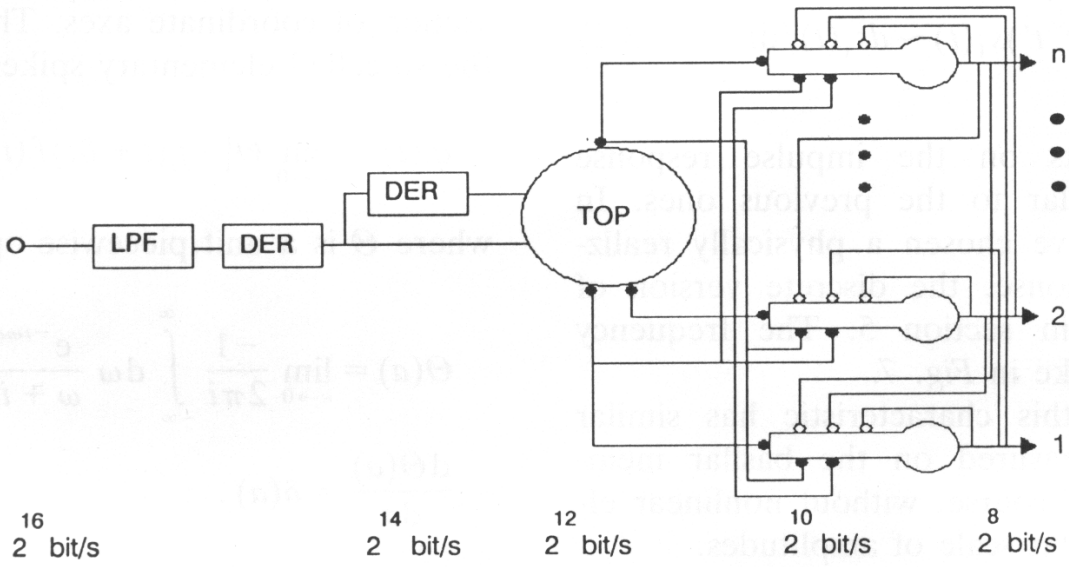
Teraz matematicky opíšeme konkrétne časti NP4 modelu.

1. Uvažujme rečový signál s(t) ako funkciu času. Prvou časťou NP4 je dolnopriepusťový filter

$$s_1(t) = \int h_1(t,t')s(t')dt' \qquad (56)$$



Obr. 29 Schéma ucha



Obr. 30 Schéma NP4 modelu

kde $h_1(t, t')$ je impulzná charakteristika dolnopriepusťového filtra. Táto impulzná charakteristika musí byť lineárna, stacionárna a jej fázová charakteristika tiež musí byť lineárna. Systémovou realizáciou takéhoto filtra je takzvaný nerekurzívny filter. Výstup z tohto filtra je $s_1(t)$. Táto časť modeluje stredné ucho.

2. Nasledujúca časť opisuje funkciu bazilárnej mambrány. Najskôr opíšeme model podobný papagájovi, bez explicitnej frekvenčnej selekcie

$$s_2(t) = \int h_2(t,t')s_1(t')dt' \approx \frac{ds_1}{dt} \tag{57}$$

Požiadavky na impulznú charakteristiku $h_2(t,t')$ sú podobné ako pri predošlej charakteristike. Pri simulácii sme vybrali fyzikálne realizovatelnú, kauzálnu odozvu, ktorej diskrétna verzia je daná Stirlingovou transformáciou, pozri časť venovanú časovým charakteristikám modelu. Takisto model s explicitnou frekvenčnou závislosťou (viac podobný ľudskej percepcii a nie percepcii papagája) je uvedený v podkapitole venovanej realizácii časových a frekvenčných charakteristík modelu.

3. Tretia časť opisuje fungovanie práve jednej vlasovej bunky, ktorá má dva typy výstupov. Prvý typ x(t) exaktne kopíruje výstupný potenciál bazilárnej membrány $s_2(t)$. Druhý výstup y(t) prevádza približne deriváciu tohto potenciálu

$$x(t) = s_2(t) = \int dt'' \int h_2(t,t')h_1(t',t'')s(t'')dt' \tag{58}$$

$$y(t) = \int dt''' \int dt'' \int h_2(t,t')h_1(t',t'')h_0(t'',t''')s(t'')dt'$$

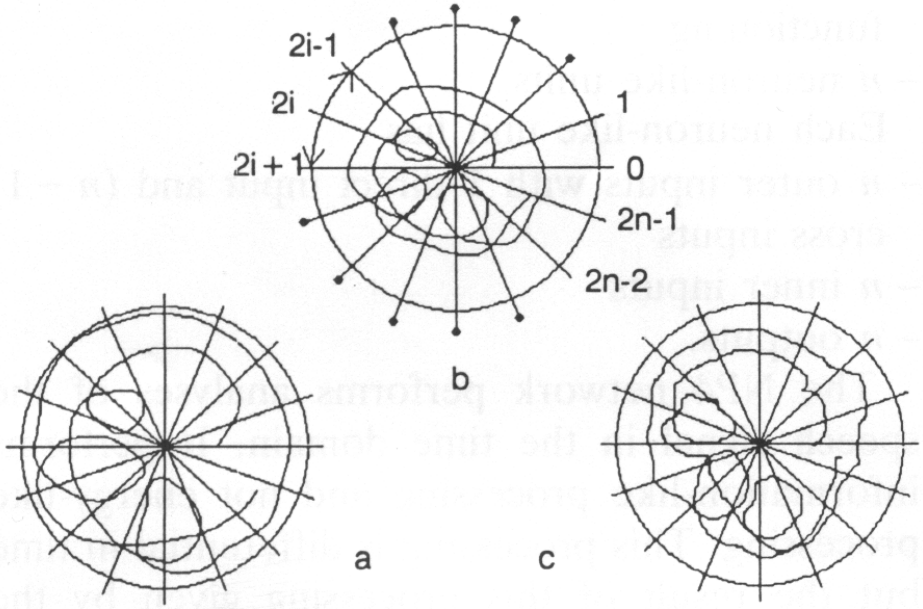
4. Vplyv synaptických ukončení a sluchových nervov opíšeme nelineárnym operátorom Θ, ktorý operuje na oboch funkciách x(t) aj y(t) a jeho výstupy sú iba hodnoty 1 a 0. Aby sme lepšie pochopili chovanie sa tohto operátora skonštrujeme takzvanú fázovú rovinu z x(t) a y(t) ako je ilustrované na Obr. 31. Táto trajektória sa nazýva Nyquistova trajektória, alebo Nyquistov graf. V tomto grafe definujeme 2n rovnomerne vzdialených osí, každá z ktorých začína v počiatku súradnicového systému. Potom môžeme definovať predpokladaný výstup pomocou takzaného elementárneho generátora spajkov

$$q_i(t) = \lim_{\delta t \to 0} \Theta[-f_i(t + \delta t)f_i(t) \tag{59}$$

kde Θ je jednotkový operátor definovaný ako

$$\Theta[a] = \lim_{\varepsilon \to 0} \frac{-1}{2i\pi} \int_{-\infty}^{\infty} d\omega \, \frac{e^{-ia\omega}}{\omega + i\varepsilon} = \begin{cases} 1 \to a > 0 \\ 0 \to a < 0 \end{cases}$$

$$\frac{d\Theta[a]}{da} = \delta(a)$$



Obr. 31 Fázová rovina x(t) a y(t)

Tu je i je rovné $(-1)^{1/2}$a δ(a) je Diracov operátor a čas dt sa dá interpretovať ako refrakčný čas nervov a v digitálnom modeli ho môžeme prirovnať vzorkovacej frekvencii. Premenná $f_i(t)$ je definovaná ako

$$f_i(t) = \text{arctg}\left(\frac{x(t) - y(t)\alpha_{2i}}{y(t) - x(t)\alpha_{2i}}\right) \tag{60}$$

kde x(t), y(t) sú dané v (4.2.3) a arctg(α_n) je uhol 2i-tej osi. Dá sa vidieť, že fyzikálny význam tohto operátora je generovať jednotkový impulz vždy keď trajektória pretína danú os. Význam faktora 2i opíšeme neskôr. Dôležitý fyzikálny význam predošlého typu spracovania je prevádzať kvázi-topologickú extrakciu čŕt, ktorá je invariantná k intenzite a k šumu. Z Obr. 31 jasne vyplýva, že počet preseknutí danej osi - počet spajkov v týchto troch situáciach, kde signálová trajektória je deformovaná topologickým spôsobom - šumom, zmenou intenzity, atď. - je rovnaký. Inými slovami operácia je invariantná voči spojitým deformáciam fázovej trajektórie, pozri dodatok A2. Práve predošlým mechanizmom modelujeme pálenie neurónov. Matematický význam týchto operácií vysvetlíme v nasledujúcej kapitole.

Náš model má dva typy výstupov. Prvý sme práve opísali, druhý sa zaoberá časovým trvaním, ktorý systém zotrváva v jednotlivom sektore fázovej roviny. Opäť tieto sektory sú definované na Obr. 31, Obr. 32. Každá (párna) os, ktorá sa používa na počítanie preseknutí má svoju oblasť - sektor, ktorý je z druhej strany ohraničený následnou (nepárnou) osou. Matematický popis takéhoto generátora je vcelku podobný predošlému. Definujeme takzvaný elementárny časový generátor ako

$$m_i(t) = \Theta[-g_i(t)g_{i+1}(t)] \tag{61}$$

Význam Θ je rovnaký ako v predošlom a funkcia $g_i(t)$ je definovaná ako

$$g_i(t) = \text{arctg}\left(\frac{x(t) - y(t)\alpha_{2i-1}}{y(t) - x(t)\alpha_{2i-1}} \right) \qquad (62)$$

Fyzikálny význam elementárneho časového generátora vyplýva z jeho interpretácie ako váhového faktora spajk generátora $q_i(t)$ vo vstupe n neurónu podobných jednotiek, ako uvidíme v nasledujúcich častiach. Účelom týchto váh je normalizovať v určitom zmysle časové trvanie a rýchlosť hovorenia. Na trochu viac lokálnejšej úrovni vyjadruje relatívnu pravdepodobnosť vstupného pálenia i-teho neurónu relatívne k druhým neurónom.

5. Prediskutujeme teraz piatu časť nášho modelu, odrážajúcu aktivity siete sluchových neurónov, ktorá sa skladá z n neurónu podobných jednotiek, ktoré vystupujú z nášho modelového „kochleárneho jadra". Pri popise dynamiky tejto siete budeme vychádzať z niektorých obecných ideí *Kohonena* (1988). Takto, aktivitu n neurónov opíšeme nasledujúcimi dynamickými rovnicami

$$d\eta_i / dt = \sum \varphi_j(z_{ij}) - \rho_i(h_i) \qquad (63)$$

kde η_i je výstupná aktivita i-teho neurónu, ζ_{ij} je frekvencia pálenia poskytovaná od nejakého iného neurónu na j-ty vstup i-teho neurónu. Funkcie φ a ρ môžu mať obecný tvar. Ak predpokladáme, že existuje inverzná funkcia k ρ a stacionárne vstupno-výstupné podmienky, môžeme získať

$$\eta_i(t) = \sigma_i[I_i(t) + \gamma_i(t)] \qquad (64)$$

kde $I_i(t)$ je vstupná aktivita i-teho neurónu, γ je offsettová hodnota, hypotetický „prah", σ je takzvaná sigmoidná aktivačná funkcia, (pre viac detajlov pozri *Kohonen* (1988)). Ako sa uvádza aj v Kohonenovi, reálny spúšťací prah závisí na kolektívnom správaní, interakciách neurónov. V tejto práci definujeme túto prahovú funkciu podobným spôsobom ako je laterálna spätná väzba u *Kohonena* (1988). Takže definujeme výstupnú aktivitu i-teho neurónu (neurónu-podobnej jednotky) ako

$$\eta_i(t + \Delta t) = \sigma_i[\Delta I_i(t) + \Sigma \Delta s_{ij}(t)\eta_i(t)] \qquad (65)$$

kde jednou z možných foriem $\Delta I_i(t)$ sa dá napísať ako

$$\Delta I_i(t) = \Delta\zeta_i(t)\Delta\mu_i(t)$$

kde $\Delta\zeta_i(t)$ a $\Delta\mu_i(t)$ sú integrálne alebo sumárne aktivity i-teho generátora spajkov alebo generátora časového trvania počas časového úseku Δt a sú definované ako

$$\Delta\xi_i(t) = \int_t^{t+\Delta t} dt dq_i / dt$$

$$\Delta\mu_i(t) = \int_t^{t+\Delta t} dt dm_i / dt \qquad (66)$$

V tejto práci nediskutujeme, či integrály sú v štandardnom tvare alebo v nejakej zobecnenej forme. Veličiny $\Delta s_{ij}(t)$ sú definované ako

$$\Delta s_{ij}(t) = [\langle \Delta\zeta_i(t)\rangle - \langle \Delta\zeta_j(t)\rangle] / [\langle \Delta\zeta_i(t)\rangle + \langle \Delta\zeta_j(t)\rangle] \qquad (67)$$

kde $\langle \Delta\zeta_i(t)\rangle$ označuje dozadu v čase ustrednenú hodnotu $\zeta_i(t)$, obecne váhovanú. V našom prípade máme potom

$$\langle \Delta\zeta_i(t)\rangle = \sum \Delta\zeta_i(t-k\Delta t)\ w(k\Delta t) \qquad (68)$$

kde váhová funkcia w je obecne nerastúca funkcia argumentu $k\Delta t$ a vyjadruje vplyvy zapamätania. Základný tvar tejto funkcie je

$$w(k\Delta t) = \Theta\ [(a - k)\ \Delta t] / a \qquad (69)$$

Doba rozpadu tejto funkcie Δt sa môže interpretovať ako retardačný čas. Neskôr, v diskusii tiež ukážeme, že Δt môžeme interpretovať ako adaptačný čas sluchových nervov a neurónov. Význam Δs_{ij} vyplýva z nasledujúcich úvah. Existujú tri hraničné hodnoty tejto funkcie

$$\Delta s_{ij}(t) = \begin{cases} 1 & \text{pre } \Delta\zeta_j(t) << \langle \Delta\zeta_i(t)\rangle \\ 0 & \text{pre } \Delta\zeta_j(t) \approx \langle \Delta\zeta_i(t)\rangle \\ -1 & \text{pre } \Delta\zeta_j(t) >> \langle \Delta\zeta_i(t)\rangle \end{cases} \qquad (70)$$

Vidíme, že táto funkcia opisuje stavy, či sa niečo stalo v i-tom neuróne vzhľadom k j-temu neurónu. Takýmto spôsobom neuróny „komunikujú", interagujú s inými neurónmi a medzi sebou. Myšlienka pre výber práve takejto funkcie pochádza z niektorých abstraktných myšlienok z oblasti vizuálnej percepcie a rozpoznávania *O´Shaughnessy* (1990), *Mozer* (1991), *Witten* (1982), *Yang, Wang, Shamma* (1992), *Liu, Andreou, Goldstein* (1991). Veríme, že tieto myšlienky sú dostatočne obecné a aplikovateľné v neurónu podobných systémoch. Taktiež vidíme, že tieto interakcie alebo synapsie nie sú obecne symetrické v indexoch i a j.

Sigmoidná funkcia $\sigma(x)$ v (64) sa dá definovať ako

$$\sigma(x) = \begin{cases} 0 \text{ ak } x \le 0 \\ 1 \text{ ak } x \ge 0 \end{cases} \qquad (71)$$

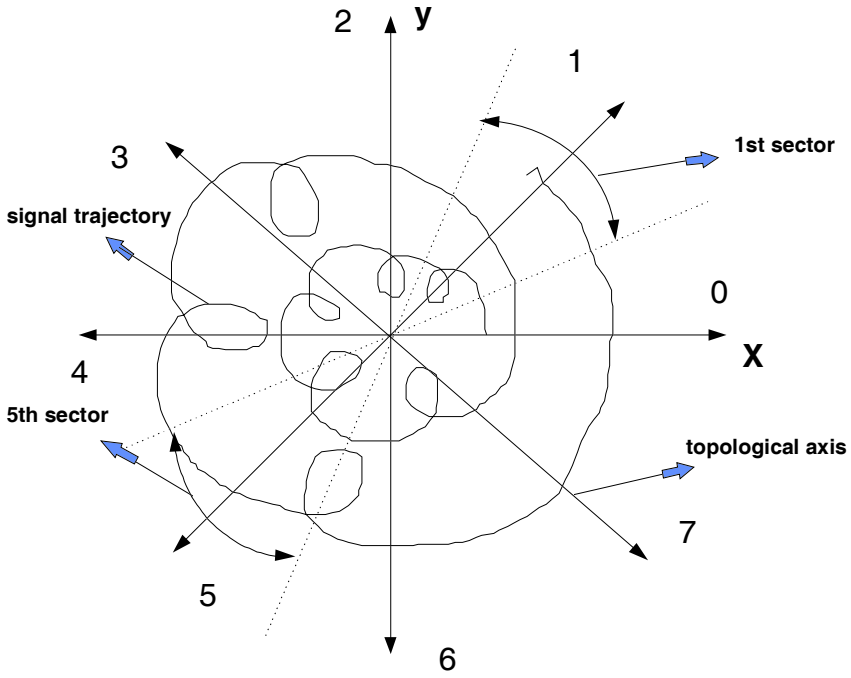
Ako uvidíme neskôr, v praxi to vedie k lineárnej aktivačnej funkcii $\sigma(x) = x$.

Na konci tejto časti zosumaruzujeme hlavné črty nášho modelu, Obr. 30. Model sa skladá z:

- 3 frekvenčným filtrom podobným jednotkám
- 1 spajk generátor s n výstupmi
- 1 časový generátor s n výstupmi
- n neurónu podobných jednotiek plne prepojenej siete, v ktorej každý neurón má:
  - 1 priamy vonkajší vstup z konkrétneho spajk a časového generátora
  - (n-1) krížových vstupov

Náš model analyzuje rečový signál v časovej oblasti (v prvom pláne). Prevádza informačné spracovanie a nie energetické spracovanie. Spracovanie je diferenciálne v čase ale výsledok spracovania daný výstupom aktivity neurónov je integrálny v čase. Cieľom návrhu a spracovania je

aby parametre modelu boli samonastaviteľné, čo je dôležité z hľadiska problematiky zovšeobecnenia. Optimálny počet neurónu podobných jednotiek prediskutujeme v kapitole venovanej výsledkom. Taktiež sme sa tu nezmienili o probléme VAD - voice activation detection, určenia začiatku-konca slova (vety). Ponecháme detajlnejšie prediskutovanie tohto problému do dodatku A1.



Obr. 32 Trajektórie a topologické invarianty

Popis nášho modelu je zložitý ale deterministický. Celá sieť sa dá vyjadriť exaktne jednou rovnicou v analytickej forme, čo je konzistentné s obecným deterministickým trendom v teóriach neurónových sietí, *Yoshifusa* (1991)*, O´Shaughnessy* (1986), *Lippmann* (1987), *Kolmogorov* (1957). Pravdepodobnostné myšlienky sú v teóriach ale aj v našom modeli prijaté len v rámci klasickej teórie pravdepodobnosti alebo klasickej fyziky.

4.2.1 Časová a frekvenčná analýza reči z hľadiska nášho modelu

V tejto časti sa dotkneme detajlnejšie niektorých základných informácií, ktoré sa súvisia s návrhom nášho systému. Abstraktný model rozoberaný v časti 4.2 špecifikujeme v digitálnej oblasti pomocou výberu konkrétnych algoritmov a ich zdôvodnenia. Najskôr sa zoznámime s vybranými časovými charakteristikami reči, v druhej časti s frekvenčnými rozšíreniami nášho modelu. V dodatku A4 je uvedených niekoľko poznámok k frekvenčnej analýze pomocou LPC analýzy.

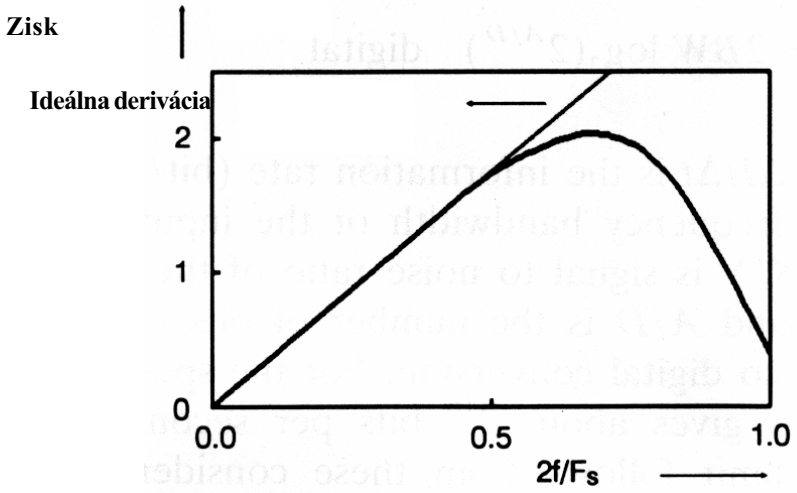
4.2.1.1 Časové charakteristiky modelu reči

V tejto časti opíšeme detajlne digitálnu aproximáciu derivácie, ktorá vstupuje do vzťahov (57), (58) v našom modeli. V numerických simuláciach sme aproximovali deriváciu pomocou Stirlingovej formuly 6-teho rádu

$$x(n) = \{[s_1(n) - s_1(n - 6)] - 9[s_1(n - 1) - s_1(n - 5)] + 45[s_1(n - 2) - s_1(n - 4)]\} / 60 \qquad (72)$$

kde $s_1(i)$ je digitalizovaný rečový signál a x(i) je Stirlingova aproximácia derivácie s frekvenčnou charakteristikou uvedenou na Obr.33 . Vidíme, že je podobná frekvenčnej charakteristike pre bazilárnu

membránu, Obr.19 (s okrem nie tak dôležitým, pre naše účely, nelineárnym efektom na reálnej charakteristike).



Obr. 33 Stirlingova transformácia - jej frekvenčná charakteristika

4.2.1.2 Frekvenčné charakteristiky modelu reči

Ako sme sa krátko zmienili v časti 4.2 navrhnutý model NP4 korešponduje viac mechanizmu percepcie ako ju poznáme u papagájov. Aby sme sa pokúsili navrhnúť model, ktorý je viac konzistentný s ľudskou fyziológiou, je užitočné sa inšpirovať už spomínanými rozdielmi medzi ľudskou percepciou a percepciou papagája.

Na základe faktov uvedených v kapitole 3 budeme diskutovať extrapolácie nášho (papagájovi podobnému) modelu NP4 na model viac podobný človeku, čo vedie v prvom priblížení k explicitnej frekvenčnej selektivite. Z nášho obecného pohľadu, prístupu k percepcii, vyplývajú dva možné základné spôsoby ako previesť túto extrapoláciu. Prvý sa zaoberá práve spomínanou frekvenčnou selektivitou ľudského ucha. Druhý je principiálne odlišný od prvého a zaoberá sa takzvanými vnútornými symetriami fonémického systému. Týmto prístupom sa budeme zaoberať v časti 7., kde sa budeme zaoberať vnútornými symetriami podrobnejšie.

Frekvenčnú selekciu budeme chápať ako explicitnú a v časovej oblasti postulovaním napríklad 4 frekvenčných priepustí, povedzme: (50 - 500) Hz, (500 - 1500) Hz, (1500 - 3000) Hz, (3000 - 5000) Hz. Alebo môžeme zvoliť škálu filtrov, ktorá modeluje bark (alebo mel) škálu, to znamená rovnako vzdialené fixné filtre do frekvencie 1 KHz a potom logaritmický nárast šírky filtra. Príklad takejto škály je v Tab.4.

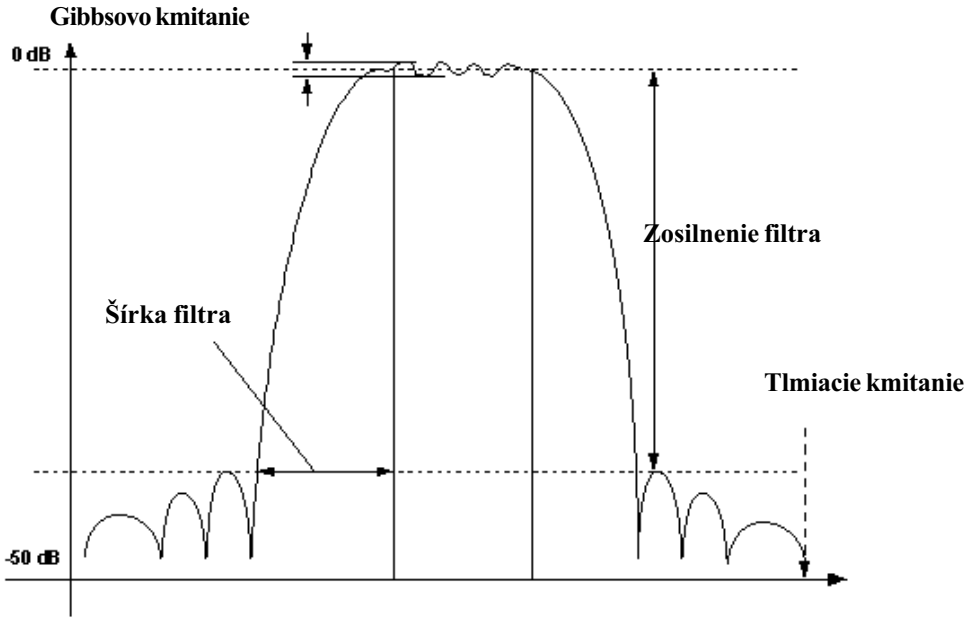
bark	1	2	3	4	5	6	7	8	9	10
kHz	0,102	0,204	0,309	0,417	0,531	0,652	0,781	0,923	1,079	1,256
bark	11	12	13	14	15	16	17	18	19	20
kHz	1,457	1,692	1,972	2,310	2,727	3,248	3,904	4,729	5,758	7,031

Tab. 4 Barkova škála, ktorá približne korešponduje percepčnému frekvenčnému rozlíšeniu ľudského ucha

Väčší počet filtrov nie je nutný. Ak napríklad použijeme 50 alebo 100 frekvenčných filtrov potom výstup z týchto filtrov, v časovej oblasti, sú približne harmonické frekvencie s frekvenciami týchto filtrov a takto prakticky zničíme, znehodnotíme všetku užitočnú informáciu.

4.2.1.3 Kaiserov nerekurzívny filter

V roku 1964 Kaiser zostrojil triedu optimálnych okien, ktorá sa dajú použiť na zostrojenie digitálnych filtrov, ktoré spĺňajú alebo dokonca prekračujú požiadavky na návrh ľubovolného filtra *Antoniou* (1983). V prvom rade sú filtre takto navrhnuté nerekurzívne, to znamená že sú kauzálne a lineárne v časovej oblasti, inými slovami ich fázová charakteristika je lineárna. To sú hlavné požiadavky na filter, ktoré boli špecifikované v časti 4.2. pre transformáciu vstupného signálu. Kaiserov filter alebo okno závisí iba od dvoch parametrov, dĺžky okna N a parametra β, ktorý kontroluje tvar okna. Jediné obmedzenie, ktoré vyplýva z návrhu filtra pomocou Kaiserovej metódy je, že kmitanie v priepustnej (tzv. Gibbsove kmitanie) a tlmiacej časti spektra filtra musia byť rovnaké, čo v praxi nie je žiaden problém, pozri Obr. 33.



Obr. 33 Kaiserov filter - frekvenčná charakteristika

V ďalšom sa sústredime iba na jeden priepusťový filter, inými slovami na filtráciu vstupného signálu v jednom frekvenčnom pásme. Impulzná charakteristika takéhoto filtru musí byť lineárna, stacionárna a jej fázová charakteristika tiež musí byť lineárnou. Systémovou realizáciou takéhoto filtra je takzvaný nerekurzívny digitálny filter, v našom prípade Kaiserov filter. Po navrhnutí filtra so zodpovedajúcou impulznou charakteristikou sa prevedie konvolúcia tejto charakteristiky so vstupným signálom, podľa vzťahu

$$y[j] = \sum_{i=0}^{N-1} x[i] * h[j - i]$$

kde x[i] ; {0 <= i <= N-1} je signál (zdigitalizovaný), a h[k] { 0 <= k <= M-1 }je impulzná odozva, alebo charakteristika, potom výstupný prefiltrovaný signál je y[j] .

Druhú alternatívu, implicitne frekvenčne selektívny model NP4, môžeme definovať pomocou dozadnej Fourier transformácie signálu, ktorý už bol prefiltrovaný dolno priepusťovým filtrom $s_1(t)$, (pozri časť 4.2) ako

$$F(\omega, t) = \int_{-\infty}^{t} s_1(t)\, \lambda\, (t - t')\, e^{-i\omega t}\, dt' = ReF(\omega, t) + ImF(\omega, t) \qquad (73)$$

kde funkciu okna $\lambda(t)$ sme zaviedli kvôli požiadavke konvergencie integrálu (73). Treba poznamenať, že (73) nie je krátko-časovou Fourier transformáciou (kde funkcia okna má konečnú dĺžku). Definujeme určitú hladiacu procedúru pre $F(\omega, t)$ ako

$$G(\omega, t) = \int F(\omega', t)\, H(\omega, \omega')\, d\omega' \qquad (74)$$

kde reálna funkcia H(x,y) je impulzná charakteristika vo frekvenčnej oblasti hladiaceho filtra. Tento filter používame na odstránenie alebo zníženie veľmi rýchlo oscilujúcich zložiek v $F(\omega, t)$. V diskrétnom obraze, v časovej (a frekvenčnej) oblasti $F(\omega, t)$ má približne 10 000 harmonických pre vzorkovaciu frekvenciu okolo 10 kHz a časové trvanie okolo 1 s. Aby sme kompresovali redundantnú informáciu v $F(\omega, t)$, môžeme použiť viacero spôsobov. Jeden z nich spočíva vo využití AD/C - analógovo-digitálného prevodníka a jeho parametra - vzorkovacej frekvencie tak, aby sa znížil počet frekvenčných zložiek z približne 10 000 na približne 500. Ak všetky zložky takéhoto systému sú fyzikálne kauzálne, potom sa dá dokázať *Nussenzweig* (1972), že $ReG(\omega, t)$ a $ImG(\omega, t)$, reálna a imaginárna časť vyhladeného, skompresovaného a zdigitalizovaného (4.4.1) sú vzájomne Hilbertove transformanty, čo je postačujúce pre naše požiadavky invariancie voči intenzite a šumu, ktoré sme spomínali v časti (4.2). V rámci tohto prístupu môžeme veľmi jednoducho vyhovieť požiadavke uzavretosti trajektórie, pozri časť (4.3), príčina je jednoducho v reálnom charaktere funkcie $s_1(t)$. Tak zvaný Nyquistov graf, *Oppenheim, Shafer* (1975), vo frekvenčnej oblasti je symetrický voči reálnej osi a trajektória tohto grafu, ktorá je daná reálnou a imaginárnou časťou $G(\omega, t)$, je uzavretá trajektória.

Je užitočné si tiež všimnúť, že tento prístup je v určitom spojení s fázovým prístupom, spomínaným v kapitole 2, fázovými zložkami

$$I(k,t) = (Re^2[G(k, t)] + Im^2[G(k, t)])^{1/2}$$
$$\varphi(k,t) = \text{arctg}(Im[G(k, t)] / Re[G(k, t)]) \qquad (75)$$

kde I(k,t) je okamžitá intenzita a $\varphi(k,t)$ je okamžitá fáza danej harmonickej zložky. Vidíme, že takýto prístup vedie veľmi prirodzene k otázke o fázovej citlivosti ucha. Podrobnejšie sa týmto v tejto prácii nebudeme zaoberať.

Druhým možným spôsobom vyhladenia frekvenčných zložiek je využitie LPC metód. Budeme sa venovať tomuto spôsobu trochu podrobnejšie v Dodatku A4. Samozrejme aj v prehľadovej časti spomínané parametrizácie PLP a MFC sa dajú chápať ako spôsoby vyhladenia spektra, u PLP až do vzťahu (11), u MFC do vzťahu (13) a namiesto rátania príslušných koeficientov sa môže vyhladené spektrum - jeho reálna a imaginárna časť brať za vstup pre výpočet topologických invariantov, nie v časovej ale frekvenčnej oblasti, pozri dodatok A4.

4.3 Pravdepodobnostná neurónová sieť

V predošlých častiach (4.2) sme ukázali, že náš model vytvára n reálnočíselných hodnôt, ktoré reprezentujú výstupné aktivity skupiny n neurónov. Znamená to, že pôvodný rečový signál (skalárna veličina) je nakoniec pretransformovaný do n-zložkového reálnočíselného vektora $h(t) \in R^n$. Konečná reprezentácia jedného slova je potom daná vývinom neurónovej aktivity skupiny neurónov na konci slova pomocou konečného vektora $x \in R^n$.

Na tomto stupni diskusie nám zostáva vyriešiť dva „chronické" problémy. Prvý z nich je navrhnúť učiaci mechanizmus na vytvorenie prototypov slov, či z hľadiska jazykovej informácie alebo informácie o hovoriacom. Druhý problém sa dotýka návrhu adekvátnej klasifikačnej schémy pre vektorové vzory do jednej triedy prototypu (z celej množiny naučených prototypov).

Z neurofyziológie vieme, že oba mechanizmy fungujú u človeka súčasne, hoci hlavný učiaci proces sa uskutočňuje v prvých rokoch života. Takže, dospelí ľudia už využívajú efektívne vytvorenú a spracovanú množinu prototypov slov, resp. hovoriacich, ktorá sa počas ďalších rokov len jemne doľaďuje, *Hertz, Krogh, Palmer* (1991). V tejto práci sme postavený pred podobnú situáciu, pri ktorej modelujeme oba mechanizmy v rámci jedinej neurónovej štruktúry Pravdepodobnostnej neurónovej siete (PNN). Namiesto zaoberania sa priamo s prototypmi patriacimi konkrétnemu slovu, alebo hovoriacemu, funguje táto schéma implicitne s kompletnými množinami pravdepodobnostných rozdelení prototypov - ktoré sú dané realizáciami vzorov patriacich do konkrétnej triedy slova alebo hovoriaceho.

Predpokladajme, že máme vektory vzorov $[x^{jk}]$, kde k označuje triedu - slova alebo hovoriaceho a j označuje konkrétnu j-tu zložku výstupného vektora z nášho modelu prediskutovaného v predošlých častiach.

V rámci takzvanej Bayesovskej filozofie alebo Bayesovského prístupu, *Gnedenko* (1976), môžeme písať pre mieru úspešnosti (anglicky cost) klasifikácie vzoru do triedy k (predpokladajúc štatistické rozdelenie do tried 1,2, ..., M)

$$C(x,k) = \sum_{l=1}^{M} P(l \mid x)\, L(k,l) \qquad (76)$$

kde $P(l \mid x)$ je podmienená pravdepodobnosť toho, že vzor x patrí do triedy l, $L(k,l)$ je jednotková úspešnosť jedného rozhodnutia, vyjadrujúca neúspešnú klasifikáciu vzoru x do triedy k, ak bol v skutočnosti z triedy l a sumuje sa cez l od 1 do M. Pre rovnako dôležité, signifikantné vzory môžeme napísať pre $L(k,l)$

$$L(k,l) = 1 - \delta_{kl}.$$

Podmienená pravdepodobnosť $P(l \mid x)$ sa dá vyjadriť ako

$$P(l \mid x) = p(x \mid l)\, P(l)$$

kde $P(l)$ je pravdepodobnosť, že ľubovolný pozorovaný vzor skutočne patrí do triedy l. Potom $p(x|l)$ je rozdelenie pravdepodobnosti vektorov-vzorov z triedy l. Podľa *Specht* (1990) predpokladáme, že toto rozdelenie pravdepodobnosti sa dá vyjadriť multivariačným gausovským rozdelením ako

$$p(x \mid l) = A(n,l) \sum_{j=1}^{N_l} \exp[- (x - x^{lj})^2 / 2 s^2] \qquad (77)$$

kde

$$A(n,l) = 1 / [(2 p)^{n/2} s^n N_l]$$

kde n je rozmer paternov, s je „disperzia" pravdepodobnostného rozdelenia, má význam parametru „vyhladenia" , l je trieda vzorov N_l je počet vzorov danej triede l. Znamená to, že odhadujeme rozdelenie pravdepodobnosti alebo hustotu pravdepodobnosti z takzvaných trénovacích, učiacich vzorov x^{lj} . Pomocou tohto rozdelenia môžeme definovať diskriminačnú funkciu $D_k(x)$ ako

$$D_k(x) = - \sum_{l=1}^{M} R(l)\, L(k,l)\, (1/N_l) \sum_{j=1}^{N_l} \exp[- (x - x^{lj})^2 / 2 s^2] \qquad (78)$$

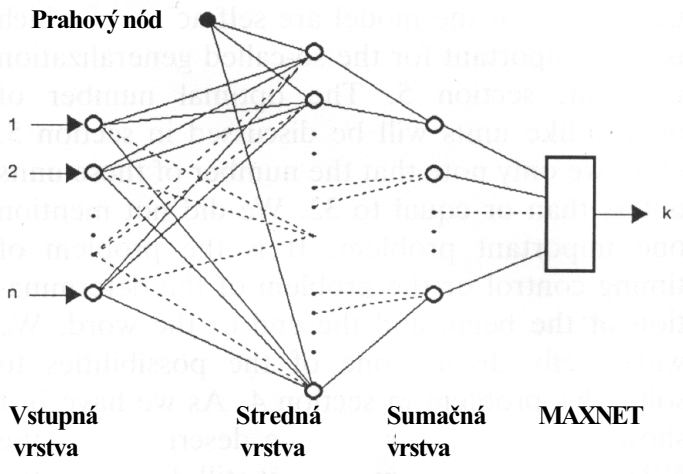
kde v prvej sume sa sumuje cez l od 1 po M, a v druhej sume cez j od 1 po N_l. Pretože pri klasifikácii sú dôležité iba relatívne úspešnosti možeme zanedbať v (77) všetky faktory, ktoré sú spoločné pre všetky triedy. Potom môžeme predefinovať (78) ako

$$D_k(x) = - \sum_{l=1}^{M} R(l)\, L(k,l)\, (1/N_l) \sum_{j=1}^{N_l} \exp[- (x\, x^{lj} - x^{lj}\, x^{lj}) / 2 s^2]. \qquad (79)$$

Základný tvar tejto diskriminačnej funkcie je, podľa podmienky (79), potom daný ako

$$D_k(x) = P(k)\, (1/N_l) \sum_{j=1}^{N_l} \exp[- (x\, x^{kj} - x^{kj}\, x^{kj}) / 2 s^2]. \qquad (80)$$

Teraz môžeme navrhnúť viacvrstvovú neurónovú sieť, ktorá modeluje diskriminačnú funkciu (79), podľa, *Specht* (1990) a je na Obr.35.



