



Översikt - talarigenkänning

- Talarverifiering - talaridentifiering
- För- och nackdelar
- Tillämpningar
- Metoder
- Två typer av fel, beslutströskel
- Forskningsläget
- Projekt på KTH

Taltekniologi 2002-2-04 (1)



Personidentifiering

- Metoder bygger på
 - något man har
 - (Ex: nyckel, passerkort)
 - något man kan
 - (PIN-kod)
 - något man är
 - (fysisk egenskap, beteende: biometri)

Taltekniologi 2002-2-04 (2)



Biometriska identifieringsmetoder



Taltekniologi 2002-2-04 (3)



Talarverifiering / -identifiering

- Talarverifiering
 - Identiteten uppges på annat sätt och verifieras med rösten
 - Binärt beslut: "acceptera eller avvisa?", "sann kund eller bedragare?"
 - Prestanda minskar ej med antalet användare
- Talaridentifiering
 - Välja 1 av N: "Vem är användaren?"
 - Med eller utan avvisning som alternativ
 - Prestanda minskar med antalet användare
- "Talarföljning" (speaker tracking)
 - Vid vilka tidpunkter talar en viss person?
 - t.ex. under telefonkonversation eller i radio- och TV-program

Taltekniologi 2002-2-04 (4)



Fördelar/problem talarverifiering

- + Tal är naturligt
- + Enkelt att mäta, ej störande
- + Billig
 - enkel utrustning
 - speciellt liten merkostnad om tjänsten redan är röststyrd
- + Ej 100% säkerhet, men
 - Så även med andra metoder
 - Kan kombineras med andra metoder
 - Höjer lönsamhetsströskeln för organiserad brottslighet
 - Avskräckande effekt
- Stor spridning för en talare vid skilda tillfällen
 - Beteende
 - Mikrofon
 - Hälsotillstånd
- Taligenkänningsproblem

Taltekniologi 2002-2-04 (5)



Applikationer - exempel

- Telekommunikation
 - Banktjänster, även som komplement till manuella
 - Kreditkortsanvändning
 - Tillgång till information
 - Debitering av telefonsamtal
 - (Begränsning av utgående telefonsamtal)
- On-site
 - Inpasserings- och behörighetskontroll
 - Internering i hemmet av brottslingar (största tillämpning i USA)
- Brottsutredningar
 - Objektiva automatiska metoder
- Talarindexering i radio- och TV-program

Taltekniologi 2002-2-04 (6)

Typen av textberoende

- Textberoende med fast lösenord
- Textberoende med kundspecifikt lösenord
- Vokabulärberoende
 - t ex siffror
- "Händelseberoende"
 - t ex tittar t ex på vissa fonem i löpande text
- Textberoende, systemet bestämmer en text
 - kombination av talar- och textverifiering
- Textberoende, kunden väljer en fri text

Minskat textberoende

Talsteknologi 2002-2-04 | 7 |

Talarspecifika parametrar

- Spektrum: talrör och röstkälla
- Både statiska och dynamiska drag
- Anatomiska
 - Längden på talröret, storleken på nasala kaviteter, stämbandsegenskaper
 - Formantfrekvenser och bandbredder, medelgrundton, spektrumlutning
- Inlärd
 - dialekt, talstil
 - grundton, talhastighet, styrka, formantfrekvenser
- Svåra att separera!

Talsteknologi 2002-2-04 | 8 |

Annat eller samma analys som i taligenkänning?

- TAL-igenkänning bör vara TALAR-oberoende
 - extrahera fonetisk information men ej talarinformation
- TALAR-igenkänning bör vara TAL-oberoende
 - extrahera talarinformation men ej fonetisk information
- Men experiment har visat att den bästa TAL-representationen också är bland de bästa TALAR-representationerna
- Varför? Kanske den optimala talsignalmodellen kan innehålla både TAL- och TALAR-information

Talsteknologi 2002-2-04 | 9 |

Moduler - talarverifiering

I stort sett samma som för taligenkänning

Parametrisering	Parametertyp: LPCC, MFCC, Energi, F0, etc Robusthet: Delta, RASTA, cepstrumnormalisering, Transformationer (LDA)
Talarmodell Matching	Teknik: HMM, DTW, Algebraiska metoder Enheter: Fonem, ord, fraser
Beslut	Normalisering: Jämförande matchning mot normaltalare Kompenserar för avvikelser (buller, brus, telefon)

Talsteknologi 2002-2-04 | 10 |

Två typer av fel

		Uppgiven identitet:	
		sann	falsk
Beslut:	acceptera	OK	felaktig acceptans (false accept, FA)
	refusa, avvisa	felaktigt ratande (false reject, FR)	OK

Talsteknologi 2002-2-04 | 11 |

Två faser i talarverifiering

Metoder

Registrering (träning, enrolment)

Träningsyttrandet från en ny kund → Spektralanalys → Träna modell → Tränad talarmodell

Verifiering

Inloggningsyttrande → Spektralanalys → Jämförelse → Släpp in / Avvisa

Hävdad identitet

Talsteknologi 2002-2-04 | 12 |

Metoder

Varför probabilistiskt synsätt

- Två möjliga metoder att mäta överensstämmelsen mellan testyttrandet och referensdata för uppgiven kund
 - Alt 1: Akustisk likhet, spektral distans
 - Alt 2: Sannolikheten för att kunden "talar så här"
- Fördelar/nackdelar med sannolikhetsmått
 - + Kan modellera talets stokastiska egenskaper
 - + Enhetligt mått vid kombination av olika metoder
 - Svårt att uppskatta statistisk fördelning med litet talmaterial
- Beslutsriterium
 - Möjligt att välja en tröskel som minimerar kostnaden för felbeslut
 - Olika kostnad för False accept / False reject

Taltekniologi 2002-2-04 | 13 |

Metoder

Sannolikhetsbaserat beslutsmått

- Bayes beslutsteori
 - Kvoten mellan klientmodellens och en antiklientmodell (bakgrundsmodell)s sannolikhet jämförs med tröskel