Obr. 35 Pravdepodobnostná neurónová sieť

V tejto neurónovej sieti, nazývanej PNN, tok hodnôt je zo vstupnej vrstvy do strednej vrstvy. Každý nód v strednej vrstve odpovedá jednému učiacemu vzoru, x^{kj} a váhy zo vstupnej vrstvy k tomuto nódu sú rovné zložkám učiaceho vektora vzorov. Váha z prahového okolia, alebo neurónu je úmerná štvorcu euklidovskej vzdialenosti konkrétneho učiaceho vzoru. Neuróny v strednej vrstve majú nelinearitu typu

$$\exp[(.) / s^2]$$

Váhy zo strednej vrstvy k sumujúcej vrstve sú rovné jednotke a každý nód v sumačnej vrstve je priradený k jednej z tried, takým spôsobom, že každý nód zo strednej vrstvy je spojený iba s nódom v sumačnej vrstve zo svojej triedy. Váhy zo sumačnej vrstvy do prvej vrstvy MAXNET sú rovné $P(k) / N_k$ a MAXNET generuje index toho sumačného nódu, ktorý má maximálnu hodnotu, alebo inými slovami rozpoznanej triedy - slova alebo hovoriaceho.

Pri bližšom pohľade vidíme, že učenie v našej sieti je jednoduché (bez akýchkoľvek iterácií), okamžité a dá sa veľmi ľahko prispôsobiť k časovo sa meniacej štatistike vzorov, *Lenz* (1991), *Forsyth* (1995). Naviac, ako sa zmieňuje aj *Girosi, Poggio* (1990), *Musavi, Kalantri, Ahmed, Chan* (1990), *Specht* (1990) PNN môže asymptoticky modelovať akúkoľvek funkciu pravdepodobnostného rozdelenia, za veľmi obecných podmienok. Dá sa to previesť jednoducho zmenou hladiaceho parametra s. Druhou výhodou tejto siete je neporovnatelná relatívna rýchlosť oproti napr. viacvrstvovému perceptrónu podobným sieťam s učením pomocou spätného šírenia chyby. V praxi je tento rozdiel až 200 000 násobný, *Yau, Manry* (1990).

Na konci tejto časti prediskutujeme v krátkosti možnosť implementácie určitých syntaktických čŕt do nášho systému. Dá sa to previesť dvomi spôsobmi. V prvom, nazvime ho explicitná implementácia, definujeme reinterpretáciu (80). Postulujeme existenciu matice syntaxe, alebo lexiky alebo obecne vzťahov S(k,l) systému slov, alebo hovoriacich - obecne tried. Potom interpretujeme prvý člen v (80) ako podmienenú pravdepodobnosť vyskytnutia sa vzoru z triedy k ak predošlý vzor bol zaradený do triedy l. Postulujeme, že matica vzťahov je úmerná tejto podmienenej pravdepodobnosti. Potom môžeme predefinovať diskriminačnú funkciu ako

$$D'_k(x) = S(k,l)\, D_k(x) \qquad (81)$$

kde $D_k(x)$ je definované v (80). Aby sme toto mohli realizovať musíme poznať umelé alebo prirodzené vzťahy (syntax) systému, či jazykového alebo personálneho.

Kvôli úplnosti spomenieme aj druhý spôsob akým môžeme implementovať určité vzťahové funkcie do nášho systému, je to využitie samo-organizujúcej sa siete typu *Kohonena* (1989), namiesto PNN. Tento spôsob implementácie nazveme implicitný. Urobili sme niekoľko experimentov s takouto sieťou a považujeme tento prístup za užitočný.

5 Výsledky - rozpoznávanie slov

V prvej časti tejto kapitoly sa budeme zaoberať motiváciou, popisom experimentálnych podmienok a prezentujeme niektoré výsledky rozpoznávania izolovaných slov nezávisle od hovoriaceho.

Hlavným cieľom uvedených počítačových simulácií bolo vyšetriť kľúčové vlastnosti navrhnutého modelu rozpoznávania reči ako je uvedené v časti 4.2, pozri aj Obr. 28, špeciálne testovanie invariantných vlastností percepcie a rozpoznávania. Toto testovanie považujeme za základné a rozhodujúce pri návrhu optimálneho a biologicky plauzibilného systému. Kvôli tomuto a kvôli časovému a materiálnemu ohraničeniu sme uprednostnili štúdium veľmi malého súboru slov, avšak nahraných pre relatívne veľa rôznych realizácií. Aby sme otestovali vlastnosti hraničných podmienok, vybrali sme fonologicky veľmi blízke slová.

Vybrali sme 5 slov, ktoré sa líšia iba v jednom dištinktívnom príznaku : lama, lame, lami, lamo. Použili sme 8 hovoriacich, 5 mužov a 3 ženy. Nahrali sme 500 realizácií týchto slov. Súbor týchto slov môžeme charakterizovať nasledujúcimi parametrami. Maximum absolútnej intenzity realizácií sa pohybovala v rozsahu od 3 V do 8 Voltov. Časové trvanie realizácií v našom slovníku bol z rozsahu 0,2 až 0,7 s. Maximum základného tónu sa pohybovala v rozsahu od približne 100 Hz do 350 Hz. Začiatok a koniec slov sme určovali naším VAD algoritmom, pozri dodatok A1.

Akustické predspracovanie v našom systéme sa skladalo z nasledujúcich častí: komerčný mikrofón, bez nejakých zvláštnych vlastností, predzosilňovač, zosilňovač, 5 KHz dolno priepusťový filter 3 rádu, so strmosťou frekvenčnej charakteristiky rovnou 24 dB na oktávu, 12 - bit AD/C so vzorkovacou frekvenciou 10 KHz.

V prvom type experimentov sme testovali vlastnosti zhlukovania nášho modelu. Na tento účel sme použili 100 realizácií (20 realizácií pre každé slovo) z našej databázy. Ako mieru podobnosti, vzdialenosti sme použili euklidovskú vzdialenosť

$$d_k(x) = (x - x^k)^{1/2}$$

kde x je výstupný vektor z nášho modelu, ktorý chceme testovať a x^k je vektorová aritmetická stredná hodnota učiacich vektorov pre k -tu triedu, slovo. Učiaca množina je tá istá ako testovacia.

V tabuľke 5. uvádzame výsledky pre rôzne funkcionálne formy spracovania, *Chudý, Chudý, Hapák* (1991). V ľavom stĺpci sú uvedené konkrétne typy spracovania, t.j. dynamické pravidlá neurónovej siete (sluchového neurónového systému) podľa 4.2. Používame nasledujúce nastavenie parametrov: počet neurónov je rovný 32, adaptačný čas (alebo dĺžka okna, pozri kapitola 4.2) je rovná 12,8 ms a retardačný čas je rovný 51,2 ms.

typ spracovania	nenormalizované	normalizované
$\eta_i(t + \Delta t) = \eta_i(t) + \Delta\mu_i(t)$	30%	20%
$\eta_i(t + \Delta t) = \eta_i(t) + \Delta\xi_i(t)$	40%	36%
$\eta_i(t + \Delta t) = \alpha\Sigma\eta_j(t)\Delta s_{ij}(t) + \Delta\xi_i(t)$	50%	44%
$\eta_i(t + \Delta t) = \eta_i(t) + \Delta\xi_i(t)\Delta\mu_i(t)$	60%	63%
$\eta_i(t + \Delta t) = \alpha\Sigma[\eta_j(t)\Delta s_{ij}(t) + \Delta\xi_i(t)\Delta\mu_j(t)]$	70%	65%

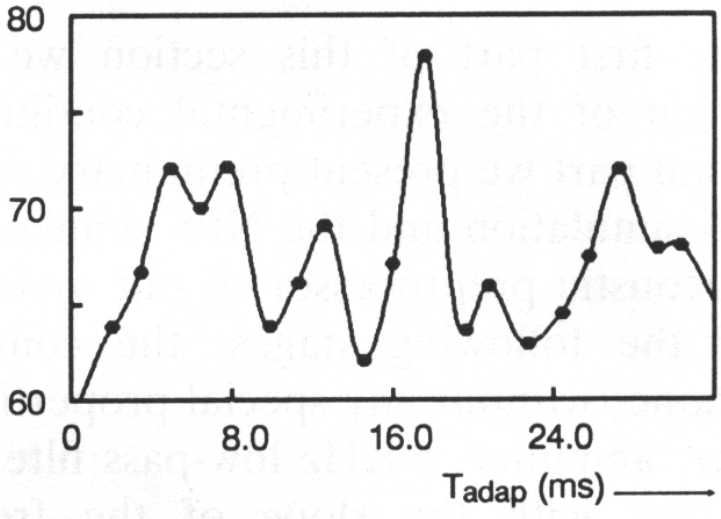
Tab. 5 Vlastnosti zhlukovania modelu rozpoznávania reči

V strednom stĺpci sú skóre úspešnosti pre nenormalizované vektory a v poslednom stĺpci pre normalizované vektory (všetky sme delili celkovým časom trvania danej realizácie). Účelom tohto experimentovania je vyšetriť experimentálne vplyv rozličných faktorov na dynamiku neurónovej siete.

V prvom, druhom, a štvrtom riadku sú uvedené funkcionálne tvary, ktoré nemajú sieťové kooperatívne chovanie (t.j. dynamika konkrétneho neurónu nie je ovplyvnená inými neurónmi). Z týchto výsledkov vidíme, aj posledne spomínané spracovanie - bez sieťovej kooperácie - má určitú klasifikačnú, zhlukovú schopnosť. Porovnaním úspešnosti vidíme, že sieťové spracovanie nielen „normalizuje" vektory podľa doby trvania slova, ale aj prevádza aj určitý typ „homotopického" zobrazenia, ktoré je lokálne v čase, dodatok A2. Je to vidieť z porovnania úspešnosti pre normalizované a nenormalizované vektory v treťom a poslednom riadku. Pozoruhodný vplyv má tiež váhový faktor $\Delta\mu_i(t)$ v externom vstupe ΔI_i , ktorý je úmerný generátoru časového trvania (pozri tretí a štvrtý riadok).

Zo skúsenosti môžeme takisto usúdiť, že úspešnosť klasifikácie sa stáva nasýtenou so zvyšujúcim sa počtom neurónov. Úplne nasýtenou je pre počet neurónov 32. V prípade väčšieho počtu neurónov, predpokladáme, že neuróny sa stávajú korelované, pretože úspešnosť s narastaním počtu neurónov už nerastie. Optimálna hodnota váhovej konštanty „a" je rovná 1/32, čo má veľmi prirodzenú interpretáciu ako normalizácia počtom neurónov.

Na Obr. 36 prezentujeme závislosť nášho modelu od adaptačného času. Vidíme, že náš model „funguje" aj pre časy okolo 1,6 ms. Maximálna úspešnosť sa dosahuje pre adaptačné časy rovné 17,6 ms, čo je v súhlase s experimentálnymi faktami z biologických experimentov, *Reddy, D.R.*, (1966), (1967).



Obr. 36 Závislosť NP4 modelu od adaptačného času

Predpokladáme, že adaptačný čas je optimálny iba v štatistickom zmysle, čo znamená, že počas spracovania dĺžka okna nie je konštantná ale, napríklad, gausovsky rozdelená okolo tejto optimálnej hodnoty (17,6 ms). Zo simulačných výsledkov môžeme urobiť záver, že náš model nezávisí na retardačnom čase. Naša skúsenosť s modelom, hovorí že pre úlohy zovšeobecnenia (prechod od malých súborov slov k väčším) a podobne, je optimálnou parametrizáciou spracovania, výber spracovania s čo najmenším počtom fitovacích parametrov (počet neurónov, refrakčný a retardačný čas nepovažujeme v našom modeli za fitovacie parametre), najlepšie samo sa prispôsobovacie, samo adjustovatelné spracovanie - bez akýchkoľvek parametrov, konštánt atď. a je to vlastne aj jedným z cieľov našej práce, nášho prístupu, *Chudý, Chudý, Hapák* (1991).

5.1 Modifikácie dynamiky neurónovej siete

Navrhnutý model ponúka veľa možností na modifikáciu, ktorá môže vylepšiť schopnosť klasifikácie a klasterizácie.

I. Najskôr by sme chceli prediskutovať vplyv predpokladu netriviálneho tvaru „synapsií“ Δs_{ij}. Doteraz sme vyšetrovali dva tvary. Prvý má triviálny tvar

$$\Delta s_{ij}(t) = a\delta_{ij} \qquad (82)$$

kde a je konštanta a d_{ij} je kroneckerov symbol. Tento tvar korešponduje situácii, kde každý neurón interaguje iba sám so sebou, neuróny sú nezávislé. Výsledky simulácie pre tento prípad sú dané v tabuľke 5 (prvý, druhý a štvrtý stĺpec). Druhý tvar (základný tvar), ktorý je prezentovaný v (4.2.9) korešponduje súboru vzájomne interagujúcich neurónov. Opäť výsledky simulácie sú dané v tabuľke 5 (tretí a piaty riadok). Teraz zavedieme takzvané vyhladené tvary synapsií definovaním sigmoidnej transformácie

$$\Delta s_{ij} \rightarrow \tanh [\Delta s_{ij} / T] \qquad (83)$$

kde T je hladiaci parameter, „teplota“. Závislosť takto modifikovaného modelu, jeho úspešnosti na teplote T je v Tab. 6.

T	10^{-6}	10^{-1}	0,25	0,5	1,0	2,0	4,0	10,0	10^{2}	10^{3}
úspešnosť %	67	68	69	70	70	71	71	68	68	68

Tab. 6 Závislosť modifikovaného modelu na hladiacom parametre

Vidíme, že hladiaci parameter T kontroluje „diskusiu“ t.j. ostrosť interakcie medzi neurónmi. Pre nízke teploty máme nižšie úspesnosti (ostrá, čierno-biela diskusia) a pre vysoké teploty máme tiež nižšie úspešnosti (mäkká, farebná diskusia). Medzi týmito hraničnými hodnotami existuje určitá optimálna teplota T, kde je úspešnosť maximálna. V našom prípade optimálna hodnota T je okolo 3. Pri tejto simulácii sme použili 32 neurónov, refrakčný čas bol 12,8 ms a retardačný čas okolo 51,2 ms

II. Teraz vyšetríme vplyv určitého typu frustrácie matice synapsií. Predpokladajme nasledujúcu modifikáciu synapsií (polovica synapsií je vynulovaná)

$$Ds_{ij} \rightarrow \begin{cases} Ds_{ij} & \text{pre } j < n/4 \text{ alebo } j \geq 3n/4 \\ 0 & \text{pre všetky ostatné synapsie} \end{cases} \qquad (84)$$

kde n je počet neurónov a Ds_{ij} je jedna z predpokladaných tvarov synapsií (základné alebo vyhladené). Myšlienkovým základom pre takúto konkrétnu formu synapsií, sú fyziologické merania (pozri *Spitzer, Hochstein* (1985)), ktoré nám ukazujú, že akustická aktivita je potlačená počas zápornej časti stimulačného signálu, keď je pod spontánnou rýchlosťou pálenia. Záporná časť stimulačného signálu odpovedá v našom modeli ľavej časti fázovej roviny, pozri Obr. 30. Simulačné výsledky ukazujú, že

tento typ frustrácie neovplyvňuje výslednú úspešnosť, ktorá je rovnaká ako v predošlom prípade (s rovnakým súborom parametrov) t.j. okolo 71 %. Takto z pragmatického hľadiska potrebujeme iba $n^2/2$ prepojení namiesto n^2.

III. Druhou prirodzenou modifikáciou siete je transformácia, ktorá sa dotýka funkcionálneho tvaru výstupu neurónov. V časti 4.2 sme zaviedli základný tvar týchto výstupov neurónov ako

$$s(x) = \begin{cases} x & \text{pre } x \geq 0 \\ 0 & \text{pre } x < 0 \end{cases} \qquad (85)$$

Tento funkcionálny tvar sme použili tiež v predošlých simulačných experimentoch. Teraz definujeme sigmoidný funkcionálny tvar

$$s(x) = s_0 / (1 + e^{-x/t})$$

kde t je hladiaci parameter a s_0 je hodnota maximálnej výstupnej aktivity (v našom prípade rovná 256). V tabuľke Tab.7 uvádzame skoře úspešnosti predošlých frustračných modelov, s refrakčným časom 12,8 ms a retardačným časom 51,2 ms. Vidíme, že podobne ako pre simuláciu s vyhladením synapsií, existuje optimálna hodnota hladiaceho parametra t, pre ktorý existuje maximálna úspešnosť. V našom prípade to znamená zlepšenie približne 5 % v porovnaní s najmešou úspešnosťou.

τ	10^{-6}	10^{-1}	0,25	0,5	1,0	2,0	4,0	10,0	10^2	10^3
úspešnosť %	67	68	69	70	70	71	71	68	68	68

Tab. 7 Závislosť modifikovaného modelu na hladiacom parametre t

Podobne ako v prípade synapsií aj tu môžeme uvažovať frustrovaný typ výstupnej aktivity

$$s(x) = \begin{cases} s_0 / (1 + e^{-x/t}) & \text{pre } j < n/4 \text{ alebo } j \text{ ł } 3n/4 \\ 0 & \text{pre všetky ostatné synapsie} \end{cases} \qquad (86)$$

Pre tento tvar sme dostali úspešnosť okolo 70 % pre optimálnu hodnotu hladiaceho parametra. Opäť, je si vhodné všimnúť, že v tomto prípade používame iba n/2 neurónov t.j. $n^2/4$ synapsií.

5.2 Vplyv šumu na náš model

Nakoniec prezentujeme výsledky simulácií, ktoré reflektujú invariantnosť nášho modelu voči šumu. Pri týchto simuláciach sme použili základný tvar synapsií a výstupných aktivít (s 32 nefrustrovanými neurónmi), refrakčný čas 12,8 ms, a retardačný čas 51,2 ms. Vplyv šumu sme simulovali nasledujúcim spôsobom

$$s(i) = z(i) + a\,(1 - \text{rand})\,z(i) + b(1 - \text{rand})\,2048 \qquad (87)$$

kde z(i) je pôvodný rečový signál v digitálnom tvare, s(i) je zašumený signál, rand je výstup z generátora pseudonáhodných čísiel - homogénne z intervalu <0,1> a je úroveň takzvaného „korelovaného“ šumu a b je úroveň „nekorelovaného“ šumu, 2048 je maximálna hodnota signálu v 12 bit AD/C. Hodnota parametru a alebo b rovná 0,2 odpovedá 20 % úrovni šumu.

a =0	0.0	0,2	0,4	0,6	0,8
úspešnosť %	69	68	74	70	71

b =0	0.0	0,2	0,4	0,6	0,8
úspešnosť %	70	61	51	54	56

Tab. 8 Závislosť modifikovaného modelu na šume

Vidíme, že najlepšie skóre úspešnosti je pre nulovú úroveň šumu a pre 40 % korelovaný šum, Tab. 8. Výsledky tiež demonštrujú invariantnosť voči šumu nášho modelu aj pre korelovaný šum.

5.3 Model NP5 - výsledky rozpoznávania

V tejto časti uvedieme výsledky ilustrujúce rozpoznávacie schopnosti úplného modelu NP5 (NP4 + PNN), čo je náš model podľa 4.2 spolu s neuronálnym modelom rozpoznávania PNN**,** *Chudý, Chudý, Hapák.* (1991).

Pre testovanie úspešnosti modelu NP5 sme náhodne vybrali súbor 100 slov z našej množiny (t.j. 20 realizácií pre každé slovo). Tieto slová - realizácie nie sú také isté ako pre predošlé experimenty. Zostávajúce slová z databázy použijeme pre „učenie“ neurónovej siete PNN. Výsledky sú uvedené v Tab. 9.

typ spracovania:												
$\eta_i(t+\Delta t) = \eta_i(t) + \Delta\zeta_i(t)\Delta\mu_i(t)$											ľavý stĺpec	
$\eta_i(t+\Delta t) = \alpha\Sigma\eta_j(t)\Delta s_{ij}(t) + \Delta\zeta_i(t)\Delta\mu_i(t)$											pravý stĺpec	
Triedy:	LAMA		LAMU		LAME		LAMI		LAMO		VŠETKO	
skóre % úspešnosti Trénovacia + testovacia:												
	92	96	97	98	91	93	98	99	90	95	94	96
Testovacia:	82	88	94	96	58	80	94	96	82	88	82	89

Tab. 9 Celkové výsledky rozpoznávania v modeli NP5

Vyšetrovali sme tu iba učiace a klasifikačné vlastnosti PNN, ktoré používali NP4 spracovanie iba s 2 najlepšími sieťovými dynamikami, pozri Tab. 6. Vidíme, že najzložitejšia sieťová dynamika, ktorú sme definovali detajlne v predošlej časti dáva dobré výsledky (pozri pravý stĺpec).

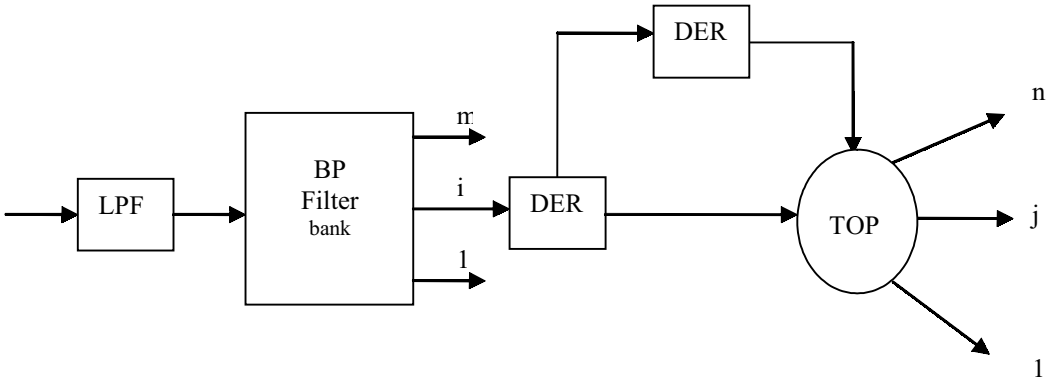
Prezentované výsledky v poslednom riadku ilustrujú zovšeobecňovacie schopnosti NP5, v tomto prípade trénovacia množina bola 400 paternov a testovacia množina 100.

Je zrejmé, že chovanie PNN modelu sa dá prispôsobiť a optimalizovať hlavne dvomi spôsobmi. Prvý spôsob je správny výber optimálneho najnižšieho počtu realizácií pre každú triedu slova a následne vhodný výber konkrétnych realizácií. Týmto smerom sa môže uberať ďaší výskum.

Hladiaci parameter siete PNN sa správa dostatočne robustne, to znamená, že dáva prijatelné výsledky pre dostatočne široký rámec hodnôt. Prezentované výsledky sú pre experimentálne vybrané optimálne hodnoty tohto parametra.

5.4 Frekvenčné a iné modifikácie modelu NP4

V tejto časti sa budeme zaoberať niektorými modifikáciami nášho modelu. Pôvodne sme navrhli a testovali NP4 a NP5 spolu s neurónovou sieťou PNN, *Chudý, Hapák, Chudý* (1991), ktorá pôsobila ako klasifikačný nástroj. Toto bolo uskutočniteľné iba pre model rozpoznávania izolovaných slov. Ak chceme použiť topologické invarianty ako črty s HMM pre rozpoznávanie nezávilé od hovoriaceho a pre súvislú reč musíme previesť niektoré zmeny, *Kačúr, Chudý* (2012). Prvou zmenou je zakomponovanie frekvenčnej závislosti percepcie, pomocou radu priepusťových filtrov, Obr. 37. Sadu filtrov sme zakomponovali za dolnopriepusťový filter a modelovali sme ich pomocou Kaiserovho návrhu filtra, 4.2.1.3.



Obr. 37 Frekvenčná modifikácia modelu NP4 - NHP3

Súčasné najúspešnejšie štatistické modely rečového signálu a percepcie používajú HMM, pozri 2.1.3, previazané kontextovo závislé (CD) fonémy spolu viacnásobnými zmiešanými Gaussovskými modelmi, *Nouza a kol.* (2005). Z hľadiska spojenia s extrakciou čŕt sa obecne ukazuje ako najoptimálnejšie spojenie HMM s MFC a PLP, pozri 2.1.1.3 a 2.1.1.2. Avšak za určitých podmienok sa využívajú aj iné statické črty napr. TIFFING, *Nadeu, Macho* (2001), *Haque, Togneri, Zaknich* (2009), ZCPA. Väčšina z týchto statických čŕt sa snaží odhadnúť amplidútovo modifikované a frekvenčne deformované spektrá, ktoré odpovedajú nasledujúcim zisteniam, *Rabiner, Juan* (1993) - v percepcii reči zohrávajú dôležitú úlohu aj vlastnosti odvodené z produkcie reči a síce rôzne polohy formantov a ich šírky, zatiaľčo celkový náklon spektier, chýbajúce frekvencie medzi prvým a tretím formantom nie sú dôležité.

5.4.1 Redukcia dimenzií pomocou LDA

Po segmentácii učiacich dát do tried foném sme pozorovali, že niektoré dimenzie v extrahovaných črtách sú korelované. Toto nie je žiadúce pre HMM modelovanie pomocou Gaussovských rozdelení, pretože matica kovariancie môže byť takto iregulárna.

Takto je nutné použiť techniky na redukciu dimenzií, pri zachovaní dôležitých z nich, ako sú PCA, ICA, LDA a HLDA. Pretože našim cieľom je nielen redukovať dimenzie ale aj zároveň zachovať separovanie medzi rôznymi triedami foném, otestovali sme metódy LDA a HLDA, kde sa dá táto informácia ľahko zohľadniť. Ako prvú sme vyskúšali heteroscedastickú LDA, *Kumar* (1997), pretože podporuje rôzne kovariančné matice tried. Použili sme iteratívny algoritmus podľa *Liu, Gales, Woodland* (2003). Prakticky sa táto metóda ukázala ako nepoužiteľná kvôli divergencii výpočtov pre inverznú ustrednenú kovariančnú maticu.

Po tomto praktickom uvedomení si situácie sme sa rozhodli pre menej náročnú procedúru LDA. Dá sa v krátkosti vyjadriť ako procedúra, ktorá lineárne transformuje vektory čŕt tak, aby sa vybrali výsledné dimenzie, ktoré najlepšie separujú relevantné triedy. Použili sme nasledovné definície

$$S_{vnútri} = \frac{1}{C}\sum_{i=1}^{C}\frac{1}{\|\aleph_c\|}\sum_{i\in\aleph_c}^{\|\aleph_c\|}(x_i-\mu)(x_i-\mu)^T$$

$$S_{medzi} = \frac{1}{C}\sum_{i=1}^{C}(\mu_i-\mu)(\mu_i-\mu)^T \qquad (88)$$

gréckymi písmenami μ sme označili očakávané hodnoty, buď pre danú triedu μ_i alebo pre celé dáta. C je počet tried, a || || označuje kardinalitu množiny. S_{medzi} je stredná kovariančná matica, zatiaľčo $S_{vnútri}$ je kovariančná matica medzi triedami. Takto matica S_{medzi} vyjadruje to ako sú separované stredy tried, zatiaľčo $S_{vnútri}$ označuje stredný rozptyl, variabilitu dát vnútri danej triedy. Potom optimálna lineárna transformácia zvyšuje separáciu medzi triedami, pričom tak aby bol rozptyl vnútri tried malý. Nech je matica lineárnej transformácie **M**, ak posledné vyjadríme formálne, musí sa maximalizovať pomer definovaný Fischerom, pričom

$$\max_M \frac{\left|\det\left(MS_{medzi}M^T\right)\right|}{\left|\det\left(MS_{vnútri}M^T\right)\right|} \qquad (89)$$

Exaktné riešenie tohto problému maximalizácie je zovšeobecnený problém vlastných hodnôt daný ako

$$S_{medzi}\,\mathbf{w}_i = \lambda_i S_{vnútri}\,\mathbf{w}_i \qquad (90)$$

kde $\mathbf{w}_i$ je i-ty vlastný vektor matice $S_{vnútri}^{-1}S_{medzi}$, ktorý odpovedá vlastnej hodnote λ_i. Potom stĺpcový vektor, ktorý odpovedá najvyššej vlastnej hodnote sú riadkové vektory matice **M**, ako je definované v (89).

5.4.2 Potlačenie korelácií vnútri klasifikačných tried

Aplikácia LDA redukuje dimenzie vektorov čŕt zaisťujúc za určitých predpokladov najlepšie diskriminujúce vlastnosti. Hoci, v novom priestore, sú aj $S_{vnútri}$ aj S_{medzi} dekorelované, nie je to presne tak pre konkrétne triedy. Takto, aj z praktických dôvodov, je vhodné nájsť nejakú jednoduchú transformáciu, ktorá by (po aplikovaní LDA) minimalizovala korelácie vnútri tried. Formálne môžeme túto transformáciu vyjadriť ako

$$\min_{M} \frac{1}{\|\aleph_t\|} \sum_{i=1}^{C} \|\aleph_c\| \sum_{i=1}^{D} \sum_{j=i+1}^{D} cor_{ij}\left(MS_c M^T\right)^2 \qquad (91)$$

kde

$$cor_{ij}(a) = \frac{a_{ij}}{\sqrt{a_{ii}a_{jj}}}$$

S_c je kovariančná matica pre triedu c, $\|\ \|$ označuje kardinalitu množiny, D je nová dimenzia po transformácii LDA, a **M** je transformácia minimalizácie. V *Demuynck, Duchateau, Compernolle, Wambacq* (1998) sa používalo iteratívne riešenie s určitými obmedzujúcimi podmienkami, my sme ho nahradili dekorelačnou metódou úplneho postupného hľadania pomocou generovania množniny matíc rotácie pre daný prvok, každú s trochu odlišným uhlom rotácie. Matica rotácie, ktorá vedie k najmenšej ustrednenej korelácii sa vyberie ako hľadaná matica minimalizujúca (91). Po tomto sa algoritmus prevedie na ďalšom prvku; je to podobné ako iteratívna Jacobiho metóda riešenia problém vlastných hodnôt. Pri konvergencii tohto procesu sa konečná transformácia M dostane ako súčin najlepších rotačných matíc, ktoré sme vybrali v každom kroku ako

$$M = R_n R_{n-1} \ldots R_2 R_1 \qquad (92)$$

kde R_n je najlepšia matica rotácie v kroku n. Spôsob ako vybrať dekorelovaný element v n-tom kroku je buď postupnou elimináciou alebo vybrať prvok s najvyššou koreláciou skrz všetky triedy v konkrétnom čase.

5.4.3 Rotácia signálovej roviny

Z konštrukcie topologických invariantov tak ako sme ich opísali v časti 4.2, pozri aj Obr. 30 je jasné, že pre jediný harmonický signál, ktorý môžeme dostať filtrovaním ideálnym priepusťovým filtrom, dostaneme kružnicu a takto počet pretnutí osí súvisí s frekvenciou daného sgnálu. Avšak, ak si predstavíme viacero priepustí napr. kritické priepuste môžeme očakávať zložitejší obraz. Tento obraz bude závisieť na počte relevantných harmonických v konkrétnej priepusti, na pomere týchto frekvencií a na ich veľkostiach a ich fázových vzťahoch.

Obecne sa dá prijať, že rozdiely fáz medzi jednotlivými frekvenčnými zložkami v stacionárnom signále nenesú akusticky relevantnú informáciu, preto sme sa rozhodli vyskúšať rotáciu Nyquistovej roviny, čo je ekvivalentné kruhovému posuvu preseknutí vo vektore čŕt. Takto preseknutia pozdĺž osí sú kruhovo posunuté pre každú priepusť tak, že najvyššia hodnota je na prvom mieste atď.. Takto zložený signál dvoch harmonických signálov, ktoré sa líšia iba vo fázovom posune sa budú reprezentovať rovnakým vektorom čŕt. Takýmto spôsobom súčasný posun medzi dvomi harmonickými zložkami je potlačený, čo je invariantné pre určité podmienky pri ľudskej percepcii.

5.4.4 Začlenenie informácie o energii signálu

Medzi invariantnými črtami pri percepcii reči sme v časti 4.1.1 zaradili aj intenzitu rečového signálu. V tejto časti sme ju chápali ako celkovú úroveň intenzity zvukového signálu - obsah rečového signálu, ktorý je vyslovený celkovo potichu, stredne alebo silne sa nemení. Samozrejme nezačlenenie akejkoľvek lokálnej energetickej informácie do rozpoznávania reči vedie k nie paluzibilnému systému percepcie.

Z fonetickej a prozodickej analýzy vieme, že pre počuteľnosť zvukového signálu je dôležitý aj dôraz, a obecne lokálna zmena energie signálu. Obecne rečový signál je veľmi dynamicky bohatý, pričom fonémy v určitej vete alebo slove sú vyslovované s rôznou relatívnou intenzitou. Takto energia v celej vete - globálne alebo v nejakom jedinom časovom okne sa môže chápať ako invariantná ale hrá dôležitú úlohu v relatívnom pohľade, teda v časovej evolúcii medzi jednotlivými časovými oknami. Kvôli tomuto sme mieru relatívnej energie zakomponovali do vektora čŕt pre každé časové okno a frekvenčnú priepusť.

5.4.5. Podmienky učenia a testovania

Proces učenia je založený na schéme MASPER *Lindberg a kol.* (2000), čo je schéma učenia navrhnutá pre vývoj viacjazykových a krížovo-jazykových referenčných systémov rozpoznávania. V tejto práci spomenieme iba hlavné vlastnosti. Všetky fonémy sú modelované 3 stavovými modelmi spolu so 4 nie rečovými udalosťami. Počas 3 učiacich cyklov sa generujú 3 typy modelov: nezávislý na kontexte (CI) v prvých dvoch opakovaniach učenia a zviazané (CD) fonémy, všetky od 1 až po 32 Gaussovských zmiešaných modelov. Táto množina modelov je dôležitá, pretože takto je možné vybrať súbor modelov, ktoré sú najvhodnejšie pre konkrétnu aplikáciu.

Všetky výpočty sa vykonali na testovacej časti MOBILDAT-SK a aby sa pokryla celá diverzita aplikácií, previedli sa 3 druhy testov rozpoznávania: jednotlivé číslice, reťazce číslic a aplikačné slová. Test s reťazcom číslic je najťažšou úlohou a preto vedie k najvyšším chybám, ale zároveň používa iba obmedzený počet CD foném. Na druhej strane, test aplikačných slov obsahuje väčšiu rozmanitosť CI a CD foném a preto poskytuje objektívnejšiu mieru o kvalite učiacich modelov.

MOBILDAT-SK databáza sa zaznamenávala cez GSM sieť a skladá sa z 1100 hovoriacich, ktorý sú rozdelení do učiacej (880) a testovacej množiny (220). Každý hovoriaci nahovorí 50 záznamov o celkovej dobe medzi 4 až 8 minútami. Celkovo obsahuje 15942 rôznych slovenských slov, 41739 využiteľných rečových záznamov v učiacom režime, 51 slovenských foném-hlások, 10567 rôznych CD foném (vnútri slova) a spolu to dáva trochu viac ako 88 hodín reči.

5.5 Experimenty a výsledky

Aby sme overili navrhnuté modifikácie pôvodných čŕt previedli sme niekoľko experimentov. Všetky nastavenia testov sa overovali 3 testovacími scenármi pomocou všetkých možných modelov z učiaceho procesu. Všetky vektory čŕt obsahujú 39 elementov, 13 statických, 13 delta a 13 akceleračných koeficientov. Takto všetky nastavenia a transformácie aplikované na topologické invarianty produkujú 13 statických elementov na časové okno, čo prispieva podstatnou mierou k zjednodušeniu vyhodnocovacieho procesu.

5.5.1 Pôvodný návrh topologických invariantov

Prvý súbor experimentov overoval pôvodný koncept, a síce počet preseknutí cez rôzne osi. V tomto súbore experimentov sme použili 16 kritických priepustí z rozsahu 200-4000Hz a preseknutia pozdĺž 2 osí v každej priepusti, t.j. veľkosť vektora bola 32. Toto by malo zaistiť korektnú detekciu

jednotlivých harmonických frekvencií a v zložitejšej situácii identifikovať rozdiely. Zistili sme, že pri takom nastavení sa nedal uskutočniť učiaci proces, ako sme ho definovali v 5.1.4.5, až do konca. Príčinou takéhoto správania sa bola singularita alebo blízkosť singularite kovariančných matíc, čo zastavilo v konečnom dôsledku učiaci proces. Podobná situácia sa opakovala aj 8 frekvenčnými priepusťami (Mel frekvenčná škála) a detekujúc preseknutia na 4 osiach. Takto sme boli nútení usúdiť, že pôvodný prístup, bez nejakých modifikácií nie je vhodný pre HMM modelovanie.

5.5.2 Aplikovanie LDA na topologické invarianty

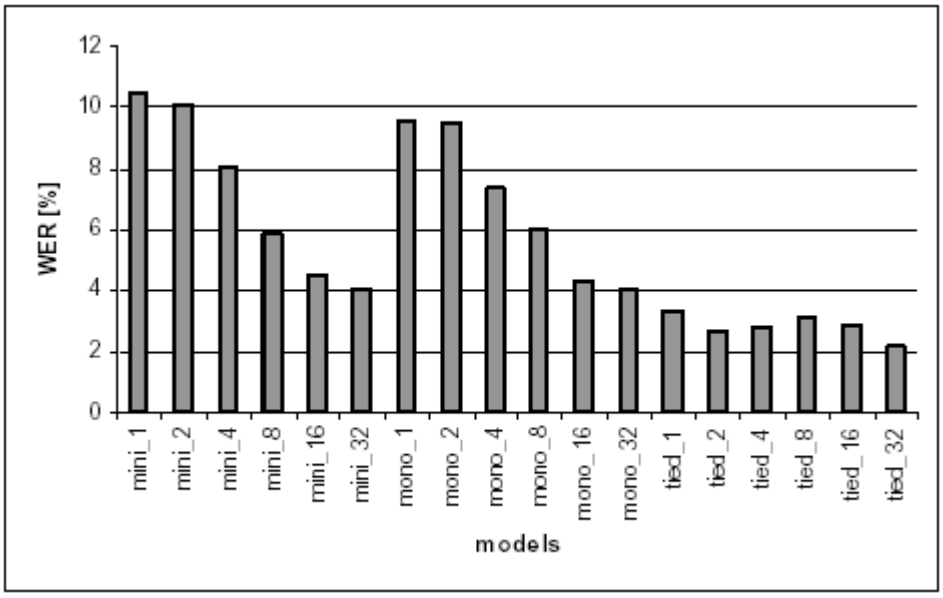
LDA transformácia sa konštruovala na časovo usporiadaných dátach a vektory čŕt sme zoskupili podľa ich príslušnosti k slovenským fonémam - hláskam (51 tried). Týmto spôsobom sme získali 13 statických čŕt na časové okno, ktoré boli z hľadiska lineárnej transformácie najdiskriminatívnejšie. LDA úspešne potlačila, v predošlom odstavci, spomínané singularity a takto sa mohli úspešne dokončiť všetky učiace cykly.

Avšak, dosiahnuté výsledky neboli uspokojujúce, pretože minimálne chyby na slovo (WER) boli nasledovné: aplikačné slová 14.23%, izolované číslice 14.9% a reťazce číslic 33.6% z WER. Pretože tieto výsledky nie sú použiteľné na súčasných aplikáciach, nebudeme sa venovať ostatným nastaveniam s takýmito jednoduchými modelmi topologických invariantov.

5.5.3 Zahrnutie informácie o energii

Na základe predošlej diskusie a diskusie v časti 5.4.4, sme pridali relatívnu energiu k topologickým invariantom. Takýmto spôsobom sa každý vektor čŕt rozšíril o toľko parametrov koľko bolo frekvenčných priepustí.

Avšak, všetky parametre sa na konci transformovali pomocou LDA na konečný statický vektor o dĺžke 13 prvkov. Tento postup viedol, pre 16 kritických priepustí a 2 preseknutia na priepusť, k veľkému zisku v úspešnosti rozpoznávania. Minimálne WER potom boli: 1.07%, 2.2%, a 2.75%, odpovedajúco pre izolované číslice, aplikačné slová a reťazce číslic. Detajlnejšia celková úspešnosť je uvedená na Obr. 38, kde sú uvedené jednotlivé WER - úspešnosti pre rôzne HMM modely a testy aplikačných slov.



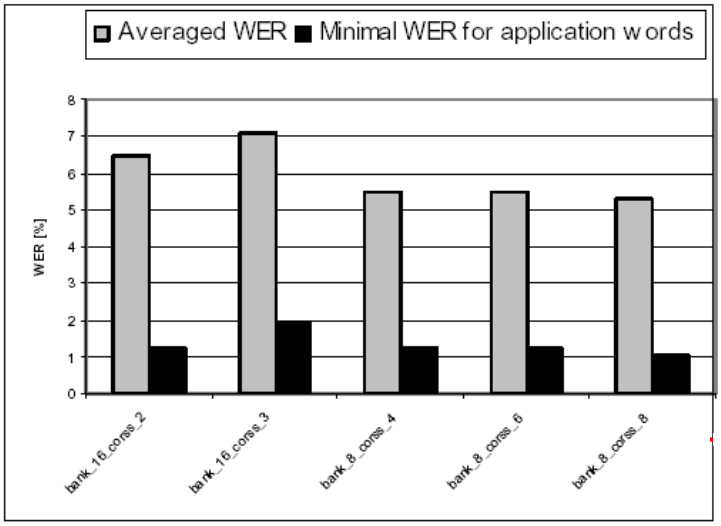
Obr. 38 Úspešnosti slov (WER) pre rôzne HMM (mini - CI 1.beh, mono - CI 2. beh, zviazané – CD 3. beh) so zmesmi v rozsahu od 1 do 32 Gaussiánov

Pretože táto modifikácia viedla k podstatnému vylepšeniu úspešnosti nášho modelu, rozhodli sme sa v ďalšom prevádzať testovanie už len s topologickými invariantami s relatívnou energiou a LDA. Ako je vidieť aj na Obr. 38 narastanie zložitosti modelu vedie k zlepšeniu výsledkov aj pre CI aj CD fonémy. Avšak, ďalšie zvýšenie nad 32 zmesí vedie len k malej výhodnosti, tu už pravdepodobne vstupuje do hry jav pretrénovania HMM.

5.5.4 Počet osí - invariantov a frekvenčných priepustí

Ako sme videli v predošlom, tieto dve nastavenia, t.j.počet osí pozdĺž ktorých sa merajú preseknutia a počet frekvenčných priepustí, do ktorých sa rozčlení primeraný frekvenčný rozsah (200-4000Hz) sú dôležité a preto sa musíme venovať ich analýze podrobnejšie. Pretože tieto parametre sú fyzikálne v blízkom vzťahu, čo vyplýva priamočiaro z Obr. 30, ich optimalizácia sa nedá previesť oddelene. Z elementárnych úvah vyplýva, že čím máme viac frekvenčných priepustí, tým je, v pragmatickom zmysle slova - pre všetky praktické účely, v danom frekvenčnom rozsahu menej harmonických zložiek. Takto tvar signálovej trajektórie (x(t), y(t)=x'(t)) - Nyquistov graf je podobný kružnici. Takto na jej popis je nutných menej osí; teoreticky jedna os pre jednu kružnicu. V takomto zjednodušenom prípade počet preseknutí cez ľubovoľnú os bude v priamom vzťahu k danej harmonickej frekvencii v danom pásme.

Na druhej strane, pri zmenšovaní počtu frekvenčných priepustí dochádza k obrátenému efektu, na popis signálu je treba viacero osía. V našich experimentoch sme nastavili ako maximálny počet filtrov-frekvenčných priepustí na 16, pretože to pojmovo odpovedá kritickým priepustiam pozdĺž Barkovej škály z rozsahu od 200 do 4000Hz. Ako opačný extrém sme použili iba 8 frekvenčných priepustí rovnomerne rozdelených od 200 do 4000Hz pozdĺž Mel škály. Toto je prakticky pravdepodobne najnižší počet filtrov ktorý je rozumne použiť v systémoch rozpoznávania. Pretože musíme uvažovať veľa kombinácií t.j. HMM modelov a 3 testovacie scénare, vybrali sme pre vyhodnotenie modelov a testov agregovanú formu, pričom sme zaznamenávali aj najlepšie výsledky.



Obr. 39 Ustrednená a minimálna úspešnosť (WER) pre rozličné počty priepustí a topologických invariantov (preseknutí)

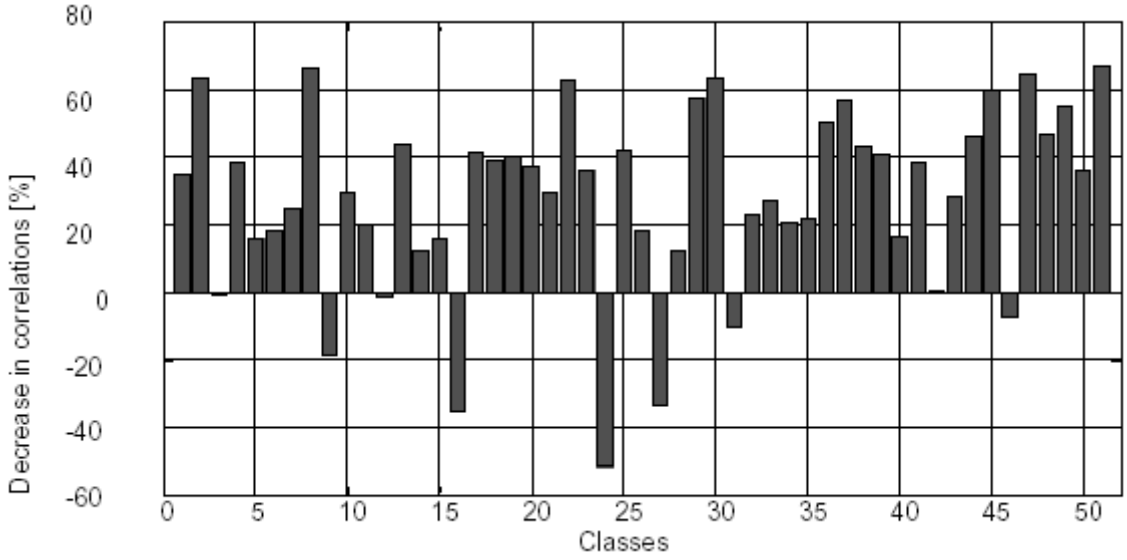
Takto na Obr. 39 sme znázornili stredné chyby pre testované počty frekvenčných priepustí (16, 8) spolu, v kombinácii s topologickými invariantami pre 3 testované scénare a učiace modely (CI a CD od 1 do 32 zmesí). Aby sme zohľadnili aj informáciu o najlepšej dosiahnuteľnej úspešnosti, uviedli sme tiež výsledky pre testy aplikačných slov. Ako vidíme z Obr. 39 aj 16 alebo 8 frekvenčných priepustí vedie k podobným výsledkom, ak sa vyberie v každom prípade primeraný počet preseknutí

(invariantov). Avšak kombinácie s menším počtom frekvenčných priepustí a väčším počtom invariantov vedú k vyššej úspešnosti.

5.5.5 Potlačenie korelácií vnútri tried

Ďalšou lineárnou transformáciou je taká transformácia, ktorá potlačí korelácie medzi elementami vnútri každej triedy, a takto sa dajú presnejšie fitovať pomocou diagonálnych kovariančných matíc. Dekorelačný proces sme previedli tak ako je uvedené v časti 5.4.2, kde sme vybrali pre dekorelačný proces element s najvyššou ustrednenou koreláciou skrz všetky triedy. Potom, sme pre tento element testovali súbor matíc rotácie tak, že rozdiel medzi susednými maticami bol okolo 3.6 stupňov. Tu sme aplikovali jednoduchý algoritmus - neopakovať znovu a znovu elimináciu rovnakého elementu ale náhodný výber odlišného elementu.

Na Obr. 40 je graficky znázornená efektívnosť tohto procesu ukázaním relatívneho poklesu maximálnych korelácií po aplikovaní dekorelačnej transformácie na každú z tried. Ako mieru efektívnosti sme použili relatívne zlepšenie.



Obr. 40 Relatívny pokles v maximálnch koreláciach vnútri každej triedy po aplikovaní dekorelačnej transformácie

V Tab. 10 sme zobrazili relatívne zlepšenie jednotlivých WER vo vzťahu k pôvodnému návrhu s LDA transformáciou. Zlepšenia sú uvedené v agregovanej forme - pre všetky testy a modely, a oddelene pre dva druhy nastavení: 16 frekvenčných priepustí a 2 invarianty a 8 frekvenčných priepustí s 6 invariantami. Taktiež sme uviedli relatívne zlepšenie najlepších WER pre najlepšie WER hodnoty a testovanie aplikačných slov. Ako je vidieť, aj pre agregované výsledky aj pre marginálne výsledky (minimálne hodnoty WER), sa pomocou dekorelačnej transformácie dosiahli dosť podstatné vylepšenia.

	16 priepustí 2 preseknutia	8 priepustí 6 preseknutí
Relatívne vylepšenie pre stredné WER	3.76%	11.44%
Relatívne vylepšenie pre minimálne WER	43.63%	18.95%

Tab. 10 Relatívne vylepšenie WER pre dekorelačnú transformáciu a dva typy čŕt

5.5.6 Rotácia Nyquistovej roviny

Znovu, aby sme videli výhody rotácie signálovej roviny, alebo Nyquistovho grafu pre naše signály previedli sme súbor experimentov s modelmi a testovacími podmienkami pre nasledovné nastavenia: 16 frekvenčných priepustí a 2 preseknutiami a 8 frekvenčných priepustí s 6 preseknutiami. Z množstva výsledkov, ktoré prezentujeme v Tab. 11 sme vytvorili, znovu agregovaním - uvažovaním všetky modely a testy oddelene pre spomenuté dve nastavenia. Výsledky sú uvedené vo forme vylepšenia oproti pôvodnému návrhu s aplikovaním LDA a dekorelačnou transformáciou. Vylepšenia sú uvedené aj pre najlepšie WER hodnoty zaznamenané skrz testy pre aplikačné slová.

	16 priepustí 2 preseknutia	8 priepustí 6 preseknutí
Relatívne vylepšenie pre stredné WER	7.69%	-6.47%
Relatívne vylepšenie pre minimálne WER	35.07%	35.07%

Tab. 11 Relatívne vylepšenie WER pre rotáciu Nyquistovho grafu a dva typy čŕt

Ako je vidieť z tabuľky navrhnutá rotácia sa ukázala úspešnou pre obe situácie v prípade16 frekvenčných priepustí a 2 preseknutí. Avšak, v prípade 8 frekvenčných priepustí a 6 preseknutí nie je jej výhoda taká zrejmá, pretože ustrednená chybovosť narástla o viac než 6%. Na druhej strane aplikácia rotácie Nyquistovej roviny je stále výhodnou v zmysle najlepšie fungujúcich modelov pre testy aplikačných slov. Tieto zistenia odpovedajú predošlej diskusii, 5.4.3, kde rotácia má jasnejšiu interpretáciu pre dve harmonické frekvencie (úzke frekvenčné priepuste), ale je o mnoho zložitejšia pre obecnejšie prípady.

6 Výsledky-verifikácia hovoriaceho

V tejto časti sa budeme zaoberať popisom a výsledkami, ktoré sa dotýkajú modelu verifikácie hovoriaceho. V prvej časti predložíme obecnejší úvod do parametrizácie reči pre verifikáciu hovoriaceho a následne v krátkosti popíšeme náš systém topologických invariantov, TIM. V druhej časti sa budeme venovať stručnému popisu klasifikačných procedúr vhodných pre úlohy verifikácie hovoriaceho a a poslednej časti uvedieme niektoré výsledky odvodené z experimentov v reálnych podmienkach.

6.1 Parametrizácia reči pre úlohy verifikácie

Akustické črty sú základom celého systému rozpoznávania, dobré črty by mali byť citlivé k rozdielom pre jednotlivých rôznych hovoriacich a mali by byť "hluché" k tým, ktoré nie sú podstatné pre náš sluchový systém na strane percepcie. Napríklad, poloha formantov v spektre a ich šírky sú dôležité pre diskrimináciu zvukov. Na druhej strane, nasledujúce aspekty nie sú tak dôležité: celková obálka spektra a jeho pokles, frekvencie pod frekvenciou prvého a nad frekvenciou tretieho formantu. Naviac, črty by mali byť necitlivé k aditívnemu a konvolučnému šumu, alebo aspoň by mali byť ľahko lokalizovateľné v priestore čŕt. Nakoniec, je zvykom aplikovať časové RASTA filtrovanie, aby sa potlačili poruchy, ktorých zdrojom sú rôzne prenosové kanály.

Produkcia reči sa dá za určitých podmienok modelovať lineárnym IIR filtrom, pozri časť 2.1.1.1 rovnica (3). Ak teda vieme prenosovú funkciu tohto filtra (napr. minimalizovaním lineárnej predikčnej chyby) môžeme odhadnúť lokalizáciu formantov a ich šírky

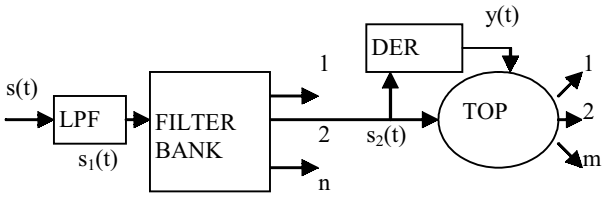
$$F = \arg(z_p)\frac{f_s}{2\pi}[Hz],\ \ B = -\log(|z_p|)\frac{f_s}{\pi}[Hz] \tag{93}$$

kde z_p je komplexný pól, F je poloha formantu pre jeden komplexný pól a B je jeho šírka. Avšak, ak sú prítomné póly v (3), potom rovnice (93) sú iba hrubými odhadmi, pretože viac pólov sa môže vzájomne ovplyvňovať.

Avšak, najpoužívanejšími črtami sú MFC, a PLP , ktoré sa používajú v spojitosti so systémami rozpoznávania. Toto sa môže zdať trochu prekvapujúce, pretože rozpoznávanie hovoriacich je v podstate opačná úloha, t.j. pri rozpoznávaní reči je cieľom potlačiť variabilitu v populácii tak aby zostala iba lexikálna časť. Avšak, sú navrhnuté tak, aby detekovali polohy a šírky formantov, ktoré sú akusticky percepovateľné a ľahko sa interpretujú a majú kompaktnú reprezentáciu, čiže napríklad jednoduchá miera vzdialenosti ako je euklidovská vzdialenosť má dobrý akustický význam.

Obe črty sú určitým typom modifikovaných kepstrálnych koeficientov, líšia sa len v spôsobe svojho výpočtu, pozri 2.1.1.2. MFC používa predzosilnené Mel a PLP Bark spektrá. PLP je zložitejšou procedúrou, pretože naviac k MFC aplikuje aj vyrovnávanie hlasitosti váhujúc napodobovanie ľudskej citlivosti, transformáciu intenzity na hlasitosť, umocnením intenzity na 0.3, pozri 2.1.12 vzťahy (10) a (11) a nakoniec len pólové modelovanie spektra, ktoré sa znovu konvertuje na kepstrálne črty, *Hönig, Stemmer, Hacker, Brugnara* (2005), *Hermansky, Hanson, Wakita* (1985).

Návrh nášho TIM (Topological Invariant Model) modelu pre verifikáciu hovoriaceho sleduje niektoré fyziologické mechanizmy rečovej percepcie tak ako sme ju opísali v časti 4 a *v Chudý, Chudý, Hapák* (1991). Takto každá časť nášho modelu má svojho partnera vo fyziológii, aj keď je toto spojenie len kvalitatívne, detajlnejšie napríklad v *Kačúr, Chudý* (2012), *Pickles, J. O.* (1988). Schématické časti parametrizácie modelu sú na Obr 41.



Obr. 41 Model TIM verifikácie hovoriaceho pomocou topologických invariantov

Na základe požiadaviek na výslednú dimenziu vektora čŕt sme zafixovali, podľa predbežných odhadov založených na experimentoch na malom korpuse dát, základný typ vektora čŕt nasledovne. Použili sme 4 frekvenčné priepuste, pozri Obr. 41, s Barkovou škálou rozdelenou ako 0.1–0.4; 0.5–0.9; 1.0–1.7; 1.9–3.2 kHz. Na aktuálne filtrovanie sme použili návrh Kaiserovho filtra s obecnými charakteristikami ako je zoslabenie filtra rovné 15dB a Gibbsove kmitanie rovné 0.1dB, časť 4.

Parametrizácia, naše vektory čŕt sa skladali z nezávislých čŕt – preseknutia definované v časti 4.2 vynásobené časovým trvaním, pozri vzťah (62) pre danú frekvenčnú priepusť. Toto násobenie vyjadruje relatívnu váhu rôznych sektorov a frekvenčných rozsahov. Prevádza určitý typ odhadu relatívnej pravdepodobnosti danej črty – preseknutia v danom frekvenčnom rozsahu a sektore Nyquistovej roviny, Obr. 30.

6.2 Metódy klasifikácie pre úlohy verifikácie

Každý systém rozpoznávania hovoriaceho musí v koncových štádiach implementovať nejaký algoritmus rozhodovania. Existuje mnoho systémov *Mitchell* (1997), od teoretických konceptov ako je Bayesovský klasifikátor až po prakticky realizovateľné ako sú GMM, ANN, rozhodovacie stromy, atď. Obecne sa tieto metódy dajú klasifikovať do niekoľkých skupín podľa rôznych kritérií.

V prvom rade sa dajú opísať ako deskriptívne metódy, ktorých cieľom je modelovať čo najpresnejšie priestor čŕt - príkladom môžu slúžiť GMM (s ML alebo MAP učiacimi kritériami, nie opravujúce sa), a diskriminačné metódy ako sú ANN, SVM, kde sa hlavný dôraz kladie na separovanie rôznych tried (namiesto presného modelovania priestoru čŕt). Iným spôsobom sa dajú rozdeliť na parametrické a neparametrické metódy (učenie na základe príkladu). Pre parametrické metódy sa učiace dáta používajú na prevedenie odhadu parametrov modelu tak, aby sa fitovalo relevantné pravdepodobnostné rozdelenie alebo minimalizovala chyba klasifikácie, napr. ANN, SVM, GMM, atď. Tieto metódy sú schopné sa vysporiadať s nedostatkom učiacich dát a previesť určité zovšeobecnenia. Obyčajne sú v reálnych podmienkach úspešné.

Na druhej strane, metódy založené na učení z príkladov len uchovávajú všetky učiace sa dáta a používajú ich na rozhodovanie o príslušnosti k danej triede, zvyčajne iba na základe vzdialenosti od uchovaných učiacich dát. Typickým príkladom je metóda kNN, ktorá bez ohľadu na jej jednoduchosť môže viesť k vynikajúcim výsledkom za predpokladu, že je dostatočný počet učiacich vzoriek. Tieto metódy nezavádzajú určité umelé predpoklady o modeloch (pravdepodobnostných rozdeleniach), predpokladajú zvyčajne iba nejaký typ miery vzdialenosti, čo pre niektoré typy čŕt nemusí byť známe alebo jasné.

V našich experimentoch sme použili špeciálny druh neurónovej siete, takže urobme malú rekapituláciu niektorých základných faktov o ANN. Existujú rôzne ANN architektúry, ale najčastejšie používanými sú siete ako viacvrstvové perceptróny (MLP) a radiálne bázové funkcie (RBF). Bolo dokázané, že obe siete sú univerzálne aproximátory, čo znamená že môžu aproximovať ľubovolnú

spojitú funkciu s ľubovolne malou chybou, *Mitchell* (1997). Preto sa dajú navrhnúť a naučiť za predpokladu daných učiacich dát tak, že vedú k najlepšej možnej separácii a môžu dobre zovšeobecňovať pre neučiace vzorky. Avšak, optimálny počet neurónov ne je známy a zatiaľ známe učiace stratégie negarantujú dosiahnutie globálneho optima. Naviac, ANN sú náchylné na jav prefitovania, takže počas učiacej fázy sa musí prevádzať krížová validácia. RBF sieť prevádza lokálne priblíženie (dané centrami siete) a pomocou vhodnej regularizácie sa dá potlačiť prefitovanie. Na druhej strane, MLP prevádza globálnu aproximáciu.

V našom prístupe použitá sieť PNN, alebo "Probabilistic Neural Network" je výraz použitý *Specht* (1990) pre kernel diskriminačnú analýzu. Môžeme si túto sieť predstaviť ako normalizovanú RBF sieť, v ktorej máme pre každý učiaci vzor nejaký skrytý neurón - jednotku, ktorá je centrovaná okolo tohto paternu. Tieto RBF jednotky sa nazývajú „jadrá" a sú to obyčajne funkcie hustoty pravdepodobnosti, také ako gaussovské funkcie. Váhy od skrytej do výstupnej vrstvy sú zvyčajne rovné 1 alebo 0; pre každú skrytú jednotku, pričom váha 1 sa použije pre spojenie idúce do výstupu, do ktorého daná vzorka patrí, pričom všetky ostatné spojenia majú nulové váhy, pozri časť 4.3.

Alternatívne môžeme tieto váhy adjustovať pre apriórne pravdepodobnosti pre každú triedu. Takto jediné váhy, ktoré treba modifikovať učením sú šírky jednotiek RBF. Tieto váhy sa nazývajú „hladiace parametre" a zvyčajne sa určujú pomocou krížovej validácie. Pretože sieť RBF môže prevádzať zložité lokálne aproximácie s ľubovolnou chybou, má rýchle inkrementálne učenie, a poskytuje jednoduchý nástroj na kontrolovanie robustnosti k zašumeným vzorkám (pomocou disperzných parametrov), rozhodli sme sa v ďalších experimentoch používať jej modifikovanú verziu a síce PNN, pozri časť 4.3.

6.3 Podmienky učenia a testovania

Na experimentoch, ktoré sme vykonali a pri zozbieraní dát sa zúčastnilo 26 ľudí. Všetci z nich boli buď natívni hovoriaci (20) alebo cudzinci (6), ale všetci hovorili plynulo anglicky. Medzi nimi bolo 11 žien a 15 mužov, vek od 23 do 66 rokov. Takisto medzi nimi boli 3 páry s rodinnými vzťahmi (2 páry boli bratia, 1 pár bol otec a syn), ktorých hlasy sa prirodzene podobali a v reálnych situáciach boli ich hlasy ťažko odlíšiteľné (aj ľudským agentom rozpoznávania).

Hardvérová platforma, na ktorej sa robila celá komunikácia a experimenty bolo telefonické rozhranie DIALOGIC/4D (1992) s 70dB SNR, a frekvenčnou odozvou v rozsahu od 300 Hz do 3500 Hz, na strane verifikačného systému, a obyčajný telefón s tónovo/impulznou voľbou na strane hovoriaceho. V tomto experimentálnom návrhu každý hovoriaci nahovoril 3 realizácie na jedno zo slov – "cry", "ocean", "daddy", "void", "voyage", "eleven" v učiacej fáze, a počas testovacej fázy sa žiadalo od každého hovoriaceho, aby povedal 3 slová z predošlého zoznamu. Začiatok a koniec slov, procedúra VAD, sme určovali naším algoritmom, pozri dodatok A1.

Hovoriaci sa testovali individuálne za normálnych zvukových podmienok a pomocou štandardného telefonického spojenia. Zber dát aj experimentovanie sa prevádzalo automaticky počas 24 hodín na dennej báze, pričom účastníci si volili prístup k systému individuálne, od prípadu k prípadu. Prístup bol pomocou obyčajného telefónu s tónovou voľbou so systémom Caller ID. Náš systém sa dá rozdeliť, z experimentálneho hľadiska, na dve základné fázy verifikačného procesu – učiacu a tetovaciu fázu.

Učiaca fáza sa skladala z troch podúloh:

• COLLECT - program pre zber PCM súborov, beží v režime reálneho času.

• EXTRACT - off-line (databázový) program pre extrakciu čŕt z databázy PCM súborov. Jeho výstupom sú vektory čŕt pre prototypy pre odpovedajúceho hovoriaceho a dané slovo.

• ADAPT - off-line (databázový) program pre adaptáciu (optimalizáciu) parametrov neurónovej siete, v našom prípade PNN.

Testovacia časť sa skladá z jedinej úlohy:

• ACCEPT - program v režime reálneho času, ktorý prevádza verifikáciu hovoriaceho v reálnom čase na základe nahovorených troch slov (buď pomocou jedného slova - 1-slovná verifikácia, alebo pomocou všetkých troch prístupných slov - 3-slovná verifikácia). Tento program zahŕňa 3 základné procesy: nahranie PCM súborov, extrakcia čŕt z týchto súborov pomocou zvolenej TIM metódy, verifikáciu hovoriaceho na základe vektora čŕt z predošlého kroku pomocou zvolenej diskriminačnej metódy.

Prirodzené časové poradie v rámci nášho systému obsahuje práve zmienené procesy (COLLECT, EXTRACT, ADAPT, ACCEPT). Každý následný proces vyžaduje úspešné splnenie predošlého procesu a existenciu už vytvorených objektov Phonebook a Wordbook. V nich je definovaná celá nutná informácia o klientoch a slovách, ktoré sa používajú pre testovanie. Takto proces je definovaný ako entita, ktorá prevádza všetky činnosti a uchováva celú informáciu o konkrétnej úlohe. Pre znázornenie celého mechanizmu zbierania dát, učnie a testovania rozoberieme podrobnejšie úlohu ACCEPT.

Procedúra Accept je podúlohou určenou na testovanie hlasov klienta pomocou verifikačného procesu. Celá informácia a klientoch a slovách, ktoré vstupujú do tohto procesu je zapísaná v konkrétnom PHONEBOOK a WORDBOOK, ktoré sa dajú vybrať z predtým už vytvoreného zoznamu phonebook-ov a/alebo wordbook-ov.

ACCEPT je procedúrou reálneho času vyžadujúcou telefonické rozhranie DIALOGIC. Môže byť prerušená a znovu obnovená (po prerušení užívateľom alebo pri prerušení z dôvodov fatálnej softvérovej alebo hardvérovej chyby). Procedúra Accept náhodne vyberá záznam klienta z fronty a rozhoduje sa s pravdepodobnostou 0.5 či sa prevedie volanie tomuto klientovi. Ak je rozhodnutie „nie“, potom program ACCEPT prejde na další záznam vo fronte.

Ak je rozhodnutie „áno“, potom ACCEPT skontroluje záznam vo fronte, či má správny dátum a čas, aby sa mohol previesť telefonát prostredníctvom rozhrania DIALOGIC. Ak sú podmienky splnené potom sa prevedie volanie. Rozhranie DIALOGIC bude monitorovať eventy „BUSY“, „NO ANSWER“, alebo „CONNECTED“. Ak nie je splnená podmienka „CONNECTED“ hovor bude ukončený a zobrazí sa stav eventu v správe o vývoji pokusu, aktualizujú sa ACCEPT RESULT - Summary Log, Details Log a vyberie sa další záznam na spracovanie. Ak je splnená podmienka, potom sa volanej strane oznámi, že má napríklad 1 minútu na opätovanie hovoru. Systém potom očakáva spätný hovor.

Ak klient opätuje hovor, systém overí prístupový kód klienta, pričom tento nemusí byť zadaný, hovoríme o voľnom prístupe alebo klient bude vyzvaný zadať DTMF číslice a tieto sa overia, alebo systém zaznamená Caller ID (ICLID). Ak sa prístupový kód zhoduje so záznamom klienta, proces bude pokračovať do verifikačnej fázy. Počas tohto procesu sa volajúci vyzve, aby povedal 3 náhodne vybrané slová (z 6 možných). Volajúcemu sa tiež oznámi ukončenie hovoru, hovor sa ukončí a DIALOGIC zavesí.

Po tejto fáze sa vybrané tri slová spracujú testovacou verifikačnou procedúrou, pomocou extrakcie čŕt cez TIM a diskrimináciou pomocou Probabilistic Neural Network, PNN. Potom systém aktualizuje správu o vývoji pokusu a ACCEPT RESULT. Po tomto systém pokračuje dalším záznamom. Procedúra ACCEPT musí prijať všetku nutnú informáciu od predošlej procedúry ADAPT. To sa prevádza výberom konkrétneho procesu zo zoznamu už vytvorených procesov ADAPT.

Správa o vývoji pokusu zobrazuje informáciu o stave pokusu, meno volaného klienta a štatistiku o súčasnom počte pokusov, chýb a volaní. Posledný blok informácie prezentuje štatistiku o pomeroch zlyhaní pre1-slovné INTRA,1-slovné EXTRA, 3-slovné INTRA a 3-slovné EXTRA prípady verifikačného procesu vtedy ak bol pokus úspešný, t.j. ak bolo úspešné volanie klientovi. 1-slovné alebo 3-slovné rozhodovacie spôsoby verifikácie odpovedajú rozhodovaniu na základe jedného vysloveného slova alebo všetkých troch naraz.

6.4 Základné experimenty na diskriminačné schopnosti

V tejto časti uvedieme výsledky niektorých experimentov pre vyššie uvedené reprezentácie topologických invariantov, 128 (4 filtre x 32 topologických invariantov) zložkový vektor, v závislosti od rôznych podmienok diskriminácie. Uvedieme výsledky opisujúce vplyv metódy diskriminácie pomocou DTW vzdialenosti, pričom tu sa počet časových okien sa mení od realizácie k realizácii, a aj pre učiace prototypy, porovnáme ho s diskrimináciou pomocou euklidovskej vzdialenosti pre fixný počet časových okien a nakoniec pomocou diskriminácie založenej na PNN znovu s fixným počtom časových okien. Počet časových okien, ktorý je v rozsahu 30 až 64, sme fixovali jednoduchou interpoláciou na 4 časové okná, túto interpoláciu sme prevádzali iba pre euklidovskú a PNN metódu diskriminácie.

Vektory čŕt, vypočítané pre každé časové okno, sú interpolované takým spôsobom, že všetky realizácie slov majú rovnaký počet výsledných (interpolovaných) vektorov čŕt, čo je vlastne určitý spôsob normalizácie podľa časového trvania. Pretože, všetky sekvencie vektorov čŕt majú teraz rovnakú fixovanú dĺžku, môžeme ich segmentovať do rovnakého počtu segmentov (pre testované slová, kvôli presnosti, výpočtovým a pamätovým požiadavkám sme použili 4 segmenty).

Vo všetkých simulačných experimentoch boli nahovorené tri učiace prototypy pre každého hovoriaceho a každé slovo.

Uvedené skóre zlyhania sa dotýkajú úlohy verifikácie, keď sa od osoby vyžaduje povedať iba jediné slovo, t.j. verifikácia pomocou jedného slova, už spomínaný, na konci predošlej časti, 1-slovný typ verifikácie.

Akceptačné rozhodnutie podľa akceptačného kritéria(špecifického pre každú metódu) sme previedli pre každú konkrétnu testovanú realizáciu v porovnaní k všetkým (3) prototypom (alebo ich kombinácii v prípade pravdepodobnostnej neurónovej siete). Vyhodnotenie výkonnosti sme urobili nasledovným spôsobom. Pre DTW a euklidovskú metódu definujeme prah vzdialenosti ako

$$prah = meandist + a*(meandist - maxdist) \qquad (94)$$

kde meandist je aritmetická stredná hodnota vzdialenosti, maxdist je maximálna vzdialenosť skrz všetky DTW alebo euklidovské vzdialenosti medzi prototypmi vo vzťahu ku konkrétnemu slovu a konkrétnemu hovoriacemu, a je empirický parameter fixovaný dopredu, na základe viacerých experimentov ako a = 0,2. Pre PNN metódu tento prah vzdialenosti korešponduje disperznému parametru, $\sigma = 1,5$. Rozhodovaciu stratégiu v tomto prípade prediskutujeme neskôr.

Potom v prípade DTW a euklidovskej diskriminácie odlišujeme dva prípady:

A. Ak prototyp a konkrétna testovaná realizácia sú z rovnakého hovoriaceho urobíme rozhodnutie o zlyhaní, ak je ich vzdialenosť väčšia než prah vzdialenosti.

B. Ak prototyp a konkrétna testovaná realizácia nie sú z rovnakého hovoriaceho urobíme rozhodnutie o zlyhaní, ak je ich vzdialenosť menšia než prah vzdialenosti.

Ako základný štatistický výsledok vyhodnotíme počet intra -hovoriaci zlyhaní a extra-hovoriaci zlyhaní pre každého hovoriaceho a každé slovo v databáze.

V prípade metódy PNN sa vytvára pre každého hovoriaceho a slovo štatistická klusterová reprezentácia založená na 3 prototypoch. V tomto prípade konkrétna realizácia sa porovnáva iba k jednej klusterovej reprezentácii pre každého hovoriaceho a slovo (a nie pre 3 prototypy oddelene).

Potom, verifikácia predpokladanej identity sa realizuje ako štatistický rozhodovací proces, ktorý testuje nasledujúce hypotézy:

A. Vstupný rečový patern patrí do triedy INTRA hovoriaceho X (t.j. patrí hovoriacemu X).

B. Vstupný rečový patern patrí do triedy EXTRA hovoriaceho X, čo je množina zostávajúcich hovoriacich.

Na základe znalosti, predpokladaná identita hlasu hovoriaceho bola pravdivá alebo nie sa potom vypočítajú úspešnosti verfikácie. Rozhodnutie, či testovaný patern patrí do INTRA/EXTRA triedy je založené na nájdení najbližšieho suseda k učiacemu paternu z celej databázy, t.j. z oboch INTRA aj EXTRA učiacich množín.

PNN metóda odhaduje hodnoty pravdepodobnosti testovaného vzoru vzhľadom k aktuálnym INTRA a EXTRA triedam. Sú dané vzťahmi

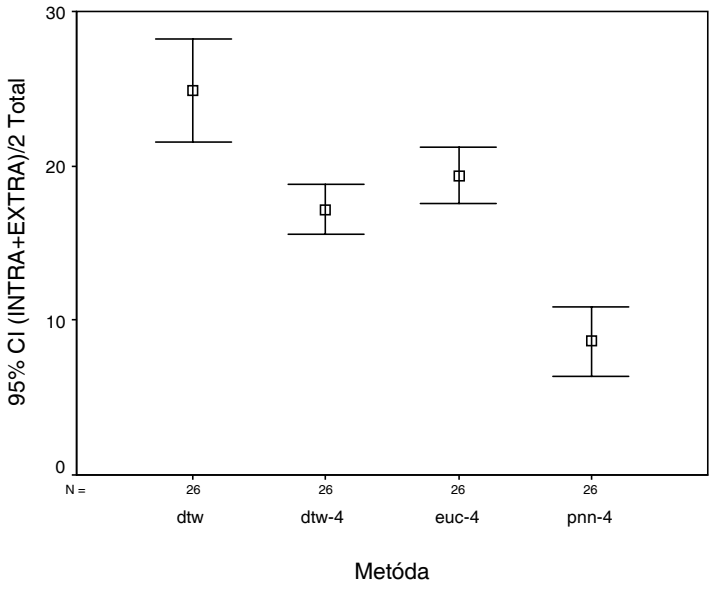
$$P_{intra,\,extra} = \frac{1}{N_{in,ex}} \sum_{i=0}^{N_{in,ex}-1} \exp\left(-\left\| x^t - x_i^{in,ex} \right\| / \sigma\right) \qquad (95)$$

kde χ^t opisuje testovaný vektor čŕt a χ_i^{in}, χ_i^{ex} zase učiace vektory čŕt aktuálnych INTRA a EXTRA tried, resp.. $\| \ \|$ vyjadruje euklidovskú vzdialenosť dvoch vektorov.

$$Ak: P_{intra} > P_{extra} \qquad (96)$$

potom je vstupný vektor čŕt priradený do triedy INTRA ináč do EXTRA triedy.

Na Obr. 42 sú uvedené intervaly spoľahlivosti pre strednú chybovosť INTRA a EXTRA zlyhaní v závislosti od metódy diskriminácie. Podobne v Tab. 12 sú zobrazené popisné štatistiky stredných chybovostí INTRA a EXTRA zlyhaní pre jednotlivé metódy diskriminácie. Treba si všimnúť, že euklidovská metóda porovnania, so zafixovaným počtom časových okien, dáva signifikantne lepšie výsledky ako DTW s premenným počtom okien a len o trochu horšie výsledky ako metóda DTW na rovnakých paternoch. Prah vzdialenosti bol určený ako a = 0.2 a pre PNN σ = 1.5.



Obr. 42 Intervaly spoľahlivosti pre strednú chybovosť INTRA a EXTRA zlyhaní v % v závislosti od metódy diskriminácie - DTW- premenlivý počet časových okien, DTW-4 - štyri časové segmenty, EUC-4 - štyri časové segmenty, PNN-4 - štyri časové segmenty

Prezentované výsledky vedú k nasledujúcim záverom:

1. V predošlej časti definované topologické parametre sú vhodnými črtami pre úlohu verifikácie hovoriaceho, pretože ak sa transformujú na nízko rozmerný komprimovaný tvar sú schopné zhlukovať podľa hovoriaceho s 80 % presnosťou iba pomocou najjednoduchšej ale zároveň najrobustnejšej a výpočtovo najmenej náročnej metódy porovnávania euklidovskej vzdialenosti, pozri (94).

(INTRA +EXTRA)/2 Total	M	SD
DTW	24,8696	8,3648
DTW-4	17,1788	4,0952
EUC-4	19,3596	4,5230
PNN-4	8,6654	5,5686

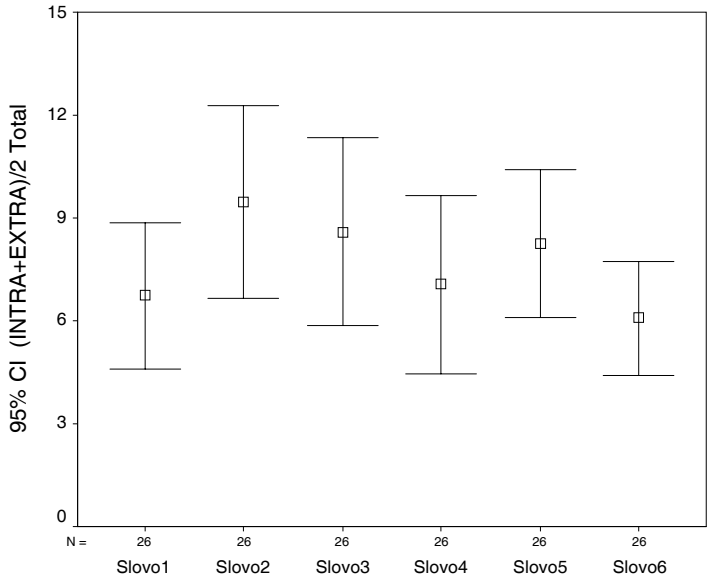
Tab. 12 Popisné štatistiky pre strednú chybovosť INTRA a EXTRA zlyhaní v % v závislosti od metódy diskriminácie - DTW- premenlivý počet časových okien, DTW-4 - štyri časové segmenty, EUC-4 - štyri časové segmenty, PNN-4 - štyri časové segmenty

2. DTW metódy nevylepšujú podstatnou mierou presnosť (alebo presnejšie vylepšujú ju iba o trochu).

Obecne, výsledky založené na týchto komprimovaných reprezentáciach sú ešte lepšie.

3. Výsledky založené na metóde PNN sú signifikantne lepšie (lepšie ako 90% presnosť) než pre iné vyšetrované metódy. Naviac, metóda PNN je menej výpočtovo náročná v porovnaní s DTW.

Na Obr. 43 sme uviedli intervaly spoľahlivosti pre stredné chybovosti INTRA a EXTRA zlyhaní v závislosti od daného slova pre PNN metódu diskriminácie (4 časové segmenty). Predpokladáme, že medzi jednotlivými slovami nebudú rozdiely v úspešnosti. Naozaj vidíme, že rôzne slová majú síce rôzne presnosti rozpoznávania, ale medzi jednotlivými slovami nie je štatisticky signifikantný rozdiel v chybovostiach, (F=1,289; sig=0,273; df=5). Kde F je Fisherova štatistika, sig znamená signifikancia daného testu a df znamená počet stupňov voľnosti. Posledný záver sme urobili pomocou GLM procedúry, opakovaných meraní v SPSS 8 za predpokladu preukázania sféricity (Chi^2=15,811; sig=0,328;df=14). V Tab. 13 sú zobrazené popisné štatistiky stredných chybovostí INTRA a EXTRA zlyhaní pre jednotlivé slová pre PNN-4 metódu diskriminácie.



Obr. 43 Intervaly spoľahlivosti pre strednú chybovosť INTRA a EXTRA zlyhaní v % v závislosti od slova pre diskriminačnú metódu PNN-4 - štyri časové segmenty

Takže, napríklad ak pracujeme iba s 3 vhodne vybranými slovami celkové výsledky sa môžu podstatne vylepšiť. Toto sa využíva pri metodike verifikácie, ktorú sme nazvali 3-slovná, keď sa diskriminačného procesu zúčastnia 3 slová naraz, pozri ďalej. Prah vzdialenosti bol určený takým istým spôsobom ako v (94) rovný a = 0.2, analogicky v rámci prístupu PNN disperzný parameter $\sigma = 1.5$, pozri vzťahy (95) a (96).

	M	SD
Slovo 1	6,7308	5,2635
Slovo 2	9,4712	6,9650
Slovo 3	8,5962	6,7628
Slovo 4	7,0577	6,3915
Slovo 5	8,2596	5,3486
Slovo 6	6,0769	4,1102

Tab. 13 Popisné štatistiky pre strednú chybovosť INTRA a EXTRA zlyhaní v závislosti od slova pre metódu diskriminácie PNN-4 - štyri časové segmenty

6.5 Diskriminačné schopnosti PNN modelu verifikácie

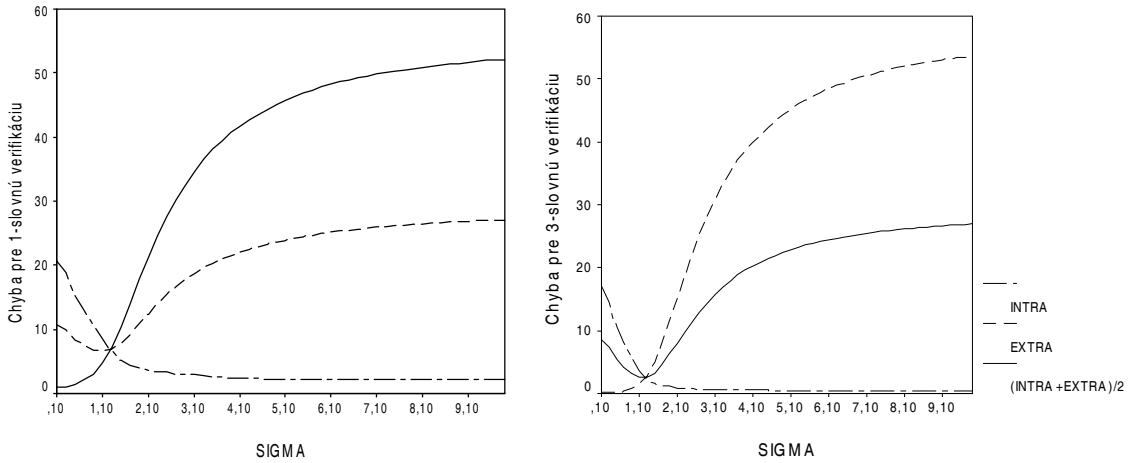
Takto sme nakoniec na základe predošlých experimentov vybrali nasledovnú reprezentáciu hlasu hovoriaceho: 4 frekvenčné priepuste, každá rozdelená do 4 segmentov v čase pre všetky nasledujúce experimenty. Konečná reprezentácia je daná vlastne počtom vyšetrovaných osí, resp. oblastí, Obr. 30 v časti 4., v ďalšom sa v našej parametrizácii uvažuje iba vrchná časť grafu. Inými slovami, stačí uvažovať iba polovicu topologických invariantov, čo výsledkovo prediskutujeme neskôr.

Treba poznamenať, že parameter PNN σ reprezentuje disperzný parameter, ktorý sa optimalizuje podľa vybraných kritérií a môže byť nejakou globálnou (pre všetkých hovoriacich a všetky slová jeden parameter) fixovanou hodnotou ako aj lokálnou (obecne rôzny pre hovoriacich a aj pre slová).

V ďalšom sme implementovali a vyhodnotili 2 typy verifikačných procesov: a) 1-slovný - identita hovoriaceho je prijatá alebo odmietnutá na základe jediného vysloveného slova, a rozhodovací proces je identický s uvedeným na predošlých riadkoch, b) 3-slovný - dentita hovoriaceho je rozhodnutá na základe troch rôznych slov - trojica slov (v ďalšom ju budeme nazývať obecne položka). Proces rozhodnutia v poslednom prípade používa rozhodnutia urobené na jednom slove a konečné rozhodnutie je založené na princípe - väčšina vyhráva.

6.5.1 Disperzný parameter a PNN diskriminácia

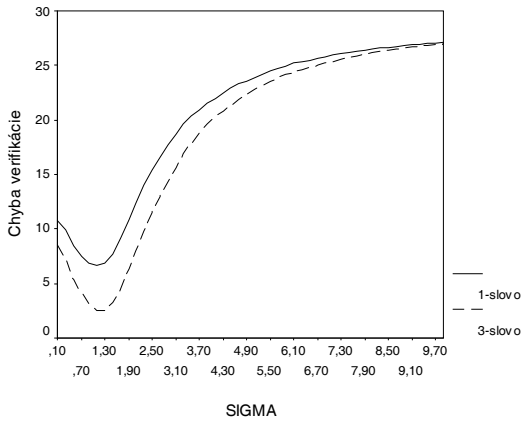
Prvý experiment, ktorý sme previedli v kontexte PNN a jej závislosti na disperznom parametre, spočíval v otestovaní celkovej úspešnosti na tomto parametri neurónovej siete, pozri kapitolu 4. Správanie sa úspešnosti z hľadiska disperzného parametra pre optimalizácie na 1-slovo a 3-slová je zobrazené na Obr.44.



Obr. 44 Chybovosťv % pre 1-slovný (vľavo) a 3-slovný (vpravo) rozhodovací proces v závislosti od sigma – parametra disperzie PNN siete

Z Obr. 44 vidíme, že chybovosť verifikácie má opačnú tendenciu pre INTRA a EXTRA diskriminácie, čo je odpovedajúce, v podstate triviálne správanie. „Plató“ je dostatočne široké k schopnosti vybrať si optimálnu hodnotu parametra v zmysle (INTRA+EXTRA)/2 chybovosti. V našom prípade to odpovedá intervalu približne od 0,75 po 2,0. V tomto intervale je aj INTRA aj EXTRA chybovosť optimálna.

Na Obr. 45 je celková chybovosť (priemer INTRA a EXTRA chybovostí) 1-slovného v porovnaní s 3-slovným rozhodovacím procesom. Z Obr. 45 vidíme, že 3-slovná chybovosť je približne 3 násobne menšia ako pre 1-slovnú, pre optimálne vybraný disperzný parameter. Pre vysoké hodnoty disperzného parametra sa chybovosti 1-slovného a 3-slovného procesu verifikácie rpibližujú k sebe, v limite veľmi vysokých hodnôt disperzného parametra sú rovnaké. Na opačnom konci, pre malé hodnoty disperzného parametra je 3-slovná verifikácia približne dvojnásobná oproti 1-slovnej verifikácii.



Obr. 45 Chybovosťverifikácie v % porovnaná pre 1-slovný (plná čiara) a 3-slovný (prerušovaná čiara) rozhodovací proces v závislosti od sigma – parametra disperzie PNN siete

6.5.2 Šum a PNN diskriminácia

V druhom type experimentov sme sa snažili vyšetriť závislosť našej verifikačnej metódy na šume, čo odpovedá interferencii šumového paternu so signálom aj multiplikatívnym aj aditívnym spôsobmi. Zašumenie pôvodného signálu sme definovali ako

$$S_n(t) = S_0(t) + \text{randGauss(dev)} * S_0(t) \qquad (97)$$

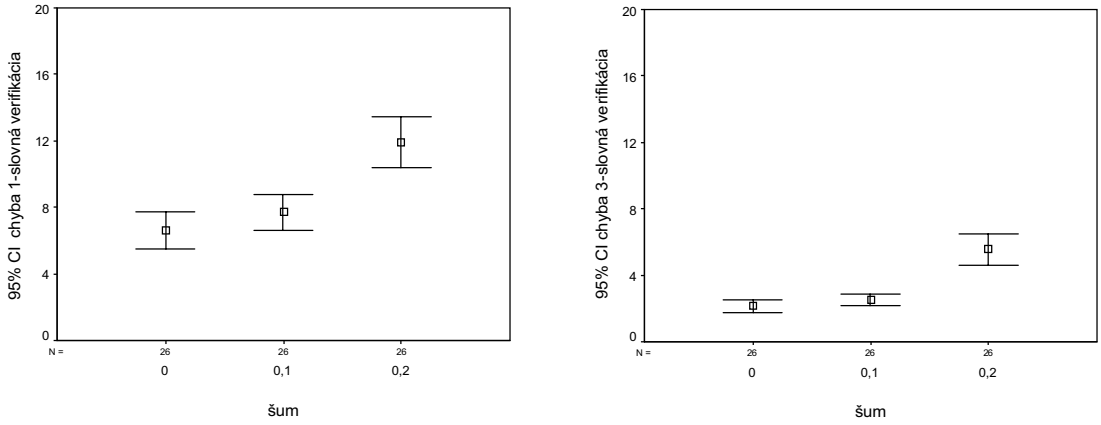
kde $S_0(t)$ je pôvodný signál, $S_n(t)$ je zašumený signál pomocou Gaussovho generátora náhodných čísiel randGauss(dev) s nulovou strednou hodnotou a štandartnou odchýlkou dev = (0,0; 0,1; 0,2).

Previedli sme experiment ako v predošlom prípade, ale so všetkými testovanými signálmi zašumenými s multiplikatívnym gaussovským šumom. Proporcie šumu sú dané v nasledujúcej tabuľke, Tab. 14, spolu s INTRA a EXTRA chybovosťami 1-slovného verifikačného procesu (pre všetkých 6 slov). Predpokladáme, že bude rozdiel v úspešnosti pre veľké šumy.

	Slovo1		Slovo2		Slovo3		Slovo4		Slovo5		Slovo6	
	INTRA	EXTRA	INTRA	EXTRA	INTRA	EXTRA	INTRA	EXTRA	INTRA	EXTRA	INTRA	EXTRA
dev 0	3.8	4.4	11.5	6.1	13.0	4.7	5.8	7.8	6.3	7.2	4.8	3.7
devR 0,1	3.8	9.6	10.1	8.8	9.1	8.1	4.8	9.3	5.8	10.8	2.9	9.3
dev 0,2	2.9	22.4	5.8	16.8	9.1	16.8	2.4	17.9	3.8	20.1	4.3	20.1

Tab.14 Chybovosť verifikácie v závislosti od šumu v % pre 3 úrovne parametru šumu dev, podľa (97) a pre všetky testované slová

Z Tab. 14 vidíme, že najvyšší stupeň zašumenia dev=0,2 sa prejavuje až trojnásobným zvýšením EXTRA chybovosti pre dané slovo, pre INTRA chybosti nepozorujeme taký veľký nárast. V priemere INTRA a EXTRA chybovosti sa pre nulový a nízky šum štatisticky nelíšia, pozri Obr.46.



Obr. 46 95% intervaly spoľahlivosti chybovosti verfikácie v % pre 3 úrovne zašumenia šum pre 1-slovné (vľavo) a 3-slovné (vpravo) spôsoby verifikácie

Na štatistické vyhodnotenia pozorovaných výsledkov sme použili štatistický program SPSS 8 a jeho procedúru General Linear Method (GLM) spolu s chybovými grafmi 95% intervalov spoľahlivosti pre relevantné veličiny, čo odpovedá úrovni významnosti rovnej 5%, *Field* (2009).

Výsledky celkovej úspešnosti vyhodnotené pomocou GLM – opakované merania, indikujú, že je signifikantný rozdiel ($F = 54,34$; $df = 2$; $sig = 0,0$), medzi chybovosťou 1-slovnej verifikácie zašumenej s úrovňami parametra dev 0, 0,1 a 0,2. V predošlých označeniach sme použili F ako testovaciu štatistiku, v tomto prípade Fisherove F, pomocou df označujeme stupne voľnosti relevantné pre špecifický test, a sig označuje dvojstrannú signifikanciu daného testu hypotéz. Podobné platí aj pre 3-slovný proces verifikácie s ($F = 128,5$; $df = 2$; $sig = 0,0$), pozri Obr. 46.

Z Obr. 46 vidíme, že nie je signifikantný rozdiel medzi šumom opísaným parametrom dev=0 a dev=0,1 v celkovej strednej chybovosti verifikácie, či pre 1-slovný spôsob alebo 3- slovný. Chybovosti pre dev=0,2 sú približne viac ako 2 krát vyššie aj pre 1-slovný spôsob alebo 3- slovný. Zároveň sú signifikantne vyššie v porovnaní s chybovosťami pre nižšie úrovne šumu.

6.5.3 Dimenzia vektorov čŕt a PNN diskriminácia

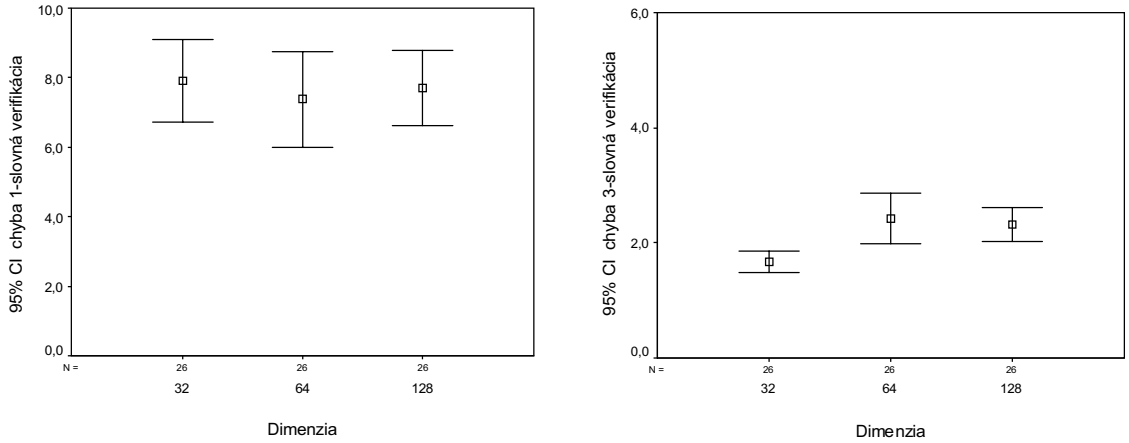
V treťom type experimentov sme skúmali závislosť našej verifikačnej metódy od dimenzionality vektorov čŕt, t.j. porovnávali sme 4, 8 a 16 osí na Nyquistovom grafe (Obr. 30, časť 4). Pretože sa v našej parametrizácii uvažovala iba vrchná časť grafu, použili sme nasledovné dimenzie vektora čŕt na patern slova: 32, 64 a 128. Predpokladáme, že úspešnosti sa nelíšia pre jednotlivé dimenzie.

Z Tab. 15 vidíme, že najmenší rozmer reprezentácie 32 sa prejavuje v strednom najmenšími hodnotami INTRA chybovosti pre dané slovo, pre EXTRA chybosti nepozorujeme taký veľký nárast s narastajúcou dimenziou. V priemere INTRA a EXTRA chybovosti sa pre všetky skúmané dimenzie štatisticky nelíšia.

	slovo1		**slovo2**		**slovo3**		**slovo4**		**slovo5**		**slovo6**	
	INTRA	EXTRA	INTRA	EXTRA	INTRA	EXTRA	INTRA	EXTRA	INTRA	EXTRA	INTRA	EXTRA
dim 32	3,4	9,5	10,6	8,5	10,6	7,8	4,3	9,7	6,3	10,5	4,8	8,7
dim 64	5,3	5,5	14,4	5,0	13,0	4,6	6,7	5,8	7,2	6,6	9,1	5,2
dim 128	3,8	9,6	10,1	8,8	9,1	8,1	4,8	9,3	5,8	10,8	2,9	9,3

Tab. 15 Úspešnosť procesu verifikácie v % pre 3 dimenzie a pre všetky testované slová

Použijúc pôvodné testovacie dáta v 1-slovnej konfigurácii vidíme, pomocou GLM - opakované merania, že nie je signifikantný rozdiel (F=1,916; df=2; sig=0,158) medzi testovanými dimenziami. Na druhej strane, vidíme, že je signifikantný rozdiel (F=7,544; df=2; sig=0,0) medzi vyšetrovanými dimenziami pre 3-slovné proces verifikácie, pozri Obr. 47.



Obr. 47 95% interval spoľahlivosti chybovosti procesu verfikácie v % pre 3 dimenzie vzorov slova 32, 64, 128 a pre 1-slovnú (vľavo) a 3-slovnú (vpravo) metódy verifikácie

Chybovosti pre 3 dimenzie vektorov nie sú signifikantne rozdielne, ako vidíme aj z grafu na OBr. 47, pre 1-slovné spôsoby verifikácie. Pre 3-slovné spôsoby verifikácie sa prekvapujúco ukázala chybovosť pre najmenšiu dimenziu najnižšia a zároveň signifikantne nižšia od chybovosti pre vyššie (64 a 128) dimenzie, ktoré sa ukázali ako štatisticky rovnaké z hľadiska chybovostí.

6.5.4 Robustnosť obecného modelu hovoriaceho a PNN diskriminácia

Ďalšou dôležitou črtou v našej metóde je tiež schopnosť zovšeobecnenia vzhľadom k počtu a výberu doplňujúcich hovoriacich pre EXTRA triedu. Hovoríme o vysokej zovšeobecniteľnosti ak iba malý podiel zo všetkých doplňujúcich hovoriacich môže vytvoriť robustnú reprezentáciu EXTRA triedy, túto schopnosť vyjadríme formálne pomocou pojmu vynechaní hovoriaci - OS (omitted speakers). V poslednom type experimentu vyšetríme práve túto vlastnosť, t.j. chybovosť v závislosti od počtu vynechaných hovoriacich. V týchto experimentoch sme vybrali dimenziu vektorov čŕt rovnú 32, parameter šumu dev je rovný 0. Potom pozorované INTRA a EXTRA úspešnosti sú dané v Tab. 16 pre počet 0 až po 20 vynechaných hovoriacich. Hodnoty v Tab. 16 sú priemery pre daný počet náhodne vybraných hovoriacich (s vrátením hovoriacich).

	#OS = 0		#OS = 5		#OS = 10		#OS = 15		#OS = 20	
	INTRA	EXTRA	INTRA	EXTRA	INTRA	EXTRA	INTRA	EXTRA	INTRA	EXTRA
1-slovo	6,7	5,1	6,9	7,1	6,6	9,6	6,0	12,6	5,2	18,3
3-slovo	1,5	1,9	2,1	3,5	2,5	3,8	2,0	4,1	1,4	7,2

Tab. 16 Celková INTRA a EXTRA úspešnosť v % pre proces verifikácie ako funkcia počtu vynechaných hovoriacich (#OS) pre 1-slovnú a 3-slovnú metódy verifikácie

A nakoniec, iba kvôli úplnosti pohľadu, výsledkov, sme predpokladali, že nie sú rozdiely v úspešnosti pre natívnych a iných hovoriacich, ani rozdiely v úspešnosti pre rodinne príbuzných hovoriacich. Naozaj, nezistili sme signifikantný efekt ($t = 3.784$; $df = 25$; $sig = 0.231$) pre natívnych hovoriacich (anglicky) v porovnaní s hovoriacimi s iným jazykovým pôvodom (čínsky, japonsky, slovensky, grécky), pričom sme použili nevyvážený t-test pre nezávislé vzorky v SPSS a t je testovacia

štatistika, v tomto prípade Studentovo t, pomocou df označujeme počet stupňov voľnosti pre daný špecifický test, a sig označuje dvoj-strannú signifikanciu daného testu hypotéz. Podobne, nezistili sme efekt, ktorý sa dotýka rodinných vzťahov hovoriacich, ktorý by sme predpokladali kvôli podobnosti hlasov členov z rovnakej rodiny (t=3,025; df=7; sig=0,114). Pre tento účel sme použili 4 páry rodinných príslušníkov v našej databáze,*Chudý, Kačúr* (2012).

7 Diskusia

V prvej časti prediskutujeme výsledky rozpoznávania reči pomocou rôznych prístupov založených na topologických invariantoch, sústredime sa hlavne na experimenty s modelovaním súvislej reči. V druhej časti rozoberieme experimenty, ktoré sme vykonali pre otestovanie schopností prístupu založeného na topologických invariantoch v oblasti verifikácie hovoriaceho. A v tretej časti, venovanej konceptuálnym otázkam, prediskutujeme percepciu reči u ľudí z hľadiska symetrií systému foném daného jazyka.

7.1 Rozpoznávanie reči

V tejto práci sme prezentovali matematický a fyzikálny pohľad na problém rozpoznávania reči pomocou topologických invariantov. Použijúc znalosti a fakty o ľudskom sluchovom systéme sme boli schopní definovať niektoré invarianty, ktoré sú matematicky dostatočne abstraktné, pričom ale stále odpovedajú základným princípom sluchového systému. Testovali sme tieto rečové črty na našej vlastnej databáze izolovaných slov pomocou modelu PNN a aj na profesionálnej databáze pomocou rôznych HMM modelov. Navyše, k pôvodnému prístupu sme navrhli a otestovali ďalšie obecné techniky spracovania signálu tak, aby sme zlepšili úspešnosť pôvodných čŕt. V nasledujúcich riadkoch zosumarizujeme hlavné zistenia:

Všetky učenia a testy sa vykonali na profesionálnej rečovej databáze MOBILDAT-SK, kde sme dosiahli nasledovné úspešnosti 97.7%, 98.7% a 98.9% pre testy na reťazce číslic, aplikačné slová a izolované číslice, resp.. V ďalšom prediskutujeme cestu akou sme sa dostali k týmto výsledkom, *Kačúr, Chudý* (2012).

Pôvodné črty, reprezentujúce preseknutia rôznych osí, nie sú vhodné v spojení s modelovaním pomocou HMM založenom na gaussovskej hustote pravdepodobnostnej funkcie. Niektoré črty sú silne korelované a niektoré vedú k nízkej variabilite vnútri tried. Aby sme eliminovali tieto nežiadúce vlastnosti, uchovajúc diskriminačný potenciál použili sme jednoduchú LDA transformáciu, čo sa ukázalo ako efektívne. Navyše k LDA, sme použili a otestovali jedinú dekorelačnú transformáciu, aby sme potlačili korelácie vnútri tried. Pri použití diagonálnych kovariančných matíc tento krok viedol k relatívnemu zlepšeniu WER v rozsahu od 3.7% až dokonca do 40%.

Topologické invarianty sú úplne nezávislé od informácie o energii signálu, ale z analýzy reči aj percepcie vieme, že evolúcia energie v čase je veľmi dúležitá, t.j. jej relatívne zmeny. Takto pre každú frekvenčnú priepusť sme pridali nový parameter, popisujúci relatívnu energiu. Avšak, po aplikovaní LDA transformácie konečná dĺžka vektora čŕt bola rovnaká. Normalizovaná energia priniesla nasledovné relatívne zlepšenie WER: 92%, 84% a 91% pre testovanie číslic, aplikačných slov a reťazcov číslic, čo preukazuje jej opodstanenosť.

Rotácia signálovej roviny sa ukázala ako ďalšou veľmi výhodnou transformáciou, špeciálne pre vyššie počty frekvenčných priepustí, t.j. úzke priepuste, kde môže eliminovať vzájomný fázový posuv medzi dvomi kvázi harmonickými signálmi. Avšak, efektívnosť takejto transformácie môže klesnúť s narastajúcou šírkou frekvenčnej priepuste, kde môže byť prítomných veľa harmonických frekvencií a v tomto prípade efekt rotácie nemusí tak jasný z interpretačného hľadiska. Pre prípad 16 frekvenčných priepustí sme zaznamenali 7.7% zlepšenie (v strednom), zatiaľčo pre prípad 8 frekvenčných priepustí naopak 6.5% zhoršenie. Na druhej strane, ak sa uvažovalo iba správanie z hľadiska najlepšie sa prejavujúcich modelov pre testovanie aplikačných slov, v oboch prípadoch sa dosiahlo zlepšenie 35%.

Čo sa týka testovania počtu frekvenčných priepustí, testovali sa dva extrémne prípady, 16 kritických frekvenčných priepustí v rozsahu od 200 do 4000Hz a 8 frekvenčných priepustí rovnomerne rozdelených pozdĺž Mel škály, ktoré pokrývali rovnaký frekvenčný rozsah. Pre 16

frekvenčných priepustí sa testovali 2 a 3 preseknutia, pričom nižšie počty dávali lepšie výsledky. Pre 8 frekvenčných priepustí sme testovali situácie s 4, 6 a 8 preseknutiami; v tomto prípade rozdiely neboli tak podstatné.

Avšak, všetky scenáre s 8 frekvenčnými priepusťami boli lepšie oproti 16 frekvenčným priepustiam v strednom o skoro 20%. Toto môže viesť k určitému úsudku, že, aspoň v prípade 8 frekvenčných priepustí, je relevantnejšia informácia uložená v topologických invariantoch než v informácii o relatívnej energii pre každú z frekvenčných priepustí. Toto sa môže vidieť z pomeru počtu elementov nesúcich informáciu o energii k celkovej pôvodnej dĺžke vektora čŕt (pred aplikovaním LDA); pre 16 frekvenčných priepustí a 2 preseknutia to je približne 40% zatiaľčo pre 8 frekvenčných priepustí a 8 preseknutí je to iba11.1%. Takto vektor čŕt obsahujúci iba 11.1% informácie o energii si vedie v strednom o 20% lepšie ako vektor čŕt s 40% elementov so vzťahom k energii.

Dosiahnuté výsledky sme porovnali s, terajšími, najpopulárnejšími črtami MFC. Použili sme štandardné nastavenia: 12 statických MFC, plus normalizovanú energiu, 13 delta, a 13 akceleračných koeficientov. Rovnaké nastavenia sme použili pre topologické invarianty. V strednom, berúc do úvahy všetky modely a scenáre testovania, MFC produkuje nižšie WER o 22%, avšak marginálne scenáre testovania zaznamenali zlepšenie topologických invariantov nad MFC o: 14%, -92%, 25%, pre aplikačné slová, reťazce číslic a izolované čislice resp.. Ako vidíme topologické invarianty zlyhali v porovnaní s MFC v scenári reťazcoch číslic, v ktorom sa vyžaduje dobré modelovanie reči ako aj dobré modely pozadia a modely nie rečových javov.

Všetky tieto fakty môžu viesť k záveru, že topologické invarianty nemajú postačujúcu schopnosť modelovať nerečové javy. Avšak, to je vecou budúcich experimentov, spolu so zahrnutím časových topologických invariantov a takto sa tomuto v tejto práci nebudeme ďalej venovať.

Podobne k vylepšeniu schopnosti vysporiadať sa aj s nerečovými javmi môže dôjsť pomocou vyšetrenia frekvenčných topologických invariantov. Túto myšlienku sme v základnej forme rozvinuli v dodatku A4. Táto myšlienka sa dá použiť samostatne alebo ako jeden z posledných krokov v štandartnom PLP prístupe, kde vyhladené spektrum v PLP sa transformuje na topologické invarianty, podobne ako je uvedené v dodatku A4. Všetky tieto prístupy a myšlienky by mohli viesť k použiteľným riešeniam pre skúmania v budúcnosti.

7.2 Rozpoznávanie hovoriaceho

V prvom rade, sme boli schopní aspoň približne replikovať úspešnosť verifikácie hovoriaceho v laboratórnych podmienkach *Reynold* (2002). Po druhé, ukázali sme, že úspešnosť metódy TIM spolu s PNN v reálnych podmienkach sa nezníži podstatným spôsobom až do hodnoty parametra šumu dev=0,1 okolo 20%, pozri 6.5.2, ale je to tak iba pre gaussovský šum, nie pre abruptný šum, podobný reálnejším podmienkam. Metódy založené napr. na DTW a topologických invariantoch sa vyporiadavajú so šumom podobným spôsobom, aj keď nie tak dobre ako euklidovská a PNN metódy.

V niektorých prípadoch je percento intraspeaker zlyhaní vyššie než percento extraspeaker zlyhaní (napr. PNN pre dev =0). Rovnováha medzi intraspeaker a extraspeaker zlyhaniami sa môže previesť pomocou voliteľného parametra prahu, (94), časť 6.4. Takže v každom prípade je možné nájsť taký prah, že percento intraspeaker zlyhaní bude menšie než pre extraspeaker zlyhania. Po tretie, experimenty napovedajú, že dimenzie vektorov čŕt vykazujú efekty iba pre 3-slovné procesy verifikácie, pričom efektívne vedú k poklesu úspešnosti s narastajúcou dimenziou vektorov čŕt. Pre 1-slovnú metódu verifikácie sme nepozorovali efekt, Obr. 47, pričom približne informácia v štyri krát väčšom vektore čŕt je porovnateľná 3 slovami v 3-slovnej verifikácii.

Zistili sme, že úspešnosť obecne, vyjadrená v %, okolo 96 % pre jednoslovnú verifikáciu a 98 % pre trojslovnú verifikáciu, je obecne porovnateľná so systémami založenými na MFC alebo

PLP črtách v úlohach, ktoré majú podobné nastavenie(počet hovoriacich a typ prehovorenej reči), *Chudý, Kačúr* (2012).

Ako sme sa už zmienili v predošlej časti problém zovšeobecnenia je dôležitým pre našu metódu verifikácie hovoriaceho. Má do činenia priamo s otázkou koľko hovoriacich zahrnieme do triedy doplnkových hovoriacich. Takže, do druhej triedy doplnkových hovoriacich sme zahrnuli iba časť prototypov z celkového počtu. Reprezentujeme to formálne pomocou počtu vynechaných hovoriacich. Naše skúmania viedli k záveru, že efekt vynechaných hovoriacich sa nemusí uvažovať až po počet rovný polovici hovoriacich, pozri Tab. 16.

A nakoniec, všetky tieto efekty a výsledky boli založené na presnosti VAD (voice activity detection). Použili sme nami vyvinutú metódu využívajúcu energiu, prechody nulou, pozri dodatok A1. Tento samo sa adjustujúci algoritmus bol optimalizovamý pre telefonickú linku. Dá sa odhadnúť (porovnaním s presnými hraniciami slov), že VAD prispieva približne 2-5% do celkovej chybovosti verifikačného systému. Tento nedostatok sa dá prekonať pomocou textovo nezávislých prístupov.

Náš systém je založený na izolovaných slovách, preto by bolo vhodné sa v budúcnosti venovať aj verifikácii, obecne rozpoznávaniu hovoriacich z hľadiska nezávislosti na slovách, aby sme vyjasnili úlohu akú zohrávajú hláskam podobné jednotky pri verifikácii hlasu hovoriaceho. Podobne je v budúcnosti nutné vyšetriť závislosť nášho systému na reálnejšom šume, abruptného typu z hľadiska zhlukovania sa primárnych vektorov čŕt.

Z celkového hľadiska, uvedené experimenty poskytujú evidenciu podporujúcu to, že nie sú efekty rodinne podobných hlasov pri verifikácii hovoriacich pomocou TIM. Navyše, sme ukázali, že vzory úspešnosti natívnych hovoriacich sú celkovo podobné vzorom úspešnosti nie natívnych hovoriacich. Nakoniec, sme poskytli evidenciu možnosti uprednostňovania minimálnej dimenzie vektorov čŕt, prinajmenšom pre 3-slovné rozhodovania.

K vylepšeniu schopnosti vysporiadať sa aj s niektorými javmi nerečovej povahy môžeme použiť prístup pomocou frekvenčných topologických invariantov. Túto myšlienku sme v základnej forme rozvinuli v dodatku A4.

7.3 Konceptuálne otázky percepcie reči

Na základe našej diskusie v predošlých častiach venovanej obecne percepcii, 3.2, môžeme usúdiť, že náš obraz je v celkovom súhlase s viacerými základnými paradigmami, ktoré podľa nášho názoru hrajú veľmi dôležitú úlohu vo všetkých úvahách za naším modelom:

P_0 - myslíme si, že jav percepcie musí byť konzistentný s tromi postulátmi Kantorovej teórie informácie, *Atmanspacher* (1988)

1. Zachovanie informácie
2. Komunikovateľnosť informácie
3. Konečná dosažiteľnosť informácie

Tieto postuláty vyplývajú z unitárnosti časového vývoja vlnovej funkcie v kvantovej mechanike alebo elementárne z nasledovného vzťahu neurčitosti

$$\Delta I \gg \Delta E \Delta t / h$$

kde ΔI je hodnota informačného prenosu, Δt je čas potrebný na prenos energie ΔE a h je Planckova konštanta, minimum akcie alebo elementárna akcia, pozri *Atmanspacher* (1988). Vidíme, že prenos tejto elementárnej akcie vedie k prenosu informácie ΔI.

P_1 - paradigma *Landauera* (1961): „*Informácia je fyzikálna, na vytvorenie, transformovanie, prenesenie 1 bitu informácie musíme vynaložiť nenulovú energiu*".

P_2 - paradigma *Chomského* (1968): „*Reč je vrodená biologická schopnosť ľudskej mysle*"

P_3 - fundamentálna paradigma porovnávacej jazykovedy: „*Všetky prirodzené jazyky majú rovnakú popisnú schopnosť, potenciu opísať vonkajší, vnútorný svet aj sami seba*"

Teraz stručne prediskutujeme posledné dve paradigmy. Druhá paradigma neznamená, že sociálne, kultúrne a historické pozadie nemá vplyv na generovanie celej jazykovej štruktúry pre daný jazyk. Táto paradigma iba predpokladá, že na základnej úrovni štruktúra jazyka je predurčená biologickou štruktúrou. Existuje veľa faktov a experimentálnych výsledkov, ktoré ilustrujú P_2, napríklad schopnosť malých detí manipulovať so symbolmi alebo rozpoznávať disštinktívne príznaky, črty *Reddy ed.* (1975), *Anderson,* (1983), *Tatham* (1984), *Chomsky* (1968). Tretia paradigma sa zaoberá schopnosťou popisu a nevzťahuje sa k aktuálnym, reálnym rozdielom a výhodám konkrétneho jazyka vzhľadom k inému jazyku. Napríklad, rozdiel v pojmovej štruktúre času v angličtine a v jazyku Hopi, alebo rozdiel v geometrických jazykových pojmoch spoločnosti žijúcej na rovine verzus vrchovine nie je relevantný z hľadiska P_3, ktorá reflektuje potenciál pre popis hocičoho na fundamentálnej, základnej úrovni, *Segall* (1986).

Máme silné dôvody domnievať sa, že predošlé úvahy aj s frekvenčne selektívnymi extrapoláciami nie sú podstatné pre rozlíšenie percepcie papagája a ľudskej percepcie. Predpokladáme, že podstatný rozdiel leží v určitej štruktúre (a odpovedajúcich typoch a mechanizmoch spracovania) nervového systému. Táto štruktúra sa navonok prejavuje ako takzvaná vnútorná symetria fonémického systému ľudského jazyka. Symetrie fonémického systému nie sú nič nezvyčajné. Používane ich každodenne, ale neuvedomujeme si tento fakt príliš silne. V tejto časti opíseme symetrie fonémického systému konceptuálnym spôsobom bez rigoróznej a presnej matematickej mašinérie. Tento spôsob je vhodný pre jeho schopnosť analyzovať javy a princípy, ktoré sú vo svojej podstate rozmazané a redundantné a pravdepodobnostné. Viac technický pohľad je uvedený v dodatku A3.

Symetrie foném sa prejavujú vo vlastnostiach ako sú podobnosť, analógia, v zhlukovaní alebo usporiadavaní niektorých foném do rôznych hierarchických štruktúr a spájaní niektorých foném s inými podľa určitých pravidiel. Fundamentálnym, základným prejavom tejto symetrie je takzvaná klasifikačná schéma daného fonémického systému jazyka.

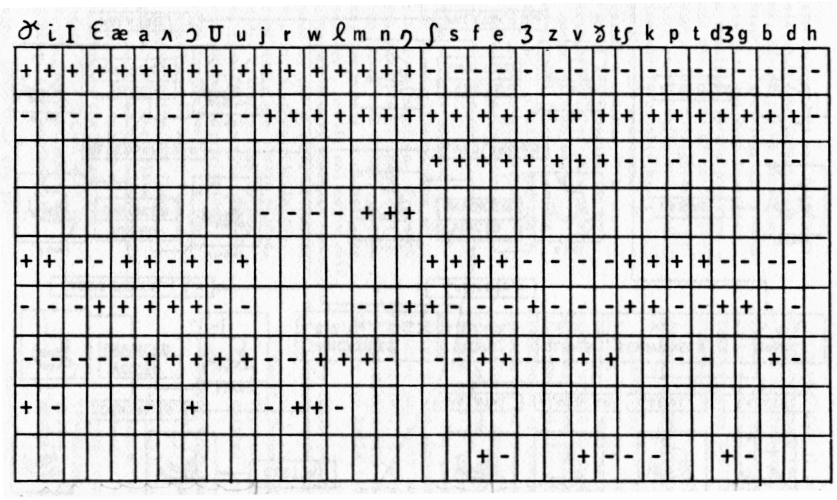
Na tomto mieste sa stručne zmienime o tom čo vlastne fonéma v „skutočnosti" je. V teórii, napríklad fonológii je fonéma definovaná ako najmenšia jednotka fonológie. Takáto interpretácia pojmu fonémy vznikla z uvedomenia si toho faktu, že presná fonetická realizácia konkrétneho zvuku reči nie je tak dôležitá ako jej funkcia v rámci zvukového systému konkrétneho jazyka. Na Obr. 48 uvádzame takýto systém pre anglický jazyk, prebraný z *Newell, Barnett* (1973). Fonetické varianty konkrétnej fonémy sa nazývajú alofóny.

Fonéma sa potom chápe ako zovšeobecnená forma konkrétneho zvuku so zanedbaním všetkých nepodstatných (dialektických, ideolektických, redundatných) variánt konkrétnych zvukov v procese abstrakcie. Tento proces sa dá napríklad vyjadriť fonémickým rozdielom medzi pármi slov, ktoré sa líšia iba v jednej foném, napríklad, *pes - les.* Každý jazyk má svoje vlastné zoskupenie foném, svoju vlastnú fonémickú štruktúru. V praxi jazykovedci používajú napríklad takzvané dištinktívne príznaky na klasifikáciu foném do konkrétnych tried, napríklad na Obr.49 je uvedený systém pre anglický jazyk z projektu DARPA (1973). Existuje veľa systémov dištinktívnych príznakov.

Pre naše účely je vhodný systém *Chomského* a *Halla*, (1968), v ktorom sú tieto príznaky umiestnené do kategórií, tried ako sú:

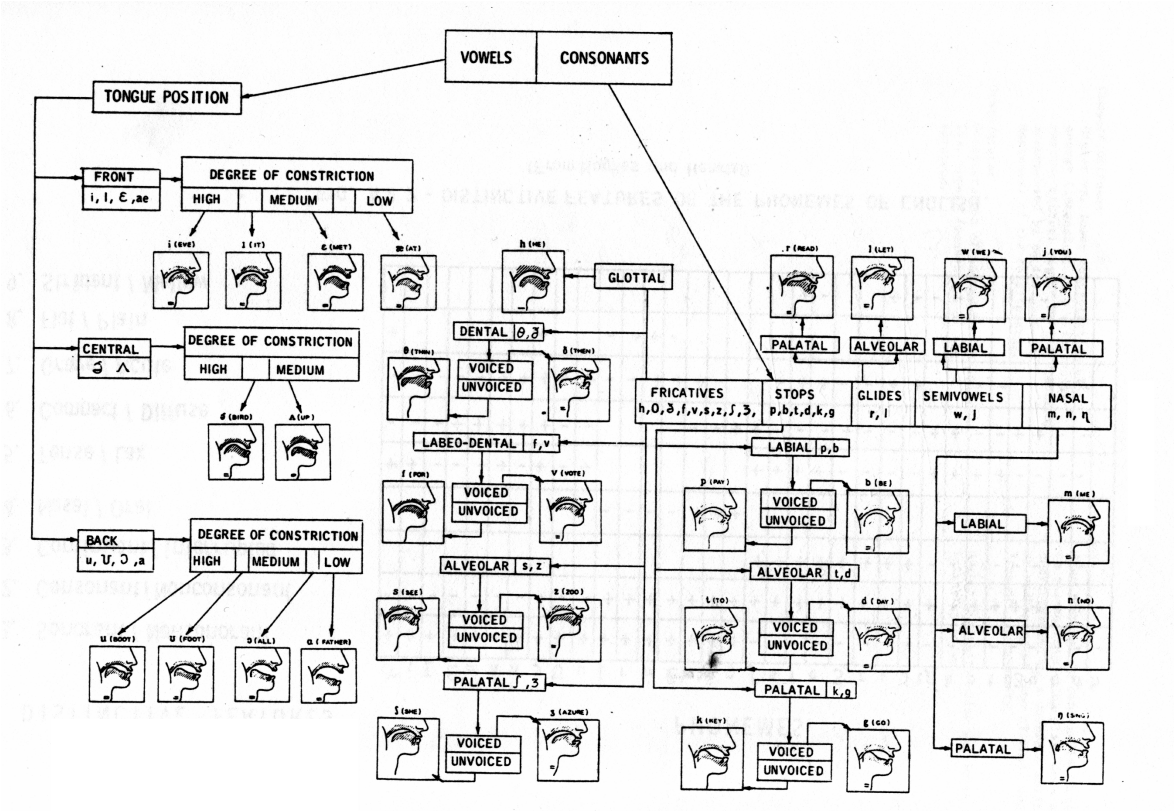
1. Hlavné príznaky - v tejto triede sú príznaky vokálny - nevokálny, sonórny - nesonórny, nosný - ústny, atď.
2. Dutinové príznaky - tieto príznaky reflektujú také črty ako miesto artikulácie a tvar ústnej dutiny, pozri Obr. 48.
3. Spôsob artikulácie - sa vzťahujú k príznakom spojitosti - nespojitosti.
4. Zdrojové príznaky - sa vzťahujú k príznakom ako je znelosť- neznelosť, atď.
5. Prozódické príznaky - a vzťahujú k príznakom ako je dôraz, základný tón, atď

Poslednú triedu príznakov neuvažujeme apriórne ako vnútorné symetrie, na rozdiel od predošlých sú iba transformáciami reči, vzhľadom ku ktorým mechanizmus rozpoznávania je invariantný.Samozrejme pre niektoré jazyky môžu zohrávať dištinktívnu funkciu.



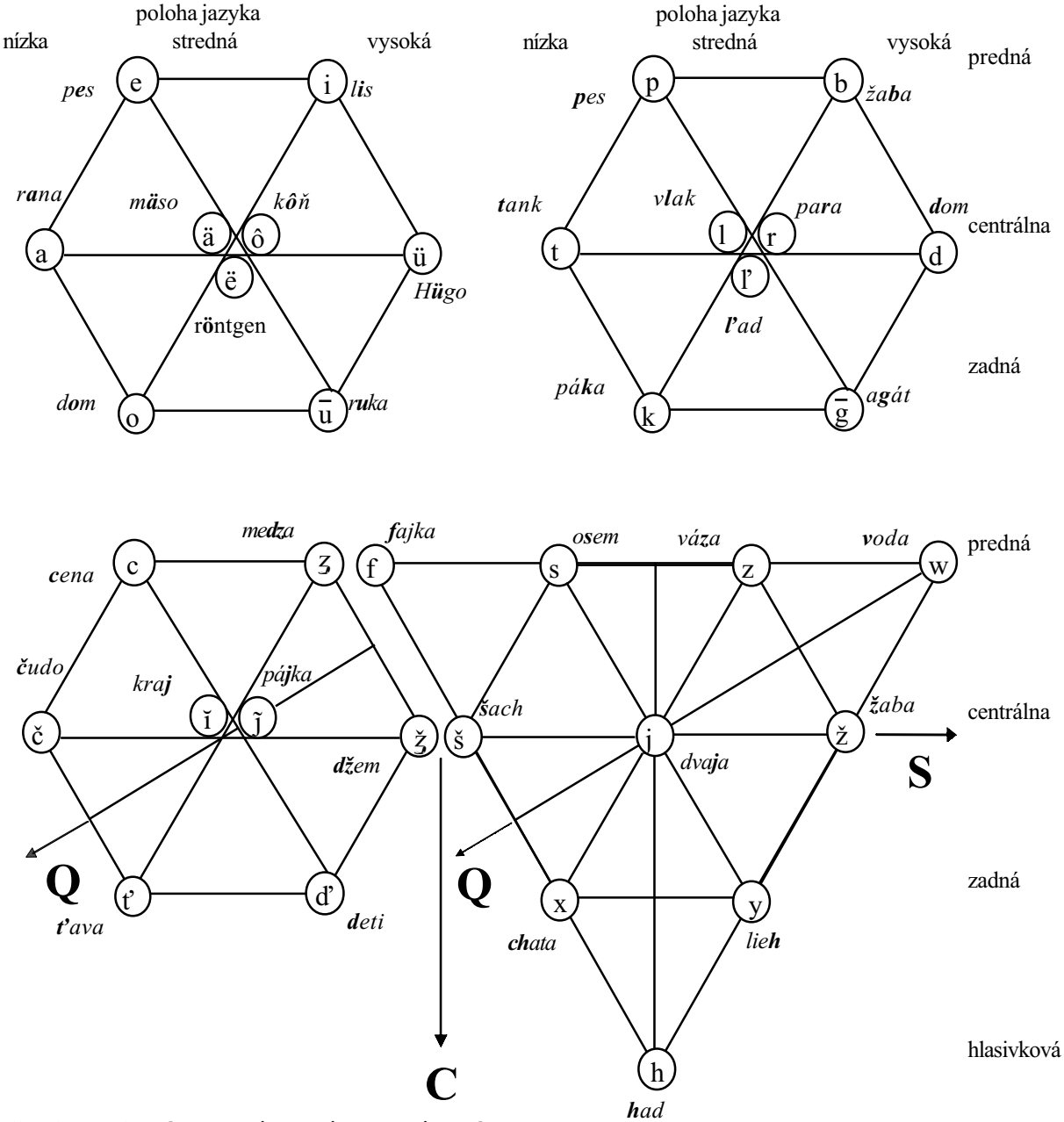
Legenda: podľa riadkov sú označené dištinktívne príznaky
1. znelé / neznelé
2.spoluhláskové / nespoluhláskové
3.spojité / nespojité (abruptné)
4.nosové / ústne
5. napäté / voľné
6. kompaktné / difúzne
7. hlboké, tupé / vysoké, ostré
8. rovné / hladké
9. škrípavé / lahodné

Obr. 48 Dištinktívne príznaky angličtiny podľa projektu DARPA

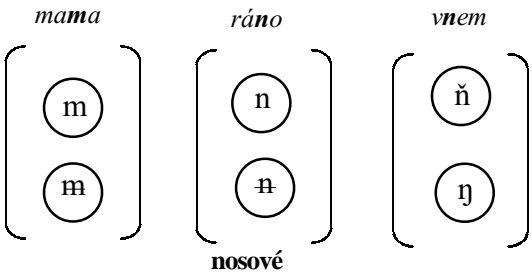


Obr. 49 Dištinktívne príznaky angličtiny podľa miesta artikulácie, projekt DARPA

Na Obr.50 uvádzame klasifikačnú schému slovenského fonémického systému, podobná schéma pre anglický jazyk je uvedená v dodatku A6. Táto klasifikačná schéma sa dá považovať tiež za heuristický princíp. Vidíme, že všetky fonémy sa klasifikujú do 5 multipletov podľa určitých čŕt, príznakov. Prvý multiplet je nonet (singlet + oktet) samohláskam - podobným fonémam. Druhý je tiež nonet (singlet + oktet) semi samohláskam podobným a stop fonémam podobným. Tretí multiplet je oktet stop a afrikatívam podobným fonémam. Štvrtý multiplet je dekaplet frikatívnych foném. Posledný multiplet je súbor nazálnych (nosových) spoluhláskových foném.



C- avity (dutinové) – predná , centrálna, zadná, glotálna
S-onorant/ nonsonorant (znelé/neznelé)
Q ? (aký je význam)



Obr. 50. Fonémický systém slovenčiny. Pre nosové fonémyV slovenčine sú fonémy iba v horných riadkoch, v dolných sú alofóny-hlásky- nemajú sémantický význam

Klasifikácia do multipletov je, podľa Chomského schémy, založená na 1.príznakoch - hlavných a 2. príznakoch - spôsoboch artikulácie. Subklasifikácia do daných multipletov je založená na 2. príznakoch - dutinových - horizontálne línie a vertikálne línie na 4. príznakoch - zdrojových a na dutinových.

Ako príklad, najskôr uvažujme nie nosové fonémy. V tomto prípade máme 4 horizontálne línie, ku ktorým priradíme nasledujúce dutinové príznaky:

horná línia - (bilabiály + labiodentály + prealveoláry)
predná stredná línia - (alveoláry + postalveoláry + prepalatály)
centrálna nižšia línia - (palatály + postpalatály + preveláry)
zadná najnižšia línia - (veláry + postveláry + glotály).

Podobne ako pre štandardnú klasifikáciu, ktorá je uvedená v predošlom v zátvorkách, aj naša klasifikácia je vecou konvencie, dohody. Ale myslíme si, že naša klasifikácia - predné, centrálne, zadné a glotálne - má veľmi zaujímavé vlastnosti. Vertikálne línie v prvom nonete opisujú dutinové príznaky, ktoré sú známe v fonológii ako vysoká, stredná a nízka podľa polohy jazyka. Vertikálne línie v zostávajúcich multipletoch opisujú znelé - neznelé vzťahy, napríklad (p - b, s - z). Fonémy v strede daného multipletu nemajú párové fonémy podľa príznaku znelosť - neznelosť.

Posledný súbor nazálnych foném (nosoviek) obsahuje tri fonémy a pre každú z nich spoločníka, alofónu, alebo fonému, ktorá nemá sémantickú oporu. Znovu, klasifikácia je podľa dutinových príznakov - predné, centrálne, zadné.

Porovnanie klasifikačnej schémy foném ako je uvedené na Obr.50 s klasifikáciou elementárnych častíc - hadrónov podľa SU(3), dodatok A5, približne vyjadrené (presnejšie SU(2)xU(1)) grupy a leptónov, *DeWitt, de Witt* (1964), *Bhagavantam, Venkatarayudu* (1951), *Ryder* (1985) nám hovorí, že môže existovať určitá analógia medzi týmito dvomi subjektami. Jeden z možných matematických popisov tejto analógie z hľadiska percepcie a neurónových sietí je prediskutovaný v dodatku A3. Vedie k určitému grupovo-teoretickému popisu nášho systému foném pomocou spojitých Lieových grúp SU(N) a ich ireducibilných reprezentácií.

7.3.1 Aplikovanie pojmov symetrie v percepcie reči

V dodatku A3 z prediskutovaných grupových vlastností transfomácií L, generátorov grupy, vyplýva aj nasledujúca interpretácia, že práve ony by mohli byť zodpovedné za kompresiu informácie v takom zmysle, že rôzne „zašumené" obrazy, vzory sa redukujú na prototypy - štandardné vzory. Aby sme to ozrejmili bližšie uvažujme, spolu s *Kammen, Yuille* (1988), *Mas, Ramos* (1989) už spomínaný problém identifikácie našich zvukových vzorov pri nejakých infinitezimálnych variáciach zvukového vzoru X, $\Delta X = (\delta x_1, \delta x_2, ..., \delta x_{2n})$. Rozvinieme $\Psi(X)$ do radu podľa bázických vektorov

$$\Psi(X) = \Psi_i(X)\, \zeta^i(X), \qquad (97)$$

kde predpokladáme sumačné pravidlo, báza je konečná. Prevedieme kroky podobné krokom v dodatku A3, len trochu iným spôsobom. Zavedieme transformácie, ktoré transformujú bázové vektory do seba, podľa vzťahu

$$\zeta'_i(X) = T^j_i[X, p_1, p_2, ...p_k]\, \zeta^j(X),$$

kde predpokladáme, že takto definované transformácie sú v dodatku A3 definované Lie-ho transformácie, a že tvoria Lie-ho grupu, s parametrami $p_1, p_2, ..., p_k$. Potom pre infinitezimálne transformácie platí

$$T^j_i [X, \varepsilon_1, \varepsilon_2, ...\varepsilon_k] = 1 + \varepsilon^j (X) L_j, \qquad (98)$$

kde L sú generátory danej Lie-ho grupy, ktoré už nie sú závislé na X a $\varepsilon^j (X)$ sú infinitezimálne parametre, závislé na X. Pre generátory platí už spomínaný komutačný vzťah (.) z dodatku A3. Teraz sa vráťme k našej variácii zvukového vzoru. Predpokladáme na základe napríklad schémy klasifikácie foném, Obr. 50, čo pokladáme za experimentálny - empirický fakt, že bázové vektory $\zeta_j (X + \Delta X)$ sa, pri infinitezimálnych transformáciach, transformujú ako

$$\zeta_j (X + \Delta X) = (1 + \varepsilon^n \delta X L_n)^i_j \zeta_i (X),$$

kde sme predpokladali, že malý parameter $\varepsilon^j (X)$ je úmerný variácii δX vzoru X. Dosadením predošlého vzťahu do rovnice (97) dostaneme

$$\Psi(X + \Delta X) = \Psi^j(X + \Delta X)\ \zeta_j (X + \Delta X),$$

kde opäť používame sumačné pravidlo. Dosadením dostaneme

$$\Psi(X + \Delta X) = \Psi(X) + [\partial_m \Psi + \varepsilon^n{}_m (X) L^i_{nj} \Psi^j] \delta X^m \zeta_i (X). \qquad (99)$$

Druhý člen v predošlom vzťahu vyjadruje odchýlku obrazu $\Psi(X + \Delta X)$ od $\Psi(X)$ pri transformácii súradníc podľa (98), tak aby sme vyhoveli pozorovanej symetrii fonémického systému. Rozdiel predošlých vzorov, pôvodného a variovaného je nový informačný obsah (v mozgu). Obecne povediac parametre transformácie môžu závisieť na X, inými slovami $\varepsilon^n{}_m (X)$ je funkciou samotného vzoru, transformácia je závislá na rečovom vzore, ktorý sa uvažuje. Odpovedá to analogickej situácii v identifikácii rečových ale aj optických obrazov, ktorá sa všeobecne popisuje ako ilúzia, *Ditzinger, Haken,* (1989), *Wechsler* (1990). Inými slovami dva odlišné vzory v konfiguračnom priestore, sa pri rovnakej transformácii zmenia odlišným spôsobom, na obrazy v mozgu. Opäť javy takéhoto druhu sa bežne pozorujú aj v zvukovej, aj v optickej oblasti - obecne pri všetkých senzorických kanáloch. A podobne ako sme postupovali v úvodnej časti dodatku A3, zavedieme kovariantnú deriváciu vzťahom

$$D_\alpha \Psi^b(x) = \partial_\alpha \Psi^b(x) + \varepsilon^n{}_\alpha L^b{}_{na} \Psi^a(x) = \partial_\alpha \Psi^b(x) + igA^b{}_{\alpha a}(x) \Psi^a(x),$$

kde sme zároveň zaviedli $igA^\alpha(x)$, väzbový koeficient. Takto dostaneme

$$\Psi(X + \Delta X) = \Psi(X) + D_\alpha \Psi^b(x) \zeta_b \delta X^\alpha.$$

Ak teraz chceme aby sa splnila podmienka identifikácie, potom norma vektora $\Psi(X + \Delta X) - \Psi(X)$ musí byť menšia ako nejaká kritická vzdialenosť identifikácie (nad ňou rozdielne, pod ňou podobné). Postačujúcou podmienkou, aby bolo toto splnené je

$$D_\alpha \Psi^b(x) = 0.$$

Čo priamo vedie k nulovosti odpovedajúceho tenzoru krivosti, ktorý v tomto prípade je

$$F^b{}_{\alpha\beta} = \partial_\alpha A_\beta{}^b(x) - \partial_\beta A_\alpha{}^b(x) + [A_\alpha, A_\beta]^b$$

kde

$$[A_\alpha, A_\beta] = A_\beta A_\alpha - A_\alpha A_\beta$$

je komutačný vzťah. V prípade ak máme splnené $[A_\alpha, A_\beta] = 0$, hovoríme o abelovských transformáciach (grupách), a ak ešte generátory infinitezimálnych transformácií nezávisia od X hovoríme o globálnej transformácii, symetrii. Je preukázané, že niektoré transformácie symetrie sú inherentné v mozgu, napríklad priestorová translačná symetria, zrkadlová inverzia, atď. iné sa učíme a následne sa tvoria v mozgu, *Giles, Maxwell* (1987), *Dodwell* (1983), *Pessa* (1988), *Mazzola, Wieser, Brunner, Muzzulini* (1989), *Földiák* (1991), *Fukushima* (1980), *Piaget, Inhelderová* (2007). V časti (3.2) alebo (4.1.1) spomínané vonkajšie symetrie môžu byť takými príkladmi. Inými slovami, ak je splnená podmienka nulovosti tenzoru krivosti, identifikácia vzorov je zabezpečená vhodnými transformáciami súradníc, ktoré vytvárajú Lieho grupu. Naopak, ak nie je splnená táto podmienka, potom sa vygenerujú v mozgu nejaké polia obrazov a im odpovedajúce vzory majú, nesú iný informačný obsah, *Guez, Protopopescu, Barhen* (1988).

Na chvíľku sa zastavme a trochu voľnejšie prediskutujme, čo vlastne hovoria spomínané symetrie - vonkajšie aj vnútorné v percepcii chápanej ako zdieľaná komunikácia- pozorovatelia - systém - na klasifikáciu a identifikáciu vzorov obecného druhu, vzorov ako takých, *Carpenter, Grossberg* (1987), *Shalkoff* (1992), *Li* (1991).

Každý deň sme schopní navzájom komunikovať o rôznych veciach, a medzi nimi aj o rôznych typoch pohybu a rozpoznávaní vzorov, obecne o procese ich kvalifikácie. Z fyziky poznáme veličinu, nazýva sa účinok, pomocou ktorej môžeme vyjadriť najsymetrickejšiu mieru pohybu, zmeny. Padnutie kameňa je zmena, jednotlivý akt percepcie je zmena, zmena hladiny hormónov v mozgu je zmena, akt myslenia jednej myšlienky je zmena (v zmysle ako sme o psychológii diskutovali v časti venovanej obecnej percepcii). Je rovnakým bez ohľadu či sa uskutoční tu alebo tam, v jednom smere alebo druhom, dnes alebo zajtra. Naozaj, (napríklad špeciálne Galileiho) účinok je číslo, ktorého hodnota je rovnaká pre každého pozorovateľa v kľude, nezávisle na jeho orientácii alebo okamžiku kedy sa robí dané pozorovanie, *Moller* (1972).

V prípade statických symetrií napríklad, pre ornament, nám symetria dovoľuje dedukovať zoznam multipletov, alebo reprezentácií, ktoré môžu byť jeho stavebnými blokmi a hovoríme, že ornament sa skladá z daných pod-ornamentov. Tento prístup je možný takisto pre akýkoľvek druh pohybu. Klasifikácia vzorov na ornamente sa dá previesť do singletov, dubletov, atď. z rôznych možných pozorovacích hľadísk.

Pre pohybujúci sa systém, stavebné bloky, odpovedajúce blokom v ornamente, sa nazývajú pozorovateľné. Pretože pozorujeme, že príroda je symetrická vzhľadom k mnohým rôznym zmenám hľadísk, môžeme podľa toho klasifikovať všetky možné pozorovateľné. Aby sme to urobili, potrebujeme len uvažovať zoznam všetkých transformácií hľadísk a dedukovať z neho zoznam všetkých ich reprezentácií.

Každodenná skúsenosť ukazuje, že svet sa nemení napríklad pri zmenách polohy, orientácie a okamihu pozorovania. Hovoríme tiež o invariancii voči priestorovej translácii, rotácii a časovej translácii. Tieto transformácie sú odlišné od transformácií na ornamente z dvoch pohľadov: sú spojité a sú neohraničené. Ako výsledok, ich reprezentácie budú obecne pojmami, ktoré sa môžu meniť spojito a bez hraníc; budú to veličiny alebo veľkosti.

Vráťme sa k opisu pohybu. Vo fyzikálnom systéme musíme vždy rozlišovať medzi symetriou celého Lagranžiánu, pozri dodatok A3 – odpovedajúcou symetrii úplného vzoru – a reprezentácii pozorovateľných – odpovedajúcou (nie moc presne) symetrii podornamentov, bočných pásov v ornamente a podobne, pozri obrázok na predošlej strane. Pretože účinok musí byť skalár a pretože všetky pozorovateľné musia byť tenzory, Lagranžiány sú súčtom a súčinom tenzorov iba v jedinej kombinácii, a síce v tej, ktorá tvorí skalár. Lagranžiány obsahujú iba skalárne súčiny alebo ich zovšeobecnenia. V krátkosti, Lagranžiány majú vždy tvar podobný ako

$$L = \alpha\, a_i\, b^i + ß\, c_{jk}\, d^{jk} + \gamma\, e_{lmn}\, f^{lmn} + \qquad (100)$$

kde indexy pri premenných a, b, c atď., vždy sa objavujú v sumačných pároch (Einšteinove pravidlo). Grécke písmená reprezentujú konštanty.

Reprodukovateľnosť pozorovaní, t.j. symetria voči zmene okamihu času alebo 'časová translačná invariancia', je prípadom nezávislosti hľadísk. Toto spojenie má niekoľko dôležitých následkov. Videli sme, že táto symetria implikuje invarianciu. Vedie to k tomu, že pre spojité symetrie, ako je časová translačná symetria, toto tvrdenie sa dá urobiť presnejším: pre hocijakú spojitú symetriu Lagranžiánu existuje asociovaná zachovávaná konštanta pohybu a naopak. Presná formulácia tohto spojenia je teoréma *Noether* (1918). Jej závery, výsledok platí nielen pre vyššie uvedený typ Lagranžiánu, ale aj pre ľubovolný typ Lagranžiánu.

E. Noether vyšetrovala spojité symetrie závisiace na spojitom parametre b. Transformácia hľadiska je symetrickou ak účinok S nezávisí na hodnote b.

$$S = \int L \partial t \ . \qquad (101)$$

Napríklad, zmena polohy ako $x \rightarrow x + b$ necháva účinok invariantným, pretože $S(b) = S_0$, táto situácia implikuje, že p = konst ; v krátkosti, symetria voči zmene polohy implikuje zachovanie hybnosti. Opak je tiež pravdou.

V prípade symetrie voči posunu okamihu pozorovania, nájdeme T + U = konštanta; že časová translačná invariancia implikuje konštantnú energiu. Opäť, opak je tiež správny. Tiež hovoríme, že energia a hybnosť sú generátormi časových a priestorových translácií.

Zachovávaná veličina pre spojitú symetriu sa niekedy nazýva Noetherovej náboj, pretože výraz náboj sa používa v teoretickej fyzike na označenie zachovávajúcich sa pozorovateľných. Inými slovami, energia a hybnosť sú Noetherovej náboje. 'Elektrický náboj', 'gravitačný náboj' (t.j. hmotnosť) a 'topologický náboj' sú iné obvyklé príklady.

V našom prípade percepcie reči, je situácia trochu zložitejšia. Predpokladáme na základe experimentálnych údajov, že grupa transformácií nie je ani ábelovská, a nie je ani globálna. Zachovávajúca sa veličina, relevantná k fonémickému systému, bude mať pravdepodobne charakter 'topologického náboja'.

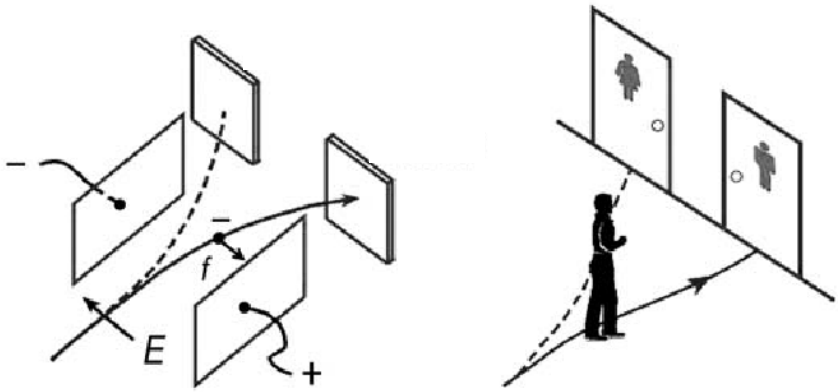
Z porovnania (A2.4) a (A3.9) môžeme usúdiť, že prípad „papagája" (A2.4) je analogický k „ľudskému" prípadu, (A3.9). Túto korešpodenciu môžeme interpretovať ako prechod z „mechanického" - konečný počet stupňov voľnosti - chápania percepcie k „poľnému" -nekonečný počet stupňov voľnosti - chápaniu. Z predošlého sa tak stáva kritickou otázka, či princíp minimálneho účinku sa dá preniesť aj na živú, poprípade mysliacu „hmotu". Pracovne ho nazvime *Princípom najmenšieho úsilia.* Nie druhoradou a ani jednoduchšou je otázka súvislosti spomínaných dvoch princípov.

Tento princíp by mal vystihovať rozdiel medzi fyzikálne ovládanými interakciami a vzormi ovládanými interakciami (tu sú fyzikálne interakcie druhoradé, sú len nosičom informácie), pozri Obr. 51. Na tomto obrázku sú znázornené dva prípady, jeden vľavo pre neživú hmotu, Stern-Gerlachov experiment, vpravo je znázornený výber robený podľa typu vzoru, ktorý percepuje agent (človek, zviera). Pravdepodobne aj pojem Shanonovej informácie sa bude musieť rozšíriť tak, aby postihoval nielen štatisticky danú informáciu (nie sémantickú) ale aj sémantický význam daných vzorov (ak je to principiálne možné).

7.3.2 Model percepcie

Ako sme videli v časti 3 a 4 percepcia sa dá chápať ako odhadovanie externého sveta z daných okamžitých senzorických údajov, informácie priradením k príkladom už percepovaných, poznaných kategórií. Zvieratá s veľmi jednoduchými nervovými systémami percepujú niekedy tak dobre a rýchlo ako zvieratá so zložitými nervovými systémami, napríklad ako sme my, a za niektorých okolností, v niektorých prípadoch možno lepšie ako my (samozrejme nehovoríme tu len o percepcii reči). Zvieratá aj my sa vysporiadavame dobre s percepciami, ktoré vyžadujú iba rozpoznávanie objektov ako príkladov neživých kategórií a nie naučených kategórií. Po dlhých rokoch skúmania

sme zistili, že zvieratá percepujú tak rýchlo a dobre, skoro akoby ani nemuseli vynakladať nejaké úsilie, že si až teraz začíname uvedomovať, že pri programovaní našich počítačov na úlohu percepcie existuje vážny problém, *Gregory* (1997).



Obr. 51 Vľavo fyzikálne ovládaná interakcia - výber, vpravo paternom ovládaná interakcia - výber

V čom spočíva problém percepcie, v predošlom načrtnutý ? Pod priamou percepciou rozumieme nejaký algoritmus, ktorý je schopný percepovať v polynomiálnom počte krokov alebo čase, niekedy sa takýto druh percepcie nazýva aj percepcia zospodu-navrch. Také algoritmy sa pokúšajú konštruovať popisy objektu (nazýva sa to vrchná úroveň) z práve prístupných senzorických dát (spodná úroveň).

Tieto prístupy buď nefungujú alebo len pre veľmi obmedzené, umelé dáta. Pri percepcii sa stretávame aj so situáciami, kde percepcia zospodu-navrch nemá jednoducho s čím začať, pretože celá explicitná informácia je potlačená, je na úrovni pozadia, zašumená alebo chýbajúca alebo chybná *Gregory* (1997), *Lieberman, Cooper, Shankweiler, Studdert-Kennedy* (1967). V takom prípade vstupuje do hry nepriama percepcia.

Algoritmy, ktoré sú úspešnejšie sa nazývajú nepriama percepcia alebo zvrchu-naspod percepcia. Začínajú na vrchnej úrovni - objektoch a hľadajú zhodu s senzorickými práve dostupnými dátami. Pretože, z povahy úlohy musia vyskúšať nie polynomiálny počet hypotéz o objektoch, ich vykonanie zaberá nie polynomiálny počet krokov alebo čas, *Pisoni, Remez* (2005).

Na druhej strane, aj keď v zvukovom alebo elektromagnetickom vzore je postačujúce množstvo lokálnej a explicitnej informácie, ani vtedy to nemusí byť postačujúce na vytvorenie vnemu. Je nutnou aj znalosť o tom čo pravdepodobne sú dané objekty. Percepcia objektu je primárna. Nepočujeme, nevidíme to čo si *myslíme*, že počujeme, vidíme - vnímame to formovaním hypotéz. Bez nejakej hypotéze o objekte je nemožné správne naložiť s práve snímanými senzorickými dátami, *Moore, Tyler, Marslen-Wilson* (2007), *Gregory* (1997).

Z trochu iného pohľadu, videli sme to aj v časti 2, a 4 tejto práce, všetky počítačové algoritmy, ktoré sú schopné úspešne riešiť nie triviálne percepčné problémy majú v sebe zakomponované hľadanie globálneho minima nejakej funkcie. To funguje pre umelý jednoduchý kvázi-svet, *Pisoni, Remez* (2005), a iba vtedy ak sa kategórie skladajú z malého počtu rigidných objektov, s iba malým počtom stupňov voľnosti, *Gross* (1996), ale ak začneme vyšetrovať neregulárne, elastické a apriori nie existujúce - podobné ako Wheelerovské kategórie, *Wheeler, Zurek ed.* (1983), pozri posledný odsek v časti 3, s veľkým počtom interných premenných a potenciálne s nekonečným počtom stupňov voľnosti - ako sú polia, také aké sa vyskytujú v reálnom svete, potom sa dimenzia priestoru riešení stáva príliš veľkou pre vyčerpávajúce hľadanie v tomto priestore.

Takto problém percepcie je efektívne ekvivalentný problému minimalizácie triedy funkcií viacerých premenných. Dá sa argumentovať, z faktu rýchlej percepcie zvierat, že pravdepodobne je postačujúce nájsť „dobré" minimum, nie globálne minimum. Doposiaľ však všetky takéto prípady, keď sa algoritmy zastavili pred nájdením globálneho minima, boli neúspešné, čo sa aj dalo očakávať z povahy problému, *Pisoni, Remez* (2005).

Ľubovolný klasický systém prevádzajúci takú minimalizáciu môže hľadať vo fázovom priestore iba lokálne, takto ak je daná funkcia neregulárna a obecná, systém musí prehľadať nie polynomiálny počet lokálnych miním predtým než nájde globálne minimum. A toto je hlavná príčina prečo „percepujúci" počítač musí obecne generovať nie polynomiálny počet hypotéz o objekte a každú z nich vyhodnotiť. S veľkou pravdepodobnosťou ale takto nepercepujeme.

Z týchto dôvodov predpokladáme, že problém percepcie môže odpovedať problému nájdenia globálneho minima funkcie viacerých premenných, kde globálne minimum je omnoho hlbšie než iné minimá. Predpokladáme, že globálne minimum je omnoho hlbšie z dôvodov konštrukcie funkcie, ktorú treba vyšetriť.

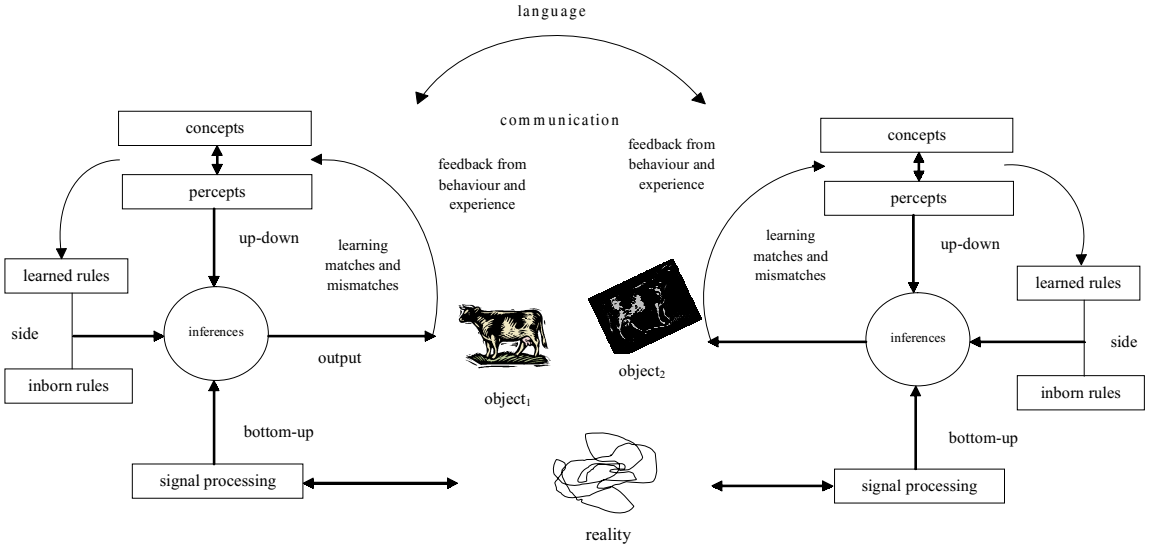
Funkcia, ktorú treba minimalizovať je mierou rozdielu, napríklad podľa (99) medzi percepciou a senzorickými dátami v danom čase. Chceme minimalizovať funkciu vzhľadom k množine premenných, ktoré tvoria nejaký vyhľadávací strom, opisujúci externý svet, o ktorom sa predpokladá, že generuje dané dáta - hypotézu o objekte. Takto budeme mať iba jedno minimum, pretože dáta majú omnoho vyššiu podmienenú entropiu než dáta vyhľadávajúceho stromu. Inými slovami dáta nie sú šumom, majú nejakú vnútornú konzistenciu, v tom že je možné že sú generované z omnoho menšieho vyhľadávajúceho stromu. Opačne, ak sú dáta nerozlíšiteľné od šumu, potom nebudeme mať nejaké dominantné globálne minimum. Tu predpokladáme, že pravdepodobnosť, že nejaký iný vyhľadávací strom bude dobre fitovať dáta je veľmi malá.

Tieto úsudky potvrdzuje aj to, že algoritmy, ktoré prerušia svoje hľadanie produkujú podstatne veľké chyby, čo môže znamenať, že správne interpretácie sú zriedkavé, a takisto že dvojznačné interpretácie sa v prírode vyskytujú zriedkavo (ako Neckerova kocka). Indikujúc, že prípady s dvomi porovnateľne alebo rovnako hlbokými minimami sú v prírode zriedkavé (ale môžu byť dôležité z hľadiska evolúcie, pri prechode od zvieracej k ľudskej percepcii).

Po týchto predbežných úvahach, pre ľudských pozorovateľov, k zvyčajným častiam ako je 'zdola-navrch' senzorickým signálom a 'zvrchu-nadol' znalostiam pre obecné popisy percepcie *Gregory* (1997), *Pisoni, Remez* (2005), pridáme 'bočné členy' ako pravidlá a komunikáciu medzi pozorovateľmi vedúcu k jazyku. Aj zvrchu-nadol aj bočné členy sú vo svojej podstate znalosti; prvé sú špecificky naučené (také ako zvuky súc príliš nahlas, alebo frikatívne alebo tvar súc kompaktný), druhé sú obecné - vrodené pravidlá aplikované na všetky objekty a scény (také ako zákony symetrie, ako sú uvedené na Obr. 50, Gestalt pravidlá, možno tiež symetrie). Na Obr. 52 je schématický načrt hlavných častí nášho modelu ako bol navrhnutý na Obr. 23, 24, teraz zdôrazňujúci špecifické časti, funkcie a procesy, ktoré sú podľa nás najdôležitejšie, *Pisoni, Remez* (2005)

Ako zvyčajne, signály z ucha alebo očí a iných zmyslov sú typu 'zdola-nahor'. Konceptuálna a perceptuálna znalosti o objekte sú znázornené v samostatných 'zvrchu-nadol' boxoch. Znalosti, ktoré sa dotýkajú naučených a vrodených pravidiel sa znázornené ako 'bočné vetvy'. Perceptuálne učenie sa tu chápe, že funguje z veľkej časti pomocou spätnej väzby od správania a činnosti.

Z hľadiska percepcie je najdôležitejšou časťou „inferenčná" časť - generátor hypotéz a ich odhadca. Je časťou podobnou hre J. A. Wheelera o kladení otázok a hádaní slova, *Wheeler, Zurek ed.* (1983), pozri posledný odsek v časti 3, v našom prípade nie so slovami ale so senzorickými dátami, ktoré pochádzajú z reality a nepriamo s dátami iných percepujúcich agentov, cez jazyk. Úloha reality je poskytnúť nejakú informáciu, inými slovami slúži iba ako nejaký bitový substrát. Iným zdrojom informácie je samotný mozog. To, ktorý z týchto zdrojov informácie sa používa v danom čase je vecou „vedomia" a/alebo pozornosti. Samotný zdroj informácie spojený s mozgom sa používa v stavoch predstavovania, snov a podobne. Pomocou obojstraných šípok medzi časťou spracovania signálov a realitou, sme chceli zdôrazniť aktívnu úlohu receptorov - vlasových buniek, tyčiniek, čapíkov, atď.. Nie sú pasívnymi prijímačmi informácie ale aktívnymi, *Lieberman, Cooper, Shankweiler, Studdert-Kennedy* (1967), *Foss, Swinney* (1973).



Obr. 52 Schéma percepcie pre systém troch entít - vnímaná realita a dva vnímajúce subjekty

Nemyslíme si, že všetky myšlienky a špekulácie tu uvedené sú perfektné v každom detaile, ale veríme, že v určitom koherentnom pohľade môžu byť užitočné a možno aj pravdivé.

8 Záver

V dizertačnej práci sme prediskutovali jeden z možných prístupov k spracovaniu rečového signálu na základe invariantných fyzikálnych, biologických a kybernetických predpokladov. Nami navrhnutý model, ktorý modeluje primárne rečové úlohy – rozpoznávanie slov a verifikáciu hovoriaceho, sme definovali konceptuálne, matematicky a realizovali sme ho programovo v reálnych podmienkach.

Hlavne ako heuristický princíp sme použili a skonštruovali aj prístup cez grupy symetrií, ktorý je implicitne zahrnutý v modeli pomocou homotopických a kalibračných grúp symetrie. Zároveň sa snažíme formulovať niektoré fundamentálne koncepcie k prístupu obecnej teórie poľa na generovanie a spracovanie symetrií v percepcii a spracovaní rečového signálu nervovým systémom človeka, obecne agenta (človek, zviera, stroj).

Diskutovali sme niektoré zaujímavé, neštandartné črty nášho prístupu z hľadiska ľudskej rečovej percepcie a symetrií prirodzených jazykových systémov, špeciálne slovenčiny a dôsledky nášho modelu z hľadiska informačného chápania symetrií.

Otvoreným okruhom otázok sú nasledujúce, ktoré chápeme ako vedecky testovateľné otázky resp. hypotézy:

1. Je percepcia reči a obecne percepcia invariantná na vonkajšie a vnútorné symetrie ako sme ich formulovali v časti 4 ? (*všetky pozorovania to potvrdzujú, samozrejme s istým stupňom idealizácie, pozri Obr. 53*)

2. Súvisia invariantnosť voči intenzite a šumu s kalibračnou symetriou ? (*Preukázanie klasickej kalibračnej invariancie sa môže previesť prevedením zvukových rovníc na Maxwellove rovnice, pretože Maxwellove rovnice sú klasicky kalibračne invariantné, Landau, Lifshitz* (1987))

3. Je klasifikačná schéma (symetria) elementárnych častíc nezávislá od klasifikačnej schémy (symetrie) foném ? (*Otázka otázok, dá sa vedecky testovať ?*)

4. Je nutné na popis symetrie foném použiť systémy s nekonečne veľa stupňami voľnosti - polia? (*Dobrým návodom by mohla byť požiadavka spontánneho narušenia symetrie*)

5. Je možné "predĺžiť" princíp minimálneho účinku z neživých systémov na živé vo forme nejakého princípu minimálneho "úsilia" ? (*Otázka otázok*)

6. Aký význam majú topologické náboje, z hľadiska percepcie a fonémického systému ?

7. Čo predstavujú fonémy, odpovedajúce prúdy topologických nábojov v mozgu (v "mysli") ?

8. Čo je analógom hmotnosti elementárnych častíc v systéme foném ?

9. Je to nejaká sémantická "hmotnosť", analogická gravitačnej/zotrvačnej hmotnosti ? (*Newtonov gravitačný zákon aj Einšteinov zákon gravitácie sa dajú odvodiť len z úvah o entropii a holografického princípu, Verlinde* (2010))

10. Dá sa otestovať náš model percepcie z hľadiska vrodených schopností využívať symetrie ? (*Áno, testuje sa niekoľko desaťročí*)

11. Dá sa otestovať náš model percepcie z hľadiska toho, že nielen percepty a koncepty ale aj objekty sú v "mysli" a nie vo vonkajšom svete ? (*Asi áno, ale veľmi ťažko*)

12. Dá sa otestovať náš model percepcie z hľadiska toho, že fyzikálna realita tu vystupuje len ako zdroj informačného toku ? (*Dá sa to, možno je to triviálne, len otázka definície*)

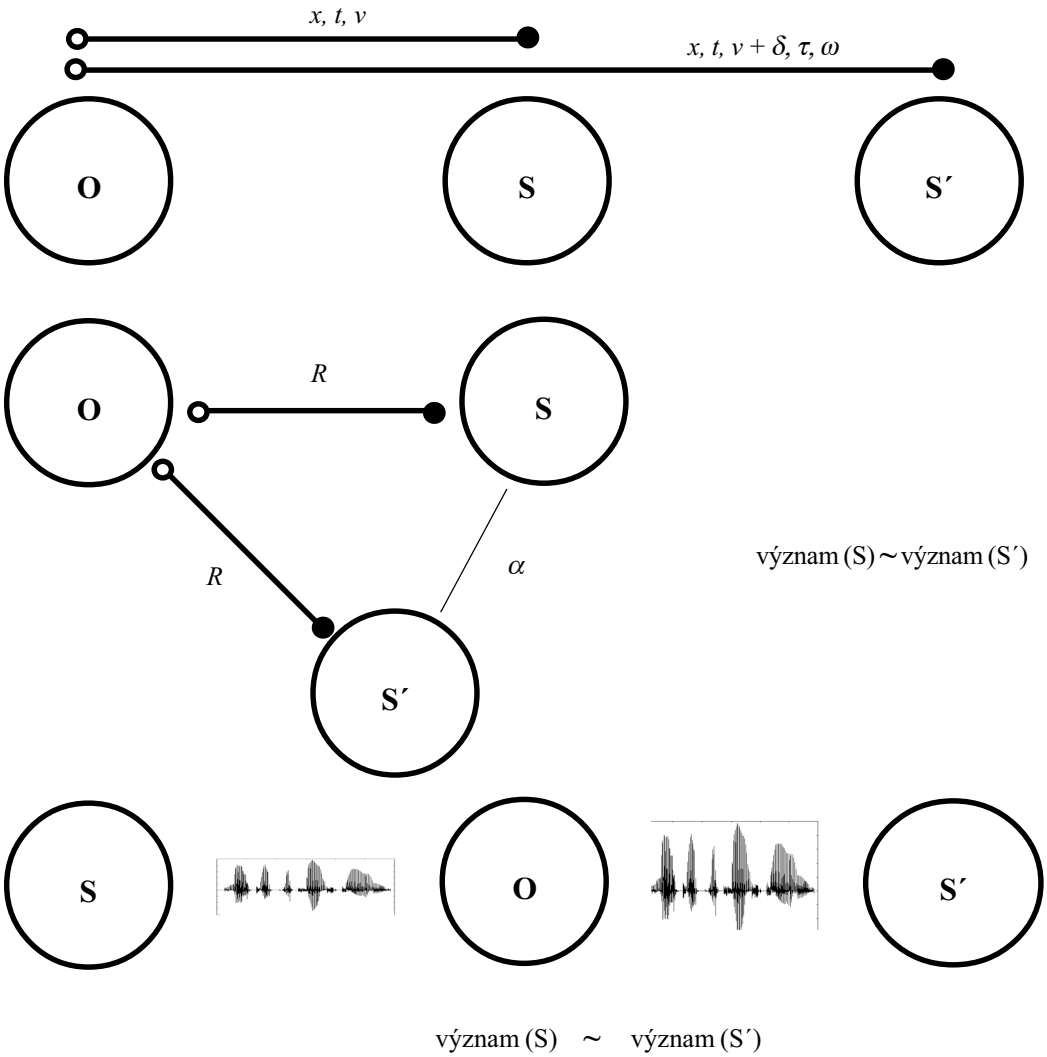
13. Je možné takýto model percepcie pochopiť ako interpretáciu kvantovej mechaniky z hľadiska vzťahu pozorovania a "vedomia" ? (*Ťažká otázka*)

14. Aký je vzťah medzi postulovanou nerozlíšiteľnosťou elementárnych častíc a postulovanou nerozlíšiteľnosťou foném ? (*Je to totožné - predpoklad, ktorý sa nedá testovať, pokladá sa za platný dovtedy kým sa nepreukáže opak*)

15. Aký je význam štatistík - Fermi-Diraca a Bose-Enšteina v systéme foném ? (*Zložité*)

16. Aký je význam foném z pohľadu rozdelenia fermiónov a bozónov, alebo leptónov ? (*Zložité*)

17. Prečo nie je "vidieť" usporiadanie na úrovni rodín kvarkov a leptónov v systéme symetrií foném? (*Môže za to jav "uväznenia" kvarkov - charakter ich interakcie ?*).



Obr. 53 Schématická ilustrácia transformácií, ktoré zachovávajú význam slov alebo viet. ***O*** *označuje objekt a* ***S, S´*** *subjekty. (a) relatívna poloha , relatívny časový posun pozorujúceho systému vzhľadom k zdroju rečového signálu, relatívny pohyb, (b) relatívna rotácia, (c) intenzita rečového signálu, akustický a fonetický šum, základný tón rečového signálu, rýchlosť hovorenia*

Z hľadiska budúceho rozvoja oblasti výskumu je pravdepodobne dôležité sa zamerať hlavne na vzťah kvantovej mechaniky a nášho modelu. Je nutné korektným spôsobom zakomponovať formalizmus kvantovej mechaniky do formalizmu neurónových sietí. Samozrejme pri tom vyvstáva otázka, či má zmysel kvantová mechanika pri popise, výskume mozgu, obecne resp. "vedomia" špeciálne, za predpokladu, že "vedomie" sa dá vedecky definovať. Je mozog kvantový systém ? Je to kritická otázka, ktorú si musíme položiť pre ďalšie uvažovanie. Aby mala táto otázka zmysel, musíme buď predpokladať alebo postulovať, že mozog je fyzikálny systém. Ak je mozog fyzikálny systém, potom je aj kvantový systém - pretože má tvar a farbu, to sú postačujúce podmienky pre každý fyzikálny systém, aby bol kvantový. Pod farbou myslíme schopnosť objektu emitovať alebo pohltiť elektromagnetické žiarenie špecifikovaných vlnových dĺžok a pod tvarom rozumieme fyzikálny objekt ohraničený v objeme alebo veľkosti. Pri tom nezávisí na veľkosti objektu, kvantový popis sa niekedy musí použiť aj na makroskopické objekty (detektor gravitačných vĺn - hliníkový valec o rozmeroch rádove metre, ale samozrejme pri veľmi nízkej teplote, okolo absolútnej nuly, *Ju, Blair, Zhao* (2000)).

Iná je otázka prítomnosti alebo detekovateľnosti kvantových javov vo vedeckom slova zmysle - exaktne, objektívne, verifikovateľne. Základná námietka proti mozgu ako kvantovému systému je jeho teplota a štruktúra. Podobne by sa ale mohlo argumentovať aj pre flash pamäte, pri ktorých

pre ich fungovanie je podstatný kvantový (presnejšie kváziklasický) jav tunelovania (a držíme ich v rukách). Pri flash pamätiach hrá dôležitú úlohu Fowler-Nordheim tunelový jav, ktorý je kvantovo-mechanický a deje sa na spojení kov polovodič, *Lopez-Villanueva a kol.* (1991). Existujú podobné štruktúry a podmienky, ako sú vo flash pamätiach, aj v mozgu ? Pravdepodobne áno, možnými kandidátmi na to sú synaptické štrbiny medzi ukončeniami neurónov. Tok neurotransmiterov a pravdepodobnosti ich prechodu cez štrbinu sa pravdepodobne (špekulatívne) nedajú popísať čisto klasicky, ale podobne ako pri flash pamätiach kváziklasicky - pomocou Schrödingerovej rovnice. Mohol by o tom, napríklad, nasvedčovať aj rozmer spomínanej štrbiny v porovnaní s vlnovou dĺžkou nosičov prúdu - neurotransmiterov.

$$\lambda = \frac{h}{p} = \frac{6.6*10^{-34}\,kgm^2s^{-1}}{mv} = \frac{6.6*10^{-34}\,kgm^2s^{-1}}{v160*10^{-26}\,kg}$$

$$v = \frac{\partial}{\tau} = \frac{10^{-8}\,m}{10^{-6}\,s} = 10^{-2}\,ms^{-1}$$

kde sme použili hodnotu Planckovej konštanty h = 6,626 069 x 10^{-34} (kg.m^2.s^{-1}) a Avogadrovu konštantu N_A = 6,022 142 x 10^{23} mol^{-1}. Za hmotnosť charakteristickú hmotnosť neurotransmiterov sme zobrali hmotnosť dopamínu rovnú približne 160g/mol. Takto jedinou veličinou, ktorej odhad je veľmi nepresný, je rýchlosť molekúl dopamínu v pri prenose synaptickou štrbinou, čo sa klasicky dá chápať ako difúzia *Jonas* (2000). Veličiny, ktoré sme pri odhade uvažovali je šírka synaptickej štrbiny 0,1 - 100 nm, a nejaký maximálny stredný čas molekúl dopamímu, v našom odhade 1 μs, ktorý potrebujú na difundovanie skrz synaptickú štrbinu, *Garris, Ciolkowski, Pastore, Wightma* (1994). Z posledných vzťahov vidíme, že deBroglie-ho vlnová dĺžka molekuly dopamínu môže byť len 1 až 2 rády vyššia ako šírka synaptickej štrbiny. A pretože náš odhad rýchlosti molekúl dopamínu je skôr podcenený, môže vlnová dĺžka dosiahnúť hodnôt šírky synaptickej štrbiny. Všetky tieto úvahy ale narážajú na otázku či otvorený ("živý") systém sa dá popísať fyzikálne, či tam má zmysel napríklad teplota - dobre definovaná pre rovnovážne a uzavreté systémy a podobne.

Ďalšou možnosťou pre pozorovanie kvantových javov v mozgu je jav kvantovej lineárnej superpozície, v jeho elementárnom chápaní, 3.2 (Časť a celok v kvantovej fyzike). Môžeme pozorovať túto superpozíciu pre nejaké javy v mozgu ? A ak, ako dlho, či je stabilná a hlavne ktoré stavy sa na jej realizácii zúčastňujú. S čím v mozgu môžeme identifikovať tieto hypotetické stavy ?

Platia tu podobné námietky ako proti mozgu ako kvantovému systému a navyše ešte námietka koherentnosti - odpovedajúcich stavov v lineárnej superpozícii, ktorej stredný čas je porovnateľný s časovými intervalmi, ktoré zodpovedajú za deje myslenia, uvedomenia a podobne. Inými slovami, podmienkou na pozorovanie kvantovej superpozície je aby stavy, ktoré sa účastnia tejto superpozície, boli koherentné - fázový rozdiel ich možností - bol konštantný dlhšie ako je dekoherenčný čas *Tegmark* (1993), *Zeh* (1970), Zurek a kol. (1993). Tu pravdepodobne končí chápanie mozgu ako "čisto" fyzikálneho systému.

Ako posledná možnosť pre pozorovanie kvantových javov, ktoré súvisia s lineárnou superpozíciou v mozgu je "kolaps" vlnovej funkcie. Jav, ktorý úzko súvisí s interpretáciou kvantovej mechaniky, interpretáciou pravdepodobností obecne a s tým, či "vedomie" súvisí s kolapsom vlnovej funkcie. Posledné experimenty, ktoré skúmajú základy kvantovej mechaniky čoraz jasnejšie napovedajú, že "vedomie" má vplyv, súvisí s kolapsom vlnovej funkcie - samozrejme nie kvantifikujúcim spôsobom ale klasifikujúcim, *Nairz, Arndt, Zeilinger* (2003). Pravdepodobne tento jav je najlepší kandidát na spojenie percepcie, tak ako sme ju opísali v časti 3, a kvantovej mechaniky**,** *Palacios-Laloy a kol.*(2010), *Lee a kol.*(2011). Pokiaľ ale nemáme vedecky dobre definované pojmy ako percepcia, vedomie a myseľ sú všetky tieto úvahy špekulatívneho, v najlepšom metafyzického charakteru.

A1 Začiatok a koniec slov

Začiatok a koniec slova sa musí určiť s čo najväčšou dôkladnosťou, bez ohľadu na typ úlohy - rozpoznávanie slov alebo verifikácia hovoriaceho. V konečnom započítaní úspešnosti modelu rozpoznávania slov alebo verifikácie hovoriaceho prispieva procedúra určenia začiatku a konca slova s viac ako 5 % do celkovej úspešnosti, *Rabiner, Samabar* (1977), *Nooteboom, Van der Vlugt* (1988), *Lamel, Rabiner, Rosenberg* (1981). Treba si zároveň uvedomiť, že vo väčšine databáz izolovaných slov je procedúra určenie začiatku a konca slov prevedená ručne, inými slovami nezahrňuje sa do celkovej úspešnosti modelu (stav do roku 1995).

V akademickom prostredí sa tento problém jednoducho nerieši, pretože sa väčšinou pracuje v podmienkach blízkych k ideálnym, z hľadiska kvality zaznamenanej reči (či kvalita prednesu samotných slov, či kvalita prenosu signálu - kvalitný mikrofón, kvalitný AD prevodník, atď.), alebo sa predpokladá implicitne, že je vyriešený. V komerčných aplikáciach sa samozrejme táto problematika musí riešiť, ale sa samozrejme nepublikuje, pretože je to kritická časť celého systému. Väčšinou sa komerčné aplikácie, na rozdiel od akademických, sústreďujú na reálne, resp. ťažké akustické a elektronické podmienky, napríklad telefónny prenos, *Atal, Rabiner* (1976).

Z tohoto dôvodu sme vyvinuli algoritmus na určenie začiatku a konca slov, ktorý sa dá použiť aj v reálnom čase, aj pre on-line implementácie konkrétnych úloh rozpoznávania alebo verifikácie a samozrejme bez ohľadu na kvalitu signálu (aj telefónnu linku). Algoritmus prevádza rozpoznávanie začiatku a konca slova pomocou základných parametrov energie a prechodov nulou, *Niederjohn* (1975), *Scarr* (1968). Algoritmus je navrhnutý ako samo-nastaviteľný, inými slovami, všetky konštanty, alebo parametre potrebné pre výpočet veličín sa vypočítavajú z prvého časového okna signálu. Vzhľadom k tomu, že v algoritme sa vyskytujú len súčty, nelinearity a nerovnosti je možné tento algoritmus triviálnym spôsobom „neuronalizovať", *Fallside* (1988). Jedinými vstupnými, voliteľnými parametrami sú prekryv okien, dĺžka rámca a rozlíšenie AD/C prevodníka.

Vo zvolenom časovom rámci máme takto definovaných NW okien. Dĺžka časového rámca sa môže odhadnúť ako stredné maximálne trvanie neznelých plozívnych a frikatívnych hlások v reči, čo je približne 160 ms, *Reddy* (1966). Dĺžka časového okna je okolo 10 ms, čo vyplýva hlavne z neurónovej štruktúry sluchových receptorov a tzv. mŕtvej doby nervových vlákien. Na každom okne sa vypočítajú energia a prechody nulou, tieto sa pre dané okno časovo dozadu (retardovane) ustrednia a klipujú - prahujú pomocou parametrov vypočítaných v prvom kroku. Výsledkom je NW logických hodnôt „Reč" - „Nie Reč". V prípade začiatku slova, ak je počet okien označených ako „Reč" väčší ako NW/2, potom máme tzv. hrubý začiatok slova pre prvé okno s hodnotou „Reč". Podobne pre koniec slova, ak je počet okien označených ako „Nie Reč" väčší ako NW, potom máme tzv. hrubý koniec slova pre prvé okno s hodnotou „Nie Reč". Algoritmus sa skladá z troch funkčných blokov, Obr. A1.1.

1. krok algoritmu: určí z prvého signálového okna parametre pre ďalšie spracovanie

Resolution = rozlíšenie AD/C prevodníka
Ret_end = NWF/2
Ret_bgn = Ret_end/5
Ret_dec = 2
Ratio = Max_amplitude(Resolution)/256

$$\text{Level} = \text{Int}\left(\sum_{i=0}^{N-1} x_i / N \right)$$

$$\text{Noise} = \text{Int}\left\{ \left(\sum_{i=0}^{N-1} \text{Abs}(\text{Int}(x_i) - \text{Level}) \right)/N \right\}$$

$$\text{Threshold} = \{\text{Abs}(\text{Level}) + \text{Int}(\text{Noise}) + \text{Resolution}/3\}/\text{Ratio}$$

Ak Threshold = 0 potom Threshold = 1

Threshold_bgn = (Threshold * Ratio - 2) / 16

Ak Threshold_bgn < 0 potom Threshold_bgn = 1 / 16

Threshold_end = Threshold_bgn

Threshold_inc = 0

2. krok algoritmu: určí hrubý začiatok a koniec slova

zober ďalšie okno

vypočítaj normalizovanú energiu

$$\text{Energy} = \sum_{i=0}^{N-1} \left\{ (\text{Int}(x_i) - \text{Level}) / \text{Threshold} \right\}^2 /N$$

vypočítaj prechody nulou

$$\text{Zero} = \sum_{i=0}^{N-1} ((x_i > 0 \;\&\&\; x_{i+1} < 0) \,||\, (x_i < 0 \;\&\&\; x_{i+1} > 0))$$

vypočítaj hodnotu pre retardované okno

$$\text{Value} = \text{Int}\{ (\text{Resolution} + 2*\text{Zero}/\text{Resolution}) * \log_{10}(1 + \text{Energy}*\text{Zero}) \}$$

zapamätaj si hodnotu do bufra pre dané okno

Buffer(okno) = Value

vypočítaj retardované hodnoty pre dané okno a začiatok resp. koniec slova

$$\text{Retard\_bgn} = \{ \sum_{i=a}^{b} \text{Buffer}(i) \} / \text{ret} ; \qquad \text{Retard\_end} = \{ \sum_{i=a}^{b} \text{Buffer}(i) \} / \text{ret}$$

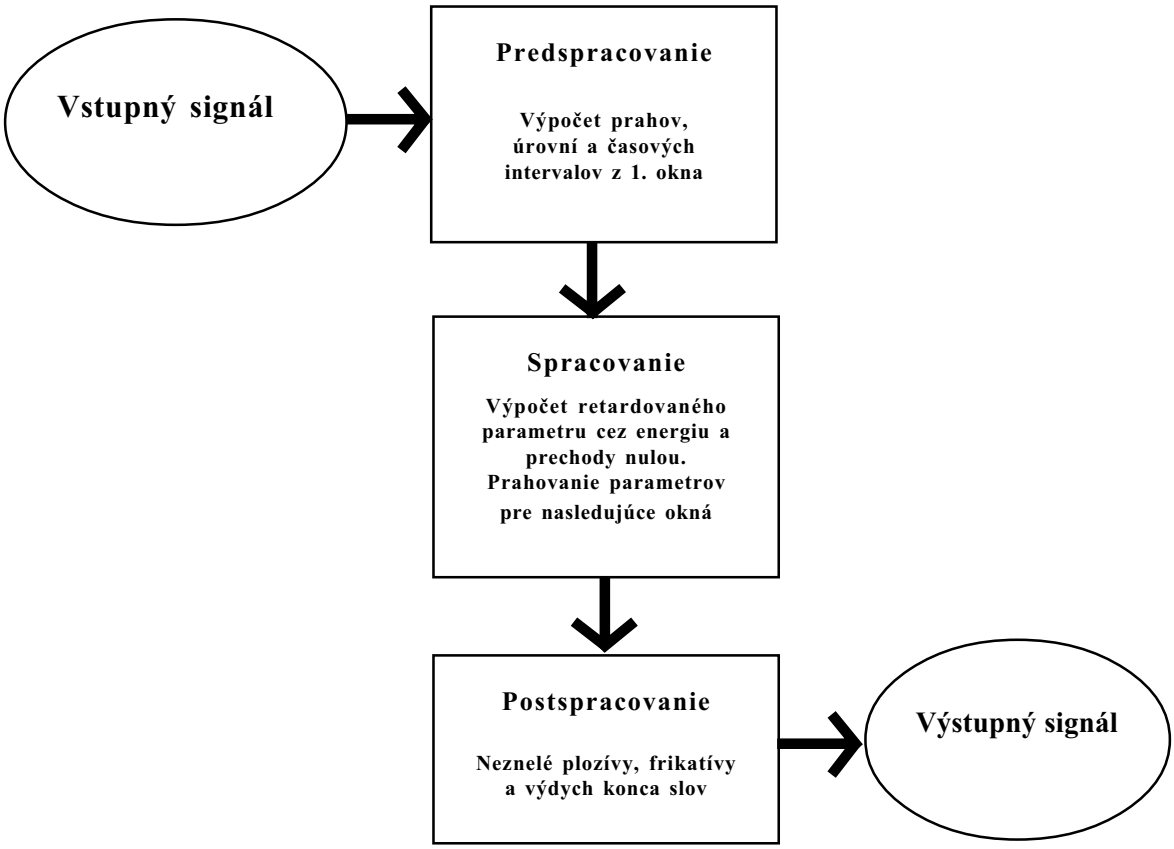
ak je okno < Ret_bgn potom (a=0; b=okno - 1) ináč (a= okno-Ret_bgn; b=Ret_bgn)

ak je okno < Ret_end potom (a=0; b=okno - 1) ináč (a= okno-Ret_end; b=Ret_end)

klipuj všetky hodnoty pre dané okno

ak Retard_bgn(okno) > Threshold_bgn potom clipp_bgn = 1 ináč clipp_bgn = 0
ak Retard_end (okno) > Threshold_end potom clipp_end = 1 ináč clipp_end = 0

ak je počet spracovaných okien menší ako rámec choď na 2. krok
ináč vyhodnoť podmienky začiatku a konca slova

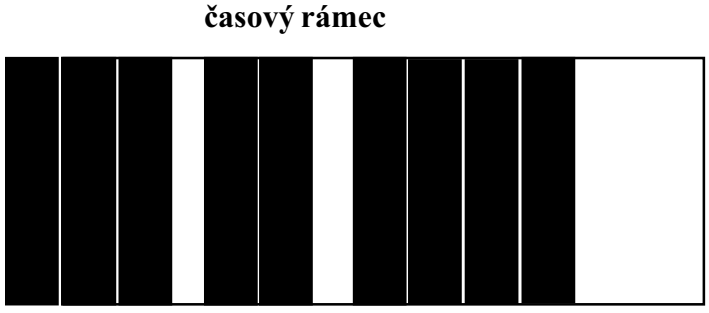


Obr. A1.1 Algoritmus pre začiatok a koniec slova

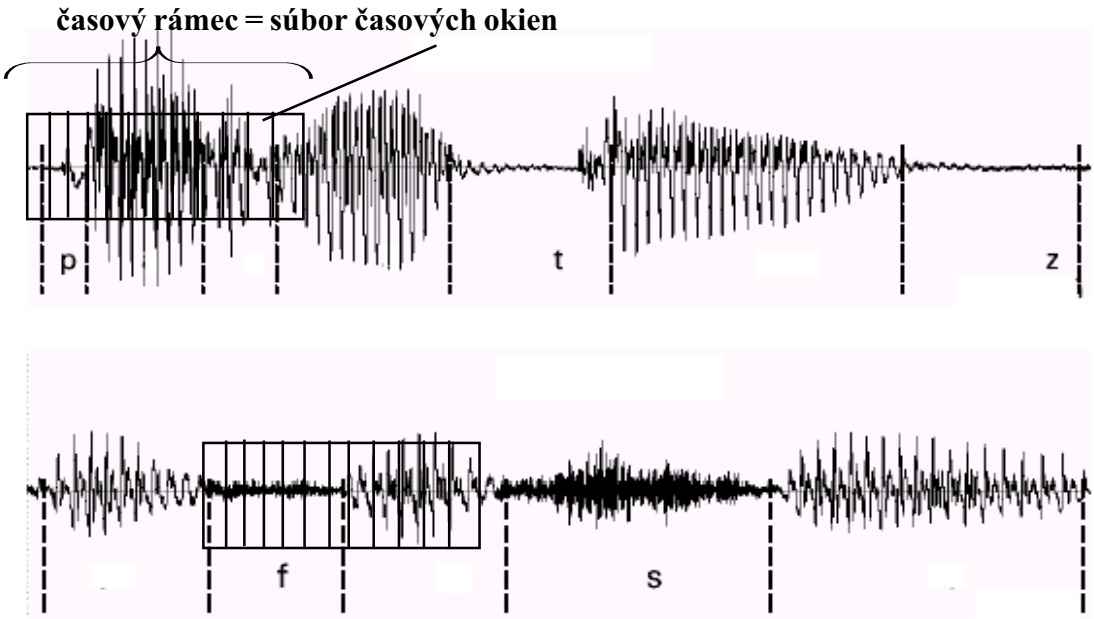
Ak nie je splnená ani jedna z nasledujúcich podmienok prechádza sa na krok 2, pričom nový rámec je posunutý o jedno okno dopredu v čase.

v prípade začiatku slova ak je počet okien v rámci označených Reč väčší ako polovica okien v rámci, potom prvé okno takéhoto rámca značí hrubý začiatok slova.
v prípade konca slova ak je počet okien v rámci označených Nie-Reč väčší ako počet okien v rámci, pozri schému na Obr. A1.1 a Obr. A1.2 a spodný obrázok, potom posledné okno takéhoto rámca značí hrubý koniec slova.

3. krok - upresnenie začiatku a konca slova



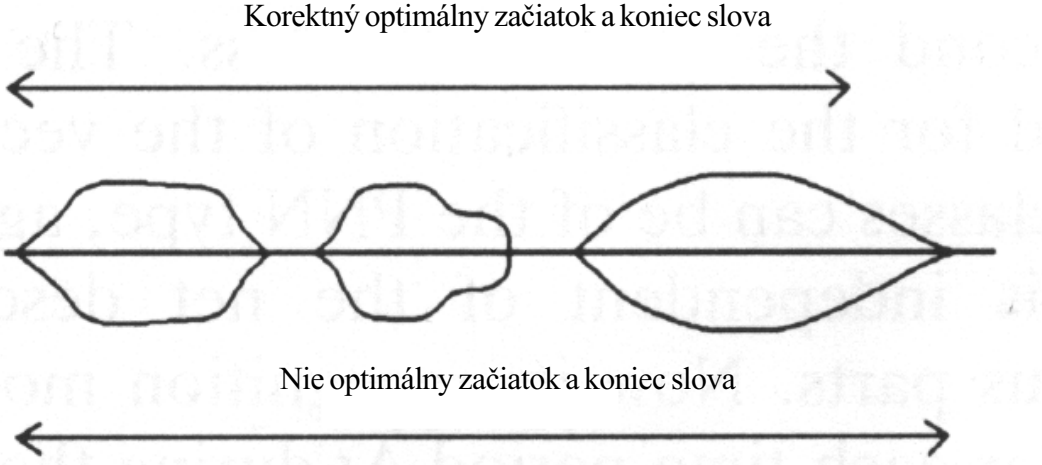
Legenda: čierne okná v rámci znamenajú - dané okno je označené za Reč
biele okná v rámci znamenajú - dané okno je označené za Nie-Reč



Obr. A1.2 Rečový signál ilustrujúci algoritmus pre začiatok a koniec slova

dôvodom tohto kroku je výdych na konci slova, fakticky prítomý pre všetky dlhé samohlásky (v slovenčine) a niektoré situácie na začiatku slova ako sú neznelé plozívy a frikatívy, *Siegel, Bessey* (1982), Obr. A1.3.

Výsledkom predošlého kroku sú dve čísla - počiatočné okno a konečné okno v predošlom



Obr. A1.3 Začiatok a koniec slova pre optimálne určený koniec slova

kroku sme si uložili do Buffra (okno) neklipované hodnoty, práve tieto použijeme na výpočet maximálnej a strednej hodnoty vzhľadom k oknu. Označíme ich ako
z týchto hodnôt vypočítame prahy pre presný začiatok a koniec slova

$$\text{mean} = \left(\sum_{i=a}^{b} \text{Buffer}(i) \right) / (b - a)$$

kde a = počiatočné okno, b = konečné okno
max = maximum z Buffer(i)

ak Resolution ≤ 8 potom prah pre začiatok = (max / 2 + mean) / 20
v opačnom prípade prah pre začiatok = (max / 2 + mean) / 10
definitoricky prah pre koniec = 3 * prah pre začiatok

potom počiatočným oknom je prvé okno pre, ktoré plati

Buffer(začiatočné okno) > prah pre začiatok

podobne pre konečné okno, prvé okno od hrubého konca pre, ktoré platí

Buffer(konečné okno) > prah pre koniec

takto na úrovni vzoriek signálu máme definitoricky

počiatočná vzorka = počiatočné okno * prekrytie okien
konečná vzorka = konečné okno * prekrytie okien + prekrytie okien

v poslednom kroku ešte urobíme jeden technický, ale potrebný krok, a síce adjustovanie signálu podľa úrovne (Level). Jednoducho od pôvodného signálu sa odčíta v 1.kroku vyrátaná hodnota Level.

signal → signal - Level

Koniec Algoritmu.

Vidíme, že náš detektor začiatku a konca slov je explicitný - to znamená nezávislý na extrakcii čŕt použitých v procesoch rozpoznávania a verifikácie. Detektor sa skladá z troch stupňov, pričom každá časť klasifikuje dané segmenty signálu. Náš detektor nepredpokladá, že hovorené slovo je prítomné v danom zázname zvuku, nemusí byť prítomné a nemusí byť dopredu známe, o ktoré slovo sa jedná. Samozrejme uvedený algoritmus sa dá implementovať aj pre podmienky reálneho času. Presnosť nášho detektoru sme testovali v režime off-line (databáza slov bola dopredu zaznamenaná bez určenie začiatku a konca slova) a aj pre on-line spracovanie, na telefonickej linke. Úspešnosť detekcie začiatku a konca slova podľa práve spomínaného algoritmu sme porovnávali s ručne (sluchom) zadanými hranicami slov. Takto zadaný etalón je nutne subjektívny (závisí od konkrétnej osoby, ktorá prevádza editovanie signálov), čím vlastne prevádzame porovnanie počítača (automatu) s človekom, hranice určené človekom pokladáme za presné, *Appel, Brandt* (1984).

Pre testovanie v režime off-line sme použili databázu slov DATA, primárne určenú na úlohu rozpoznávania hovoriaceho :

1 ... osem
2 ... žaby
3 ... puto
4 ... kraj
5 ... cena
6 ... voda

DATA je databáza primárne slúžiaca na úlohu verifikácie hovoriaceho. Skladá sa z dátovych súborov typu „dat“, v ktorých sú uložené zdigitalizované signály nahovorených slov. Využívame 26

hovoriacich - A...Z, 6 slov - 1...6 a 8 realizácií z každého slova - 0...7. Vzorkovacia frekvencia bola 10 KHz a zrezávacia frekvencia antialiasing filtra bola 4 KHz, s kompresorom dynamiky, nastaveným na úroveň keď šum celého zariadenia dával najväčšiu hodnotu 17, v rozlíšení +10/-10 V a 12 bitov tj. 17x(10/2048) = 83 mV, pri mikrofóne AIWA nevybudenom, a osciloval okolo 16, čiže dával 3 hodnoty (15, 16, 17). Túto hodnotu (16) treba odrátať od signálových hodnôt.

V nasledujúcej tabuľke Tab. A1.1a sú uvedené výsledky pre práve špecifikovanú databázu. Vidíme, že výsledky pre slová začínajúce na samohlásky, respektíve znelé spoluhlásky sú o niečo lepšie. Celková úspešnosť pre týchto 6 slov je 95,4 percent.

k	**slovo**	**Úspešnosť (%)**
1	osem	98,5
2	žaby	97,7
3	puto	91,3
4	kraj	94,8
5	cena	92,3
6	voda	97,6
Spolu		95,4

Tab. A1.1a Úspešnosti určovania začiatku a konca slova pre slová DATA

BATA je databáza slúžiaca na úlohu verifikácie hovoriaceho po telefóne v režime on-line. Skladá sa z dátových súborov typu „dat“, v ktorých sú uložené zdigitalizované signály nahovorených slov. Využívame 26 hovoriacich - A...Z, 6 slov - 1...8 a 8 realizácií z každého slova - 0...7. Vzorkovacia frekvencia bola 3 KHz a zrezávacia frekvencia antialiasing filtra bola 1 KHz. Pre testovanie v režime on-line sme použili databázu slov BATA :

1 ... cry
2 ... ocean
3 ... patty
4 ... voyage
5 ... eleven
6 ... human
7 ... Danube
8 ... Washington

V nasledujúcej tabuľke Tab. A1.1b sú uvedené výsledky pre práve špecifikovanú databázu. Vidíme, že výsledky pre slová začínajúce na samohlásky, respektíve znelé spoluhlásky sú o niečo lepšie. Celková úspešnosť pre týchto 8 slov je 91,9 percent.

k	**slovo**	**Úspešnosť (%)**
1	cry	91,2
2	ocean	90,6
3	patty	90,5
4	voyage	92,3
5	eleven	93,5
6	human	90,9
7	Danube	92,1
8	Washington	94,7
Spolu		91,9

Tab. A1.1b Úspešnosti určovania začiatku a konca slova pre slová BATA

A2 Homotopická grupa ako grupa pozdĺž trajektórie

V tejto časti prediskutujeme niekoľko pojmov z topológie, ktoré majú vzťah k zobrazeniam medzi dvomi variétami. Celé pojednanie nebude striktne matematické. Tieto pojmy sú dôležité pre pochopenie mechanizmu invariantnej extrakcie čŕt v našich modeloch.

Najskôr definujme zobrazenie f: X ® Y, kde variéty X, Y sú kompaktné n rozmerné, orientované a hranice X a Y sú nulové. Variéta Y je súvislá. Potom existuje celé číslo a toto číslo sa nazýva stupeň zobrazenia f. V určitom elementárnom zmysle toto číslo opisuje koľkokrát Y pokrýva X pomocou zobrazenia f. V našom prípade máme dve reálne funkcie x(t), y(t) z priestoru R^1 . Definujeme trajektóriu v komplexnej rovine podľa

$$z(t) = d(t)\,[x(t) + i\,y(t)] \qquad \text{(A2.1)}$$

kde parameter d, obecne závislý na čase sa nazýva parameter deformácie trajektórie. Nateraz uvažujme parameter d ako nezávilý od času. V tomto prípade označujeme takúto deformáciu ako globálnu na rozdiel od lokálnej deformácie, kde d je závislé od času. Ak trajektória nepretína počiatok súradnicového systému, potom môžeme definovať fázu trajektórie Φ(t). Platí

$$\varphi(t_a) = \varphi_a,\ \ \varphi(t_b) = \varphi_b,\quad t_a \le t \le t_b$$

Trajektória môže byť preseknutá sama sebou, čo znamená že

$$z(t_1,d) = z(t_2,d) \qquad t_1 \neq t_2$$

Trajektória sa nazýva uzavretou ak platí

$$z(t_a,d) = z(t_b,d) \qquad \text{(A2.2)}$$

Pretože z nie je rovné nule pre $t_a \le t \le t_b$, potom pre každý čas t môžeme definovať fázu φ(t), ako sme už spomínali predtým, ale s nejednoznačnosťou $2k\pi$. Aby sme sa vyhli tejto nejednoznačnosti môžeme zafixovať počiatočnú fázu $\varphi(t_a)$. Fáza φ(t) je spojitou funkciou času. Preto môžeme definovať rozdiel fáz ako

$$\Delta\varphi = \varphi(t_a) - \varphi(t_b) = 2k\pi \qquad \text{(A2.3)}$$

Celé číslo k nazývame stupeň trajektórie z(t,d). Fyzikálny význam stupňa trajektórie je jasný z nasledujúcich úvah a z Obr. A2.1. Uvažujme najskôr elementárny prípad uzavretej trajektórie

$$z(t,d) = c + d(\cos t + i \sin t), \qquad \text{kde } 0 \le t \le 2\pi \qquad \text{(A2.4)}$$

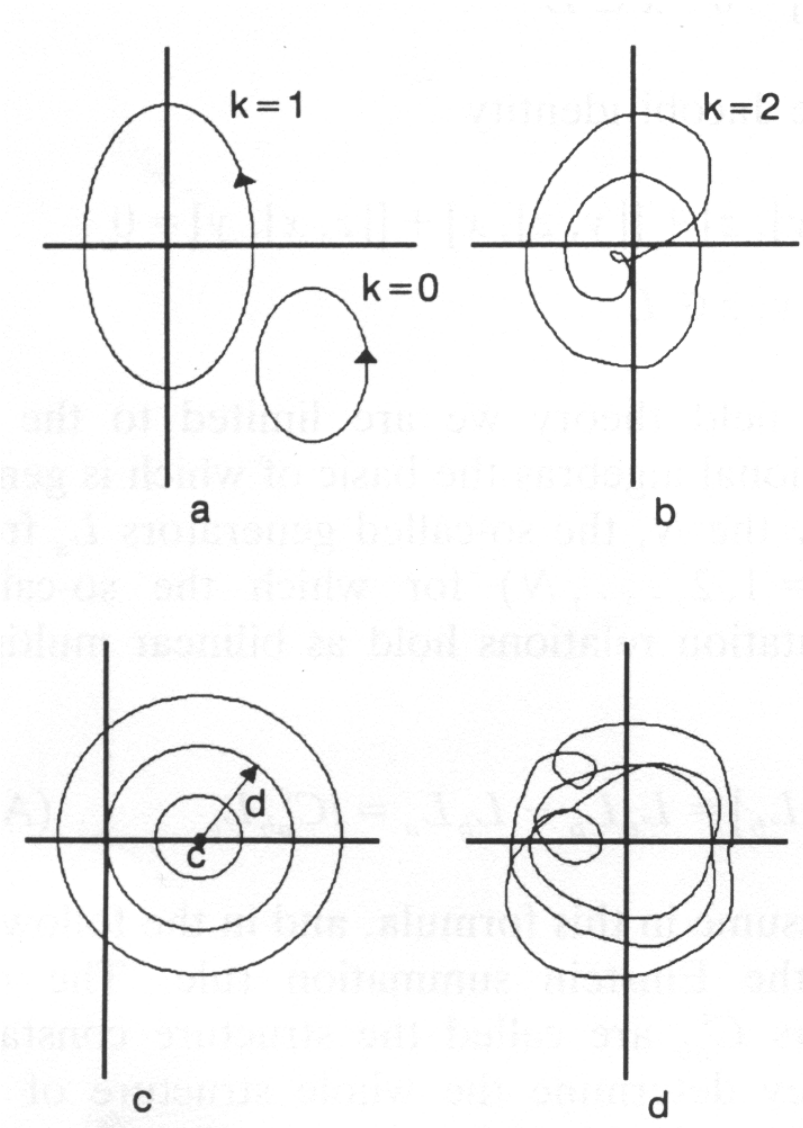
Vidíme, že trajektória je kružnica so stredom v bode c a s polomerom d.

Ak d < c, potom počiatok súradnicového systému je mimo tejto kružnice a stupeň je nulový.

Ak d = c, potom trajektória pretína počiatok súradnicového systému a stupeň nie je definovaný.

Ak d > c, potom stupeň trajektórie je rovný jeden.

Ak sa parameter d mení, potom hovoríme, že trajektória (A2.1) alebo (A2.4) je deformovaná. Všimnime si, že stupeň sa nemení, kvôli celočíselnej povahe stupňa, je invariantný pri spojitých



Obr. A2.1 Trajektórie a stupne trajektórie ako topologické invarianty

deformáciach trajektórie. Pravdaže pri deformáciach trajektórie sa nesmie pretnúť počiatok súradnicového systému, alebo bod kde fáza trajektórie sa nedá definovať.

Ak nasledujeme elementárne geometrické úvahy, potom môžeme vyjadriť rozdiel fáz $\Delta\varphi$ definovaný v (A2.3) v integrálnej forme ako

$$\frac{\Delta\varphi}{2\pi} = \oint_C d(arctg[y'(t)/x'(t)]) = \int_L \boldsymbol{A}(t)dl \qquad \text{(A2.4)}$$

kde

$$A(t) = [y'(t)x''(t) - y''(t)x'(t)] / [x'^2(t)] + y'^2(t)]^{3/2}$$
$$dl = [x'^2(t)] + y'^2(t)]^{1/2}$$

a

$x' = dx / dt$, $y' = dy / dt$

Z týchto úvah môžeme usúdiť, že celá množina trajektórií sa dá klasifikovať do rôznych tried podľa ich stupňa trajektórie. Tieto triedy majú rôzne topologické vlastnosti. V našom modeli percepcie sa zaoberáme s kvázi-topologickými vlastnosťami, ktoré sú limitované na veľké stupne, pretože neuvažujeme presne podmienky uzavretia trajektórie, alebo vo Fourier obraze, kde podmienka uzavretosti je splnená z definície.

Všetky pojmy, o ktorých sme sa zmienili, sa môžu preformulovať presne pomocou pojmov množiny tried ekvivalencie trajektórií, homotopických tried alebo homotopickej invariancie *Hirsch* (1976), *Massey, Stallings* (1967). V prípade uzavretých trajektórií má množina tried ekvivalencie grupové vlastnosti. Táto grupa sa nazýva fundamentálnou grupou alebo Poincaré grupou variéty Y v jej bode y - p(Y,y). Formulácia fundamentálnej teorémy, ktorá opisuje náš prípad, môže byť nasledovná:

Ak $f: X \rightarrow Y$ je homotopická ekvivalencia (alebo homotopické zobrazenie), potom pre ľubovolný bod x z X homomorfizmus $f^*: p(X,x) \rightarrow p(Y,f(x))$ je izomorfizmus. Znovu, exaktné definície pojmov, ktoré sme tu spomínali sú uvedené v *Massey, Stallings* (1967) a *Adámek, Koubek, Reiterman* (1977). Z týchto dôvodov nazývame stupeň trajektórie homotopickým invariantom nášho zobrazenia. Musíme ale zdôrazniť, že stupeň zobrazenia, nie je jediným homotopickým invariantom daného zobrazenia. Stupeň zobrazenia je jediným homotopickým invariantom v prípade zobrazenia n-rozmernej, kompaktnej, súvislej variéty do S^n, kde S^n je n-rozmerná jednotková guľa. Všetko toto sa dá tiež preformulovať pre ľubovolný n-rozmerný topologický priestor. V tomto prípade nazývame grupu homotopickou grupou a nie Poincaré grupou.

Z hľadiska nášho prístupu k percepcii reči, môžeme povedať, že naša neurónová sieť realizuje homotopické zobrazenie f(x). Zobrazenie f závisí na x alebo inými slovami na stave neurónovej siete a tiež na čase t. Je to lokálne v čase invariantné zobrazenie.

Všetko toto sa dá lepšie vidieť z explicitného tvaru siete a z výsledkov simulácie, ktoré uvádzame v kapitole venovanej výsledkom a diskusii.

A3 Kalibračná grupa ako grupa pozdĺž trajektórie

V tejto časti uvedieme niektoré predbežné myšlienky a návrhy na priame zavedenie vnútorných symetrií do modelov ľudskej percepcie reči a rozpoznávania. Z hľadiska toho, že táto práca nie je určená špeciálne fyzikom, budeme sa niektorým otázkam venovať podrobnejšie, ale bez veľkého dôrazu na matematickú rigoróznosť a exaktnosť formalizmu. Ako sme uviedli v predošlej časti tieto symetrie sa odrážajú v symetriách fonémického systému a veríme, že vyjadrujú jeden z najdôležitejších rozdielov medzi ľudskou a papagájovou percepciou a rozpoznávaním a sú vôbec dôležitým krokom k pochopeniu percepcie reči ako takej.

Náš prístup bol silne inšpirovaný formalizmom kalibračných teórií poľa, kde princípy symetrie sú inherentné a prejavujú sa na rôznych stupňoch a vo veľkom rozsahu.

Uvažujme 4 - rozmerný priestor X, tento priestor je generovaný 4 - vektormi x = (ct, x). Ďalej prejdeme do systému jednotiek s c = 1. Postulujeme existenciu zobrazenia

$$f: X \rightarrow \Psi, \Psi = f(x) \qquad \text{(A3.1)}$$

kde X je 4 - rozmerný vektorový priestor a Ψ je 4 - rozmerný vektorový priestor vzorov akustického poľa. Inými slovami, postulujeme, že pre každý bod z X existuje práve jeden bod y z Ψ. Ak uvažujeme zvukový vzor ako fyzikálnu veličinu s nejakými vlastnosťami invariancie pri patričných transformáciach grúp, potom sa môžeme pokúsiť zaviesť nejakä obecnú teóriu identifikácie týchto vzorov. Fyzikálne relevantnou veličinou v našom prípade môže byť napríklad 4 - vektor akustickej intenzity akustického poľa, ktorý pôsobí na bazilárnu membránu alebo mikrofónnu membránu, *Davis* (1983), *Ashmore* (1990), *Nobili, Mammano, Ashmore* (1998). Čiže, presnejšie, je dobre chápať X ako konfiguračný priestor, a nie len ako euklidovský, alebo minkovského priestor. Podobne ako pre vlnovú funkciu v kvantovej mechanike $\Psi(x)$, kde x je priestorová súradnica (spolu s časom) len pre jednu časticu, pre viac častíc je to konfiguračný priestor. Inými slovami predpokladáme, že x je $(x_1, x_2, ..., x_n, p_1, p_2, ..., p_n)$, kde p sú časové derivácie x, $p = \delta x/\delta t$. Samozrejme môžeme to označiť aj ako $(x_1, x_2, ..., x_{2n})$

Takto môžeme postulovať existenciu systému polí vzorov v neurónovom systéme, v mozgu, pre ktorý každý bod z priestoru X alebo Ψ je vyjadrený vektorovým poľom $\Psi(x)$. Obraz $\Psi(x)$ zvukového vzoru x (obecne aj zrakový, čuchový, chuťový, hmatový vzor) odpovedá napr. vlnovej funkcii elementárnej častice alebo systému častíc. Všeobecne grupy tranformácie indukujú potenciál spomínaného vzoru, ktorý odpovedá napr. potenciálu elektromagnetického poľa. Podobne ako v elektrodynamike a v diferenciálnej geometrii, ak rozdiel medzi dvomi poľami Ψ - obrazmi vzorov, nie je nulový pri vhodnej transformácii v konfiguračnom priestore, potom vzniká (v mozgu) toto pole. Ako ukážeme na záver tejto časti, rozdiel v informácii ktorú nesú spomínané dve polia sa dá vyjadriť nejakým poľom, jeho excitáciami (v mozgu). Postulujeme, že systém polí je kalibračne invariantný a zložky $\Psi(x)$ sú zoskupené do multipletov, ktoré sa transformujú podľa danej ireducibilnej transformácie z Lieho grupy G. Práve takýmto spôsobom vyjadríme vnútorné symetrie pričom predpokladáme, že práve tieto symetrie sú prítomné v našom fonémickom systéme.

Teraz sa stručne venujeme pojmom z Lieho grúp. Ako prvé zavedieme pojem Lieho algebry. Takže, nech je F komutatívne pole. Lie algebra nad F je vektorový priestor L nad F kde platí bilineárné násobenie v tvare

$$[\ ,\ ] : L \times L \rightarrow L$$

ktoré vyhovuje vzťahom

$$[\ x, x] = 0 \text{ pre } x \in L$$

a Jacobiho identite

$$[\,[\,x, y], z] + [\,[\,x, z], y] + [\,[\,z, x], y] = 0 \quad \text{pre } x, y, z \in L$$

V teórii poľa sme limitovaní na N rozmerné algebry, ktorých báza je generovaná pomocou N takzvaných generátorov L_a z L, (a = 1, 2, ..., N) a pre ktoré platia komutatívne vzťahy ako sú bilineárne násobenia

$$[\,L_a, L_b] = L_a L_b - L_b L_a = iC^c{}_{ab} L_c \qquad \text{(A3.2)}$$

V tejto a nasledujúcich formulách predpokladáme, že platí Einsteinovo sumačné pravidlo. Reálne čísla $C^c{}_{ab}$ sa nazývajú štruktúrne konštanty a určujú celú štruktúru Lie algebry L. Teraz prvok spojitej grupy G pre danú algebru opísanú pomocou [,] je definovaný ako

$$U = \exp(-i\omega_a L_a) \qquad \text{(A3.3)}$$

kde $\omega_a(x)$ sú nejaké reálne čísla. Netriviálny výraz [,] a jeho grupové vlastnosti vyplývajú z teorémy Campbell - Hausdorfa, *Adams, J. F.* (1969). Ak ľubovolná zo štruktúrnych konštánt je rôzna od nuly, potom grupa G nie je abelovská. V prípade N = 3 máme SU(2) grupu, ktorá je neábelovskou grupou najnižšieho rozmeru. Ak opíšeme generátory v maticovej forme hovoríme o maticovej reprezentácii grupy G. V teórii poľa sa zvyčajne ohraničujeme na unitárne konečno rozmerné matice, *De Witt, C., De Witt, B. eds* (1964).

Teraz stručne prediskutujeme kalibračné transformácie. Postulujeme existenciu „ lagrangeovej hustoty“ L_0 systému obrazov, alebo neurónových vzorov. Fyzikálne-informačný význam tohto lagranžiánu prediskutujeme neskôr. Takto máme

$$L_0 = L_0 (\Psi, \partial^\alpha \Psi)$$

kde $\partial^\alpha Y = \partial \Psi / \partial x_\alpha$, a kalibračná invariancia znamená, že

$$L_0(U\Psi, \partial^\alpha U\Psi) = L_0(\Psi, \partial^\alpha \Psi) \qquad \text{(A3.4)}$$

kde

$$U(x) = \exp(-i\omega(x))$$

je prvok Lie-ho grupy a $\omega(x)$ je prvok Lie-ho algebry, s ktorou pracujeme. V tomto prípade hovoríme o lokálnej kalibračnej invariancii. Ak U nezávisí od x hovoríme o globálnej kalibračnej invariancii.

Potom

$$\Psi(x) \rightarrow U(x)\, \Psi(x),$$

$$\partial^\alpha \Psi(x) \text{ ® } U(x)\, (\partial^\alpha \Psi(x)) + (\partial^\alpha U(x))\, \Psi(x)$$

a pre infinitezimálne transformácie máme

$$\partial \Psi(x) = -i\omega(x)\, \Psi(x),$$

$$\partial\, [\, \partial^\alpha \Psi(x)\,] = -i\omega(x)\, \partial^\alpha \Psi(x) - i[\, \partial^\alpha \omega(x)]\, \Psi(x)$$

Definujeme kovariantnú deriváciu ako

$$D^{\alpha}\Psi(x) = [\,\partial^{\alpha} + igA^{\alpha}(x)]\,\Psi(x) \qquad \text{(A3.5)}$$

kde $A^{\alpha}(x)$, prvky Lie-ho algebry, sú definované ako

$$A^{a}(x) = A^{\alpha}{}_{a}(x)\,L_{a}$$

Ak požadujeme, aby sa kovariantná derivácia transformovala ako $\Psi(x)$ pri lokálnych kalibračných transformáciach, potom polia $A^{\alpha}(x)$ sa musia trasformovať ako

$$\partial A^{\alpha}(x) = (1/g)\partial^{\alpha}\omega(x) - i[\omega(x), A^{\alpha}(x)]$$

alebo

$$\partial A^{\alpha}(x) = (1/g)\partial^{\alpha}\omega(x) + C_{\alpha bc}\,\omega_{b}(x)\,A^{\alpha}{}_{c}(x) \qquad \text{(A3.6)}$$

Takýmto spôsobom máme definované N kalibračných polí, takzvané „Yang-Mills" polia. Veľmi ľahko sa dá potom ukázať, že ak nahradíme derivácie v (4.6.4) kovariantnými deriváciami stane sa Lagranžián invariantným relatívne k lokálnym kalibračným transformáciam, *Abers, Lee* (1973),

$$L_{0}(U\Psi, D^{\alpha}U\Psi) = L_{0}(\Psi, D^{\alpha}\Psi)$$

V kalibračnej teórii sa polia $A^{\alpha}(x)$ sa interpretujú ako dynamické premenné. Preto zavedieme hustotu voľného lagranžiánu kalibračných polí ako

$$(-1/4)\,F^{\alpha\beta}{}_{a}F_{a\alpha\beta}$$

kde

$$F^{\alpha\beta}{}_{a} = \partial^{\alpha}A^{\beta}{}_{a}(x) - \partial^{\beta}A^{\alpha}{}_{a}(x) - gC_{abc}A^{\alpha}{}_{b}(x)A^{\beta}{}_{c}(x)$$

Konštanta úmernosti g sa interpretuje ako väzbová konštanta, spolu s C_{abc} vyjadruje relatívnu váhu konkrétneho poľa. Takto nakoniec máme lokálne kalibračne invariantnú a Lorentzovsky invariantnú hustotu lagranžiánu

$$L = (-1/4)\,F^{\alpha\beta}{}_{a}F_{a\alpha\beta} \quad + L_{0}(Y, D^{\alpha}\Psi) \qquad \text{(A3.7)}$$

Teraz, podobne ako v časti o homotopických grupách, vyšetríme variácie polí Y pozdĺž nejakých trajektórií vo vonkajšom koordinačnom priestore X alebo Y. Ak máme kalibračné polia (A3.6), potom v každom bode X existuje nezávislý výber orientácie koordinačného systému vo vnútornom priestore polí vzorov. Kalibračné transformácie A^{α} sú korelované so zmenou orientácie lokálneho vnútorného systému vzhľadom k prenosu z bodu x do bodu x + dx. Pole $\Psi(x)$ sa lokálne nelíši od poľa $U(x)\,\Psi(x)$ a kalibračné transformácie $A^{\alpha}(x)$ tiež nemenia vlastnosti systému, o ktorom uvažujeme. Pre „Yang - Mills-ove" polia, *Wu, Yang* (1975), požadujeme, aby takzvané paralelné prenosy boli nulové

$$dx_{\alpha}D^{\alpha}\Psi(x) = dx_{\alpha}[\,\partial^{\alpha} + igA^{\alpha}(x)]\,\Psi(x) = 0$$

Predpokladajme, že $\Psi(x)$ je paralelne prenesený pozdĺž trajektórie P, ktorá je parametrizovaná

reálnym parametrom τ z intervalu $\langle 0, 1 \rangle$. Potom

$$x = x(t), \qquad A^{\alpha} = A^{\alpha}(x(t)), \qquad \Psi = \Psi(x(t))$$

$$(dx_{\alpha} / dt)[\partial^{\alpha} + igA^{\alpha}(t)]\,\Psi(x) = 0$$

s riešením

$$\Psi(t) = T[\exp\{-ig \int_0^{\tau} dt'(dx_{\alpha} / dt')\, A^{\alpha}(t)\}]\,\Psi(0),$$

kde T je operátor „časového usporiadania“ definovaný ako

$$T[a(t)\, b(t')] = a(t)\, b(t')\, Q(t - t') \pm b(t')a(t)\, Q(t' - t)$$

Horný znak platí pre bozónové polia a dolný znak pre fermiónové polia. Potom môžeme definovať, pre každú trajektóriu P, maticový operátor

$$W(P) = T \exp\{-ig \int_P dx_a\, A^a(t)\}$$

Fyzikálny význam tohto operátora je nasledovný; prevádza kalibračnú transformáciu Y(x(0)) do $\Psi(x(1))$ pozdĺž trajektórie

$$\Psi(x(1)) = \Omega(P)\,\Psi(x(0)) \qquad \text{(A3.8)}$$

kde x(0) je počiatočný a x(1) je konečný bod trajektórie P.

Teraz dokážeme, že pre uzavretú trajektóriu C platí, že Spur $[\Omega(C)]$ je kalibračne invariantná veličina. Pre kalibračnú transformáciu ľubovolného Ψ máme

$$\Psi'(x) = U(x)\,\Psi(x)$$

a z (4.6.7)

$$\Psi'(x(1)) = \Omega'(P)\,\Psi'(x(0)),$$

čo vedie k

$$U(x(1))\Psi(x(1)) = \Omega'(P)\, U(x(0))\Psi(x(0)),$$

$$U(x(1))\Omega(P)Y(x(1)) = \Omega'(P)\, U(x(0))Y?(x(0)),$$

a nakoniec máme

$$\Omega'(P) = U(x(1))\Omega(P)U^{-1}(x(0))$$

Z toho priamo vyplýva, že

Spur [Ω(C)] = kalibračne invariantný (A3.9)

Najdôležitejšou vlastnosťou Ω(C) je nasledovná. Ak Ω(C) je známe pre ľubovolnú uzavretú trajektóriu C, potom v Ω(C) je skoncentrovaná celá fyzikálna informácia o kalibračných poliach, a čo je veľmi dôležité, všetká redundantná informácia je mimo, *Ryder* (1985), *Masahiro* (1988).

Teraz sa môžeme vrátiť k fyzikálno - informačnej interpretácii „lagranžovej hustoty" L danej vzťahom (A.3.7). Nasledujúc predošlé úvahy a *Ryder* (1985) môžeme postulovať, že pole vzorov Ψ sa „percepuje" alebo vyhodnocuje podľa princípu najmenšej akcie

variácia (L) = 0.

Tento princíp je potrebný na to, aby sme kompresovali rýchlosť informácie, inými slovami maximalizovali úspešnosť klasifikácie daného vzoru pomocou minimalizácie excitácií poľa vzorov. Vzor pripravený takýmto spôsobom bude topologicky najviac podobný zapamätanému prototypu. Z (A3.7) môžeme usúdiť, že polia vzorov (alebo neurónov) sú excitované iba v tom prípade ak dva vzory akustického poľa sa nedajú transformovať prostredníctvom danej kalibračnej transformácie.

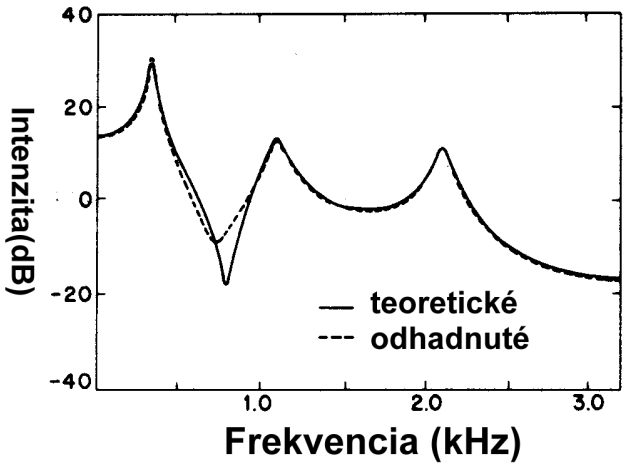
A4 Topologické invarianty vo frekvenčnej oblasti a LPC

LPC techniky sa za posledné roky považujú za výkonné nástroje pre analýzu reči. Používajú sa pre také algoritmy ako je odhad základného tónu, formantov reči, tvaru a parametrov vokálneho traktu, pre odhad cepstrálnych koeficientov alebo na bitovú kompresiu rečového signálu, *Markel, Gray* (1976), *Makhoul* (1975), *Makhoul* (1975). Reč sa modeluje ako výstup lineárneho, časovo premenného systému, ktorý sa budí buď periodickým signálom alebo šumom, Obr. 2 v kapitole 2. Je zvykom nazývať tieto dva zdroje signálu znelým (v reálnom trakte odpovedá kmitaniu hlasiviek) a neznelým (odpovedá to neprítomnosti kmitania hlasiviek) zdrojom.

Prenosovú funkciu 1 / H(z) (3) môžeme pokladať za vyhladzovací filter pre rečový signál. Frekvenčná odozva tohto filtra sa dá vypočítať pomocou Fourier transformácie, špeciálne diskrétnej Fourier transformácie, DFT, ktorú definujeme ako

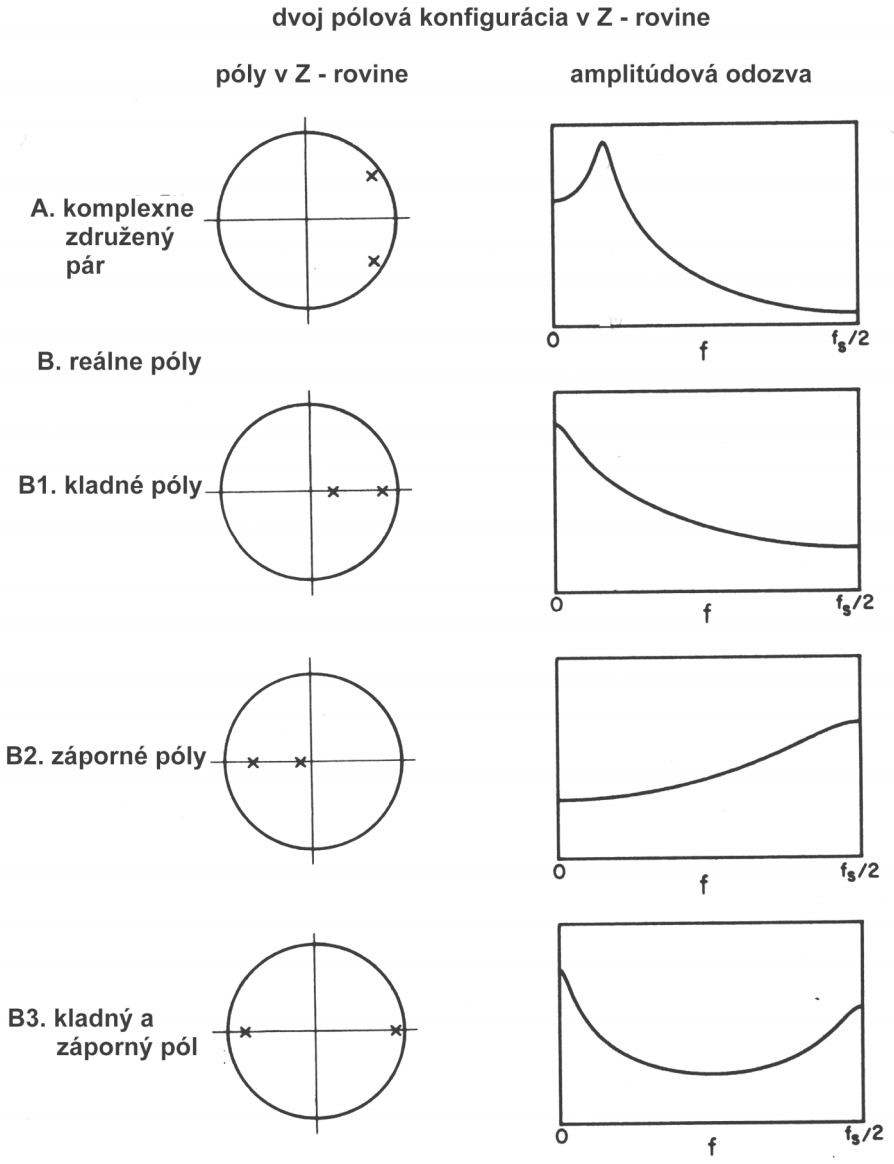
$$X_k = \sum_{n=0}^{N-1} x(n) \exp\{-i(2\pi nk / N)\} \qquad \text{(A4.3)}$$

kde x(n) je obecne komplexný signál, X_k sú komplexné Fourier komponenty, na Obr. A4.1 je znázornené spektrum vypočítané predošlou metódou porovnané s pôvodným spektrom. Z viacerých vlastností Digitálnej Fourier Transformácie - DFT, ako je linearita, symetria pri časovom a frekvenčnom posuve, je pre náš prístup podstatná vlastnosť o symetrii pre reálny signál, inými slovami ak je x(n) reálnou funkciou, potom sú príslušné DFT transformanty symetrické ohľadom reálnej osi, *Čížek* (1981), *Oppenheim, Shafer* (1975). Ak vynesieme do grafu v komplexnej rovine reálnu a imaginárnu zložku X_k, pre vstupný reálny signál x(n) = {a_1, a_2,... a_p, 0,0, ...0}, kde a_i sú LPC koeficienty, dostaneme graf, ktorý je triviálne symetrický vzhľadom k reálnej osi, čím je zároveň splnená podmienka uzavretia trajektórie pre topologické invarianty.

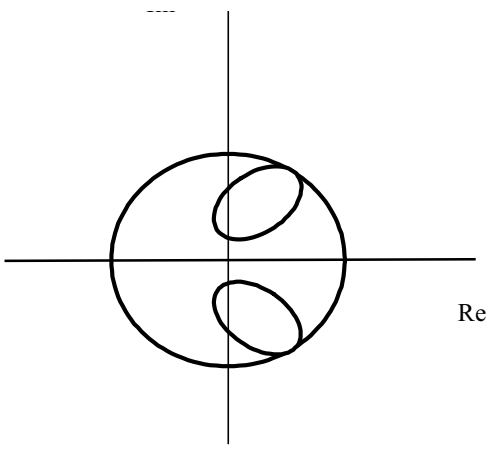


Obr. A4.1 Vyhladenie spektra pomocou LPC

Súvislosť rozmiestnenia pólov nášho modelu, alebo koreňov prechodovej funkcie ilustrujeme pre najjednoduchší prípad na Obr. A4.2a. Zvolili sme si LPC druhého rádu a zostrojili všetky 4 možné situácie a im prislúchajúce spektrá, jasne demonštrujúc vyhladzovací efekt LPC techniky. Podobne si môžeme predstaviť aj odpovedajúce grafy reálnej a imaginárnej zložky, vo fázovej rovine, pre náš prípad dostaneme graf s jedným závitom na hornej polovici a symetricky umiestneným na dolnej, pozri Obr. A4.2b.

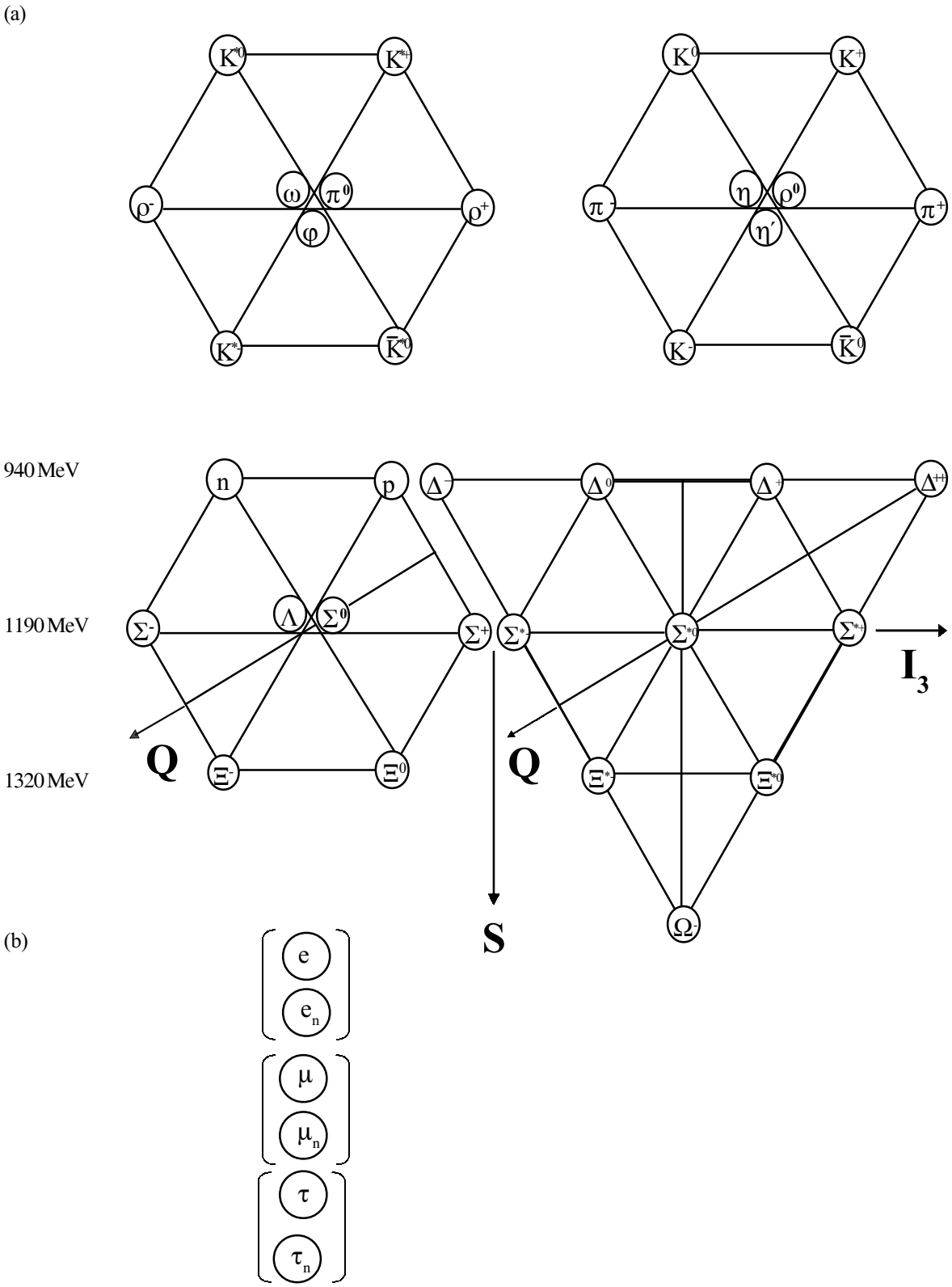


Obr. A4.2a Rozmiestnenie pólov LPC modelu a tvar spektra



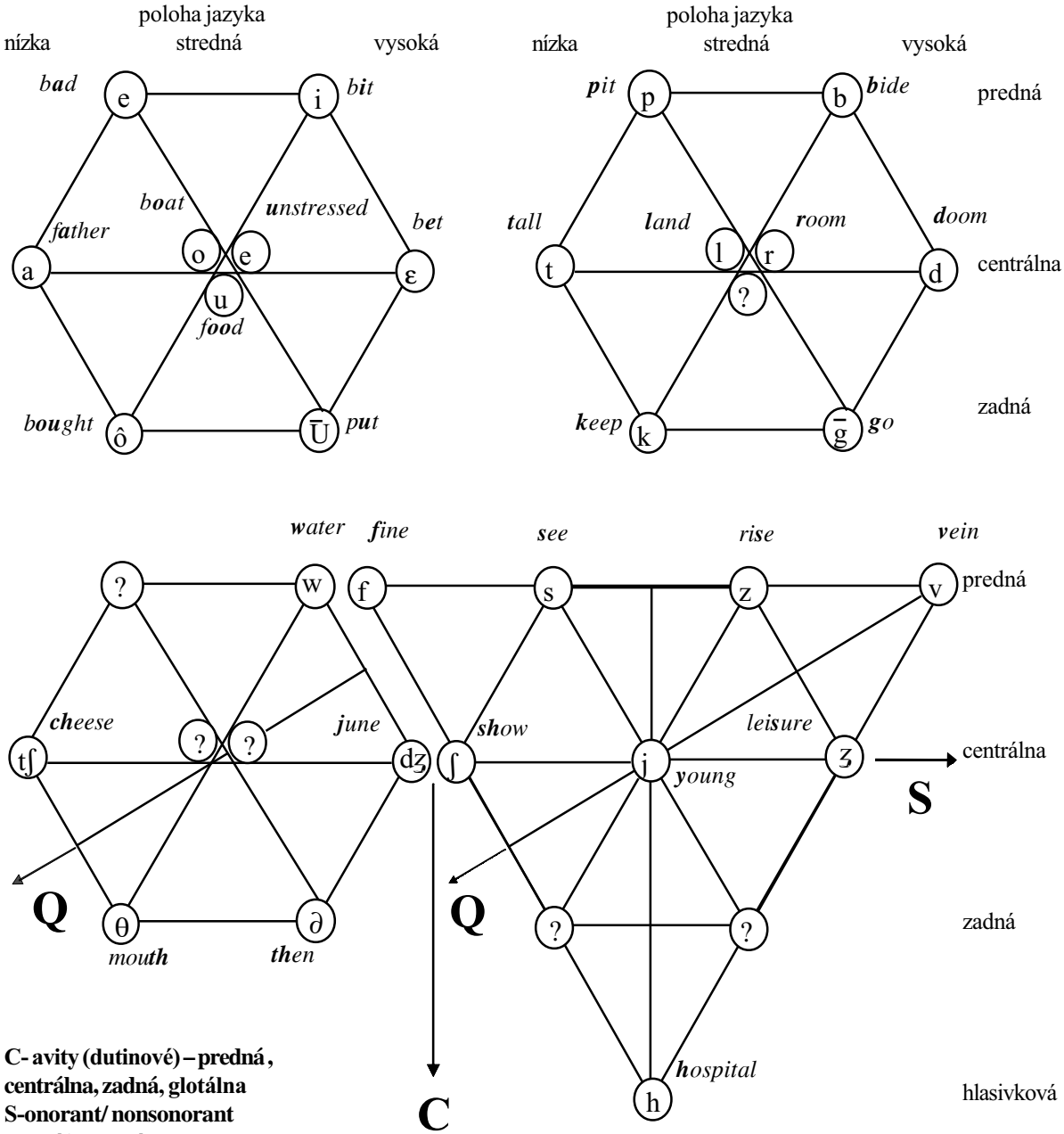
Obr. A4.2b Nyquistov graf vo frekvenčnej oblasti, podľa LPC modelu 2 rádu

A5 Systém symetrie elementárnych častíc



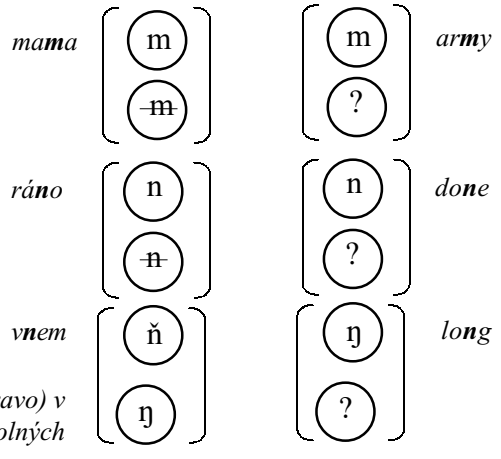
Obr.A5.1 (a) Multipletová štruktúra elementárnych častíc - hadróny, vľavo je udané z ilustračných dôvodov hmotnostné spektrum pre baryónový oktet. ***Q*** *označuje elektrický náboj,* $\boldsymbol{I_3}$ *označuje 3-tiu zložku izospinu a* ***S*** *označuje spin. Existuje vzťah, nazývaný Gell-Mann, –Nishijima, medzi elektrickým nábojom,izospinom a hypernábojom* ***Y****:* $\boldsymbol{Q = I_3 + Y/2}$*. (b) Tri dublety leptónov. Druhí protihráči- neutrína nemajú hmotnosť alebo iba sa predpokladá, že veľmi malú*

A6 Systém symetrie anglických foném



C- avity (dutinové) – predná, centrálna, zadná, glotálna
S-onorant/ nonsonorant (znelé/neznelé)
Q ? (aký je význam)

Obr. A6.1 Multiplety foném angličtiny



Obr. A6.2 Tri dublety nosových foném vľavo) v slovenčine vpravo) v angličtine. V slovenčine sú fonémy iba v horných riadkoch, v dolných sú alofóny-hlásky- nemajú sémantický význam

A7 Rýchla topologická transformácia - funkcia FTT

```
/
*************************************************************************/
** FUNCTION: FTT()
** PURPOSE: Fast Topological Transformation for data x and y
** RETURN: dsp_error=-101 if array size is less than 1
** REMARKS: num_features is radix 2
*************************************************************************/
void FTT(
  DSPDBL *x,            /*** I; x-axis array ***/
  DSPDBL *y,            /*** I; y-axis array ***/
  DSPINT winlen,        /*** I; size of window ***/
  DSPINT num_features,  /*** I; number of axes in plot ***/
  DSPSHR *tw,           /*** O; short; number of axis intersections ***/
  DSPSHR *cw)           /*** O; short; duration in given section ***/
{
  extern DSPINT dsp_error;
  DSPDBL *endwin;
  register DSPINT ind1, ind2;
  register DSPINT index1, index2;
  register DSPINT fea, numfeam2;
  DSPDBL an1, an2, anb;

  dsp_error=DSP_OK;
  if (winlen <= 0) { dsp_error=DSP_ERROR_ARRAY_SIZE_MUST_GT_0; return;}

  for (fea=0; fea < num_features; fea++)
     tw[fea]=cw[fea] = 0;

  numfeam2=num_features*2;
  anb=DSP_PI/(DSPDBL)num_features;
  index2=0;
  endwin=x+winlen;

  if (fabs(*y) > DBL_EPSILON)                    /* computing angle1 and its
index1 */
            index1=(DSPINT)((an1=(atan2(-*y++, -*x++)+DSP_PI))/anb);
  else if (fabs(*x) > DBL_EPSILON) {
      y++;
      if (*x++ > 0.0) {
           an1=0.0;
           index1=0;
       }
      else {
           an1=DSP_PI;
           index1=num_features;
       }
   }
```

```
 else {
        an1=0.0;
        index1=0;
        x++;
        y++;
  }
 while (x < endwin) {
    cw[((index1+1) & (~numfeam2)) >> 1]++;
     ind1=index1 >> 1;
    if (fabs(*y) > DBL_EPSILON)
         index2=(DSPINT)((an2=(atan2(-*y++, -*x++)+DSP_PI))/anb);
    else if (fabs(*x) > DBL_EPSILON) {
                y++;
          if (*x++ > 0.0) {
              an2=0.0;
             index2=0;
           }
          else {
              an2=DSP_PI;
            index2=num_features;
           }
      }
     else {
          an2=0.0;
          index2=0;
                x++;
          y++;
      }
    ind2=index2 >> 1;
    if (index1 < index2)     {
          if (index2-index1 < num_features)
            while (ind1 != ind2)
                  tw[++ind1]++;
          else if (index2-index1 > num_features) {
             while (ind1 >= 0)
                  tw[ind1--]++;
                        while (++ind2 < num_features)
                  tw[ind2]++;
           }
          else if (an2-an1 < DSP_PI)
            while (ind1 != ind2)
                  tw[++ind1]++;
          else {
             while (ind1 >= 0)
                  tw[ind1--]++;
            while (++ind2 < num_features)
                  tw[ind2]++;
           }
      }
```

```
            else
        if (index1-index2 < num_features)
          while (ind2 != ind1)
              tw[++ind2]++;
        else if (index1-index2 > num_features) {
          while (ind2 >= 0)
              tw[ind2--]++;
          while (++ind1 < num_features)
              tw[ind1]++;
         }
        else if (an1-an2 < DSP_PI)
          while (ind2 != ind1)
              tw[++ind2]++;
            else {
          while (ind2 >= 0)
              tw[ind2--]++;
            while (++ind1 < num_features)
              tw[ind1]++;
        }
  an1=an2;
  index1=index2;
   }
  cw[((index2+1) & (~numfeam2)) >> 1]++;
}
```

Literatúra

Abers, E. S., Lee, B. W. (1973). Physics Reports, 9, 1.

Adams, J. F. (1969). Lectures on Lie groups. W. A. Benjamin, Inc.

Adámek, J., Koubek, V., Reiterman, J. (1977). Základy obecné topologie. SNTL Praha.

Ahalt, S. C. , KrishnaMurthy, A. K., Chen, P., Melton, D. E. (1990). Competetive Learning Algorithms for Vector Quantization. Neural Networks, 3, 277-290.

Allen, J. B., Rabiner, L. R. (1977). A unified approach to short-time Fourier analysis and synthesis, Proc. IEEE, 65, 1558-1564.

Anderson, J.A. (1983). Cognitive and Psychological Computation with Neural Models, IEEE on SMC, 13.

Antoniou, A. (1983). Digital Filters - Analysis and Design. McGraw-Hill Book Company.

Appel, U., Brandt, A. (1984). A Comparative Study of Three Sequential Time Series Segmentation Algorithms. Signal Processing, 6, 45-60.

Artieres, T., Bennanni,Y., Gallinari,P., Montacie,C. (1991). Connectionist and conventional models for free-text talker identification tasks. Neuro-Nimes, France.

Ashmore, J. F. (1990). A Fast Motile Response in Guinea-pig Outer Hair Cells; the Cellular Basis of the Cochlear Amplifier. J. Physiol., 383, 323-347.

Atal, B. S., Rabiner L. R. (1976). A Pattern Recognition Approach to Voiced-Unvoiced-Silence Classification with Applications to Speech Recognition. IEEE on ASS, 14.

Atmanspacher, H. (1988). MPE Preprint, 123.

Bennanni,Y., Fogelman,F., Gallinari, P. (1990). A connectionist approach for speaker identification. Proc. Internat. Conf. Acoust. Speech Signal process. '90, S5.1, Albuquerque, NM, 265-268.

Bennanni,Y., Gallinari, P. (1992). Task decomposition through a modular connectionist architecture: A talker identification system. IEEE ICANN.

Bennanni,Y., Gallinari, P. (1995). Neural networks for discrimination and modelization of speakers. Speech Communication 17, 159-175.

Bhagavantam, S., Venkatarayudu, T. (1951). Theory of Groups and its Application to Physical Problems, S. E. Andhra University Waltair.

Bridle, J.S. (1990). Alpha-nets: A reccurent 'neural' network architecture with a Hidden Markov Model interpretation. Speech Communication, 9, 83-92.

Carpenter, G. A., Grossberg, S. (1987). A Masively Parallel Architecture for a Self-organizing Neural Pattern Recognition Machine. Computer Vision, Graphics and Image Processing, 37, 54-115.

Casar M., Fonllosa J. (2007). Double layer architectures for automatic speech recognition using HMM, in book Robust Speech recognition and understanding, I-Tech education and publishing, ISBN 978-3-902613-08-0, Croatia.

Connine, C. M., Titone, D. (1996). Phoneme Monitoring. Language and Cognitive Processes, 11, 635–645.

Cooley J. W., Cochran, W. T. (1967). What is the Fast Fourier Transform ? IEEE on AEA, 15.

Cristianini, N., Shawe-Taylor, J. (2000). An introduction to support Vector Machines: and other kernel-based learning methods. Cambridge University Press, New York, NY, USA.

Čížek, V. (1981). Diskrétní Fourierova transformace a její použití. SNTL Praha.

Davis, H. (1983). An Active Process in cochlear mechanics. Hearing Res, 9, 79-90.

Demuynck, K., Duchateau, J., Compernolle, D., Wambacq, P. (1998). Improved feature decorrelation for HMM based speech recognition, International Conference on Spoken Language Processing (ICSLP), Sydney.

Deng, L. (1992). Processing of Acoustics Signals in a Cochlear Model Incorporating Laterally Coupled Suppressive Elements. Neural Networks, 5, 19-34.
De Witt, C., De Witt, B. ed. (1964). Relativity, Groups and Topology. Willey.
Dialogic Corporation. (1992). DIALOGIC Hardware and Software Manual.
Diehl, R. L. (2004). Speech perception. Ann. Rev. Psych. 55, 149–79.
Ditzinger, T., Haken, H. (1989). Oscillations in the Perception of Ambigous Patterns. Biol. Cyber. 61, 279-287.
Dodwell, P. C. (1983). The Lie Transformation Group Model of Visual Perception. Perception & Psychophysics, 34, 1-16.
Eimas, P. D. (1974). Auditory and linguistic processing of cues for place of articulation by infants. Perception & Psychophysics, 16, 513–521.
Erhart, A. (1984). Základy jazykovědy. Státní pedagogické nakladatelství Praha.
Fallside, F. (1988). On the Analysis of Linear Predictive Data such as Speech by a Class of Single Layer Connectionist Model. Technical Report, CUED/F-INFENG/TR.27, Cambridge University.
Field, A. P. (2009). Discovering Statistics Using SPSS. SAGE Publications Ltd.
Fischbach, G. D. (1992). Mind and Brain. Scientific American, September.
Flanagan, J. L. (1972). Speech Analysis, Synthesis and Perception, S.E. Springer-Verlag.
Forsyth, M. (1995). Discriminating observation probability (DOP) HMM for speaker verification. Speech Communication 17, 117-129.
Foss, D. J., Swinney, D. A. (1973). On the Psychological Reality of the Phoneme:Perception, Identification, and Consciousness. Journal of Verbal Learning and Verbal Behavior. 12, 246-257.
Földiák, P. (1991). Learning Invariance from Transformation Sequences. Neural Computation, 3, 193-199.
Fukushima, K. (1980). Neocognitron - A Self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position. Biological Cybernetics, 36, 192-202.
Furi, S. (1997). Recent Advances in Speaker Recognition, Paterrn Recognition Letters, 18.
Garris, P. A., Ciolkowski, E. L., Pastore, P., Wightma M. (1994). Efflux of Dopamine from the Synaptic Cleft in the Nucleus Accumbens of the Rat Brain. The Journal of Neuroscience, 14(10). 6064-6093.
Giles, C. L., Maxwell, T. (1987). Learning, Invariance and Generalization in High-order Neural Networks. Applied Optics, 26
Girosi, F., Poggio, T. (1990). Networks and the Best Approximation Property. Biol. Cyber. 63, 169-176.
Gnedenko, B. V. (1976). The Theory of Probability. Mir Publishers, Moskva, 1976.
Golden, R. M. (1988). A Unified Framework for Connectionist Systems. Biol. Cybernetics, 59, 109-120.
Gregory, R. L. (1997). Knowledge in perception and illusion. Phil. Trans. R. Soc. Lond. B. 352, 1121–1128.
Gross, D. J. (1996). The role of symmetry in fundamental physics. Proc. Natl. Acad. Sci., 93, 14256–14259.
Guez, A., Protopopescu, V., Barhen, J. (1988). On the Stability, Storage Capacity, and Design of Nonlinear Continous Neural Networks. IEEE SMC, 18.
Gupta, M. M., Jin, L., Homma, N. (2003). Static and dynamic Neural Networks. IEEE Press.
Haque S., Togneri R., Zaknich A. (2009). Perceptual features for automatic speech recognition in noisy environments, Speech Communication, Vol. 51, ELSEVIER.
Hecht-Nielsen, R. (1984). Kolmogorov's mapping neural network existence theorem. Proc. 2nd IEEE International Conference on Neural Networks, San Diego, California, III11-III13.
Heisenberg, W. (1966). Fyzika a filosofie, Svoboda, Praha.

Heisenberg, W. (1969). Der Teil und das Ganze, R. Piper & Co. Verlag München.
Helmholtz, H. (1954). On the Sensations of Tone, Dover, New York.
Hermansky, H., Hanson, B. A., Wakita, H. (1985). Perceptually based linear predictive analysis of speech, IEEE.
Hermansky, H. (1990). Perceptual linear predictive (PLP) analysis of speech. J. Acoust. Soc. Am., 87, 1738-1752.
Hermansky, H., Morgan, N., Bayya, A., Kohn, P. (1991). Compensation for the effect of the communication channel in auditory-like analysis of speech (Rasta-PLP). Proc. Eurospeech-91, Genova, 24-26, 1367-1370.
Hertz, J. , Krogh, A., Palmer, R. G. (1991). Introduction to the Theory of Neural Computation. Addison-Wesley Pub. Company.
Hinton, G., Anderson, J. A. (1981). Parallel Models of Associative Memory. Lawrence Erlbaum Associates, Inc.
Hirsch, M. W. (1976). Differential Topology. Springer-Verlag.
Hoggs, T., Huberman, B. A. (1987). Artificial Intelligence and Large Scale Computation -A Physics Perspective. Phys. Reports, 156, 5, North-Holland.
Hönig, F., Stemmer, G., Hacker, Ch., Brugnara, F. (2005). Revising Perceptual linear Prediction (PLP), Proceedings of INTERSPEECH 2005, s. 2997-3000, Lisbon, Portugal.
Childers, D. G., Krishnamurthy, A. K., Rocchieri, E. L., Naik, J. M. (1985). Vocal Source and Tract ModelsBased on Speech Signal Analysis. Mathematics and Computers in Biomedical Applications, Elsevier Science Pub.
Chomsky, N., Halle, M. (1968). The Sound Patterns of English. Harper & Row. N. Y.
Chudý,V., Chudý, L. Hapák, V. (1991) Isolated word recognition in Slovak via neural nets. Neurocomputing, 3, 259-282. .
Chudý,V., Chudý, L. Hapák, V. (1991) Invariant speech perception and recognition by neural nets. Neural Network World, 1, 4, 227-243.
Chudý,V., Kačúr, J. (2012). Speaker identification in real conditions using topological invariants. International Journal of Research and Reviews in Computer science (IJRRCS). Science Academy Publisher, 3, 2, 1514-1520.
Churchland, P. S., Sejnowski, T. J. (1992). The Computational Brain. A Bradford Book, MIT Press.
Jonas, P. (2000). The Time Course of Signaling at Central Glutamatergic Synapses. News Physiol. Sci. 15. 83-88.
Ju, L., Blair, D. G., Zhao, C. (2000). Detection of gravitational waves. Rep. Prog. Phys. 63, 1317–1427.
Kačúr,J., Chudý, V. (2012). Topological invariants as speech features for automatic speech recognition. International Journal of Signal and Imaging Systems Engineering (IJSISE). V tlači.
Kachar, B., Brownell, W. E., Altschuler, R. Fex, J. (1986). Electrokinetic changes of cochlear outer hair cells. Nature, 322, 365-368.
Kaiser, I. ed. (1957). Manual of Phonetics. North-Holland Publishing Company.
Kak, S. (1992). Can we build a Quantum Neural Computer ? Technical Report ECE/LSU 92-13.
Kammen, D. M., Yuille, A. L. (1988). Spontaneous Symmetry - Breaking Energy Functions and the Emergence of Orientation Selective Cortical Cells. Biol. Cyber., 39, 23-31.
Kao,Y., Rajasekaran, P., Baras, J. (1992). Free-text speaker identification over long distance telephone channel using hypothesized phonetic segmentation. Proc. IEEE ICASSP, II. 177- II. 180.
Kohonen,T. (1989). Self-organisation and Associative memory. Springer-Verlag.
Kolers, P. A., Eden, M. (1968). Recognizing Patterns. MIT Press, Cambridge, MA.

Kolmogorov, A. N. (1957). On the Representation of Continuos Functions of Many Variables by Superpositions of Continous Functions of One Variable and Addition. Dokl. Akad. Nauk. USSR, 114, 953-956.

Kumar, N. (1997). Investigation of Silicon Auditory Models and generalization of linear discriminate analysis for improved speech recognition, PhD thesis, Baltimore.

Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. Perception and Psychophysics, 50, 93–107.

Lamel, L. F., Rabiner L. R., Rosenberg, A. E., Wilpon, J. G. (1981). An Improved Endpoint Detector for Isolated Word Recognition. IEEE on ASSP, 29.

Landau, L. D., Lifshitz, E. M. (1987). Fluid Mechanics, Pergamon, Oxford.

Landauer, R. (1961). IBM J. Res. Develop., 5, 183.

Lauterborn, W., Parlitz, U. (1988). Methods of Chaos Physics and their Application to Acoustics. J. Acoust. Soc. Am, 84.

Lee, K. C., Sprague, M. R., Sussman, B. J., Nunn, J., Langford, N. K., Jin, X. M., Champion,T., Michelberger, P., Reim, K. F., England, D., Jaksch, D., I., Walmsley, A. (2011). Entangling Macroscopic Diamonds at Room Temperature diamond structure. Science, 334, 2.

Lenz, R. (1990). Group Invariant Pattern Recognition. Pattern Recognition, 23, 199-217.

Lenz, R. (1991). On Probabilistic Invariance. Neural Network, 4, 627, 642.

Li, Y. (1991). Application of Moment Invariants to Neurocomputing for Pattern Recognition. Electronic Letters, 27, 587-588.

Lieberman, A. M., Cooper, F. S., Shankweiler, D. P., Studdert-Kennedy, M. (1967). Perception of the speech code. Psych. Rev., 74, 431-461.

Lieberman, P. (2007). The Evolution of Human Speech, Current Anthropology, Vol. 48, 1, 39-66.

Lindberg, B., Johansen, F. T., Warakagoda, N., Lehtinen, G., Kacic, Z. Zgang, A., Elenius, K., Salvi, G. (2000). A Noise Robust Multilingual Reference Recognizer Based on SpeechDat(II), In Proceedings of ICSLP 2000, Beijing, China.

Lippman, R. P. (1987). An Introduction to Computing with Neural Nets. IEEE ASSP Magazine.

Lippman, R. P. (1988). Neural Network Classifiers for Speech Recognition. MIT Lincoln Lab..

Lippman, R. P. (1989). Review of Neural Networks for Speech Recognition. MIT Lincoln Lab.

Liou, H. S., Mammone, R. J. (1995). Speaker verification using phoneme-based neural tree networks and phonetic weighting scoring method. in Neural Networks for Signal Processing V, Proceed.of the 1995 IEEE Workshop.

Liu, W., Andreou, A. G., Goldstein, M. H. (1991). Analog VLSI Implementation of an Auditory Periphery Model. 25th Annual Conference on Information Science and Systems, Baltimore.

Liu, X., Gales, M. J. F., Woodland, P.C. (2003). Automatic complexity control for HLDA systems, ICASSP.

Locher, P., Nodine, C. (1989). The Perceptual Value of Symmetry. Computers Math. Applic., 17, 475-484.

Lopez-Villanueva, J. A., Jimenez -Tejada, J. A., Cartujo, P., Bausells, J., Carceller, J. E. (1991). Analysis of the effects of constant-current Fowler-Nordheim-tunneling injection with charge trapping inside the potential barrier. J. Appl.Phys., 70,7, 3712-3720.

Luce, D. R., Bush, R.R., Galanter, E. (1987). Handbook of Mathematical Psychology. J. Wiley & Sohns, Inc.

Lyon, R. F., Mead, C. (1988). An Analog Electronic Cochlea. IEEE ASSP, 36.

Maeda, S. (1982). A Digital Simulation Method of the Vocal-tract System. Speech Communication, 199-229.

Makhoul, J. (1975). Spectral Linear Prediction - Properties and Applications. IEEE on ASS 23.

Makhoul, J. (1975). Linear Prediction - A Tutorial Review. Proc. IEEE, 63.

Markel, J.D., Gray, A. H. (1976). Linear Prediction of Speech. Springer-Verlag.

Maršala, J. (1985). Systematická a funkčná neuroanatómia. Vyd. Osveta.

Martin, T. (1975). Applications of limited vocabulary recognition systems. In: Rec. 1974 Symp. Speech Recognition, D. R. Reddy, Ed. New York, Academic Press, 55-71.

Mařík, V. Štěpánková, O., Lažanský, J. a kol. (1993-2003). Umělá inteligence I, II, III, IV. Academia Praha.

Mas, J., Ramos, E. (1989). Symmetry Processing in Neural Network Models. J. Phys. A: , 22, 3379-3391.

Masahiro, A. (1988). Phys. Rev. D, 12, 4415.

Massey, W. S., Stallings, J. (1967). Algebraic Topology - An Introduction. Harcourt, Brace & World, Inc.

Matsui, T., Furui, S. (1991). A text-independent speaker recognition method robust against utterance variations. Proc. IEEE ICASSP, 377-380.

Matsui, T., Furui, S. (1992). Comparision of text-independent speaker recognition methods using VQ-distortion and discrete/continuous HMMs. in Proc. IEEE ICASSP, II. 157-164.

Mazzola, G., Wieser, H. G., Brunner, V., Muzzulini, D. (1989). A Symmetry -oriented Mathematical Model of Classical Counterpoint and Related Neurophysiological Investigations by Depth EEG. Computers Math. Applic., 17, 539-594.

Meddis, R. (1986). Simulation of Mechanical to Neural Transduction in the Auditory receptor. J. Acoust. Soc. Am., 79.

Meddis, R., Hewitt, M. J., Shackleton, T. M. (1990). Implemetation Details of a Computation Model of the Inner Hair-cell/ Auditory - Nerve synapse. J. Acoust. Soc. Am., 87.

Minsky, M. I., Papert, S. (1969). Perceptron. MIT Press, Cambridge, MA.

Mitchell, T. M. (1997). Machine Learning, McGraw-Hill, ISBN 0-07-042807-7.

Moller, C. (1972). The Theory of relativity. Clarendon Press, New Haven, CT.

Moore, B. C. J., Tyler, L. K., Marslen-Wilson, W. (2007). Introduction. The perception of speech: from sound to meaning. Phil. Trans. R. Soc. B.,363, 917–921.

Mozer, M. C. (1991). The Perception of Multiple Objects- A Connectionist Approach, MIT Press.

Musavi, M. T. , Kalantri, K., Ahmed, W., Chan, K. H. (1990). A Minimum Error Neural Network (MNN). Neural Networks, 6, 397-407.

Nadeu C., Macho D. (2001). Time and Frequency Filtering of Filter-Bank energies for robust HMM speech recognition, Speech Communication. Vol. 34, Elsevier.

Naik, J. M., Lubenski, D. M. (1992). A hybrid HMM -MLP Speaker verification algorithm for telephone speech. Proc. Interat. Conf. Acoust. Speech Signal Process., I, 153-156.

Nairz, O., Arndt, M., Zeilinger, A. (2003). Quantum interference experiments with large molecules. Am. J. Phys. 71, 4, 319-325.

Nakagawa, S. Ueda, Y., Seino, T. (1992). Speaker-independent, Text-independent Language Identification by HMM, ICSLP.

Newell, A. Barnett, J., a kol. (1973). Speech Understanding Systems. North-Holland Publishing Company.

Niederjohn, R. J. (1975). A Mathematical Formulation and Comparision of Zero-Crossing Analysis Techniques which have been Applied to Automatic Speech Recognition. IEEE ASSP, 23.

Niranjan, M., Fallside, F. (1988). Neural Networks and Radial-Basis-Functions in Classifying Static Speech Patterns. Technical Report, CUED/F-INFENG/TR.22, Cambridge University.

Nobili, R., Mammano, F., Ashomore, J. (1998). How well do we understand the cochlea ? Trends. Neurosci., 21, 159-167.

Noether, E. (1918). Invariante Variations probleme, Nachr. d. König. Gesellsch. d. Wiss. zu Göttingen, Math-phys. Klasse, 235–257.

Nooteboom, S. G., Van der Vlugt, M. J. (1988). A Search for a word-beginning superiority effect. J. Acoust. Soc. Am. 64.

Nouza, J., Zdansky, J., David, P., Cerva, P., Kolorenc, J., Nejedlova, D. (2005). Fully Automated System for Czech Spoken Broadcast Transcription with Very Large (300K+) Lexicon. Proceedings of Interspeech 2005, ISSN 1018-4074, Lisboa, Portugal, 1681-1684.

Nussenzweig, H. M. (1972). Causality and Dispersion Relation, Academic Press, N. Y.

Oglesby, J., Mason, J.S. (1991). Radial basis function networks for speaker recognition. Proc. Internat. Conf. Acoust. Speech Signal process. '91, S6.7, Toronto, Canada, 393-396.

Oliver, D., He, D. Z., Klocker, N., Ludwig, J., Schulte, U., Waldegger, S., Ruppersberg, J. P., Dallos, P., Fakler, B. (2001). Intracellular Anions as the Voltage Sensor of Prestin, the Outer Hair Cell Motor Protein. Science, 292, 2340-2343.

Oppenheim, A. V., Shafer, R. W. (1975). Digital Signal Processing. Prentice-Hall, Inc.

O'Shaugnessy, D. (1986). Speaker Recognition. IEEE ASSP Magazine.

O'Shaugnessy, D. (1990). Speech Communication. Addison Wesley Publishing Company.

Palacios-Laloy, A. a kol. (2010). Nature Phys. 6, 442–447.

Paliwal, K. K. (2004). Usefulness of Phase in Speech Processing. JNRSAS ISSN:1547-0407.

Park, J., Sandberg, I. W. (1991). Universal Approximation Using Radial-Basis-Function Networks. Neural Computation, 3, 246-257.

Pessa, E. (1988). Symmetry Breaking in Neural Nets. Biol. Cyber., 59, 277-281.

Piaget, J., Inhelderová, B. (2007). Psychológie dítěte, vyd. Portál.

Pickles, J. O. (1988). An Introduction to the Physiology of Hearing, S.E. Academic Press.

Picone, J. (1990). Continous Speech Recognition Using Hidden Markov Models. IEEE ASSP Magazine.

Pisoni, D. B., Remez, R. E., ed. (2005). The handbook of speech perception. Blackwell Publishing Ltd.

Pitts, W., McCulloch, W. S. (1947). How we Know Universals - the Perception of Auditory and Visual Forms. Bulletin of Mathematical Biophysics, 9, 127-147.

Poeppel, D., Idsardi, W. L. , van Wassenhove, V. (2008). Speech perception at the inter face of neurobiology and linguistics, Phil. Trans. R. Soc. B., 363, 1071–1086.

Ptáček, M., Buček, A., Dvořák, P. (1985). Zařízení hlasového vstupu HR 01. Tesla VÚST, A.S. Popova.

Rabiner, L. R., Sambar, M. R. (1977). Voiced - Unvoiced Detection Using the Itakura LPC Distance Measure, Bell Lab.

Rabiner, L., R., Wilpon, J. G. (1979). Speaker Indepedent Isolated Word Recognition for a Moderate Size (54 Word) Vocabulary. IEEE on ASS, 27.

Rabiner, L. R. (1989). A tutorial on HMM and Selected Applications in Speech Recognition. In:[WL], PROCEEDINGS OF THE IEEE, 77, 2, 267-296.

Rabiner, L., Juan, B. (1993). Fundamentals of speech recognition, ISBN 0-13-015157-2, Prentice Hall PTR, New Yersey.

Reddy, D. R. ed. (1975). Speech Recognition. Academic Press.

Reddy, D. R. (1966). Segmentation of Speech Sounds. J. of Acoust. Soc. Am., 307-312.

Reddy, D. R. (1966). Phoneme Grouping for Speech Recognition. J. of Acoust. Soc. Am., 1295-1300.

Reddy, D. R. (1967). Computer Recognition of Connected Speech. J. of Acoust. Soc. Am., 329-347.

Reddy, R. R., Erman, L. D., Neely, R. B. (1973). A model and a System for Machine Recognition of Speech. IEEE on AEA, 21.

Reynold, D. A. (2002). An Overwiev of Automatic Speaker Recognition Technology, In Proc. Of International Conference on Acoustic, Speech, and Signal Processing, Orlando USA.

Reynolds, D. A., Rose, R. C. (1995). Robust text-independent speaker identification using gaussian mixture speaker models. IEEE Trans. Speech and Audio Processing, 3, 72-83.

Rhode, W. S. (1971). Observations of the Vibration of the basilar membrane in squirrel monkeys using Mössbauer technique. J.Acoust. Soc. Amer., 49, 1218-1231.

Ritter, H., Kohonen, T. (1989). Self-organizing Semantic Map. Biol. Cybernetics, 61, 241-254.

Robinson, A. J., M., Fallside, F. (1988). A Dynamic Connectionist Model for Phoneme Recognition - Preliminary Results. Technical Report, CUED/F-INFENG/TR.14, Cambridge University.

Robinson, A. J., M., Fallside, F. (1990). Phoneme Recognition from the TIMIT database using Recurrent Error Propagation Networks. Technical Report, CUED/F-INFENG/TR.42, Cambridge University.

Rumelhart, D. E., Zipser, D. (1985). Feature Discovery by Competetive Learning. Cognitive Science, 9, 75-112.

Russell, I. J., Nilsen, K. E. (1997). The location of the cochlear amplifier; spatial representation of a single tone on the guinea-pig basilar membrane. Proc. Natl. Acad. Sci. U.S.A., 94, 2660-2664.

Ryder, L. H. (1985). Quantum Field Theory. Cambridge University Press, Uk.

Sakoe, H., Chiba, S. (1978). Dynamic Programming Algorithm Optimization for Spoken Word Recognition. IEEE on ASSP, 26.

Scarr, W. A. (1968). Zero Crossings as a Means of Obtaining Spectral Information in Speech Analysis. IEEE on AEA, 16.

Secker-Walker, H. E., Searle, C. L. (1990). Time domain analysis of auditory-nerve-fiber firing rates. J Acoust. Soc. Am., 88, 1427-1436.

Segall, P. (1986). Úvod do syntaxe a sémantiky. Academia Praha.

Sejnowski, T. J., Rosenberg, Ch. R. (1986). NetTalk- A Parallel Network that Learns to Read Aloud. Technical Report JHU/EECS-86/01, J. Hopkins University.

Sellick, P. M., Patuzzi, R., Johnstone, B. M. (1982). Measurement of basilar membrane motion in the guinea pig using the Mössbauer technique. J.Acoust. Soc. Amer., 72, 131-141.

Shalkoff, R. J. (1992). Pattern Recognition - Statistical, Structural and Neural Approaches. J. Wiley & Sohns, Inc.

Siegel, L.J., Bessey, A. C. (1982). Voiced/Unvoiced/Mixed Excitation Classification of Speech. IEEE ASSP, 30.

Smola, A. J., Schölkopf, B. (2004). A tutorial on support vector regression, Statistics and Computing, Vol. 14, No. 3.

Specht, D. F. (1990). Probabilistic Neural Network. Neural Networks, 3, 109-118.

Specht, D. F. (1990). Probabilistic Neural Networks and Polynomial Adaline as Complemetary Techniques for Classification. IEEE NN, 1.

Spitzer, H., Hochstein, S. (1985). A Complex-cell Receptive Field Model. Journal of Neurophysiology, 53, 1266-1286.

Stern, R. M., Lasry, M. J. (1987). Dynamic Speaker Adaptation for Feature-Based Isolated Word Recognition, IEEE on ASSP, 35.

Strube, H. W. (1985). A computationally Efficient Basilar Membrane Model. Acoustica, 58.

Tatham, M. A. A. (1984). Towards a Cognitive Phonetics. Journal of Phonetics, 12, 37-47.

Tegmark, M. (1993). Foundations of Physics Letters, 6, 6, 571-590.

Thevenaz, P., Hügli, H. (1995). Usefulness of the LPC-residue in text-independent speaker verification. Speech Communication 17, 145-157.

Tirakis, A., Delopoulos, A., Kollias, S. (1992). Cumulant-based Neural Network Classifiers. In : Proceedings of ICANN-92, North-Holland.

Tishby, N. Z. (1991). On the application of mixture AR hidden Markov models to text independent speaker recognition. IEEE trans.Signal Processing, 39, 563-570.

Torkkola, K. (1991). Short-time Feature Vector Based Phonemic Speech Recognition with the Aid of Local Context. PhD. Thesis. Helsinky University of Technology.

Verlinde, E. (2010). On the Origin of Gravity and the Laws of Newton. arXiv.1001.0785v1. [hep-th].

Von Békesy, G. (1960). Experiments in Hearing. McGraww-Hill, N. Y.

Wechsler, H. (1990). Computational Vision. Academic Press, Inc.

Wheeler, J. A., Zurek, W. H. ed. (1983). Quantum Theory and Measurement.Princeton University Press, Princeton.

Witten, H. I. (1982). Principles of Computer Speech. Academic Press.

Wu, T. T., Yang, C. N. (1975). Phys. Rev., D, 12, 3845.

Yang, X., Wang, K., Shamma, S. (1992). Auditory Representations of Acoustics Signals. IEEE IT, 38.

Yau, H., Manry, M. T. (1990). Iterative Improvement of a Gaussian Classifier. Neural Networks, 3, 437-443.

Yoshifusa, I. (1991). Approximation of Functions on a Compact Set by Finite Sums of Sigmoid Function without Scaling. Neural Networks, 4, 817-826.

Zeh, H. D. (1970). On the Interpretation of Measurement in Quantum Theory. Found. Phys., 1, 69.

Zhen, J., Shen, W., He, D. Z., Long, K.B., Madison, L. D., Dallos, P. (2000). Prestin is the Motor Protein oc Cochlear Outer Hair Cells. Nature, 405, 149-155.

Zurek, W. H., Habib, S., Paz., J. P. (1993). Coherent States via Decoherence. Phys. Rev. Lett. 70, 9, 1187.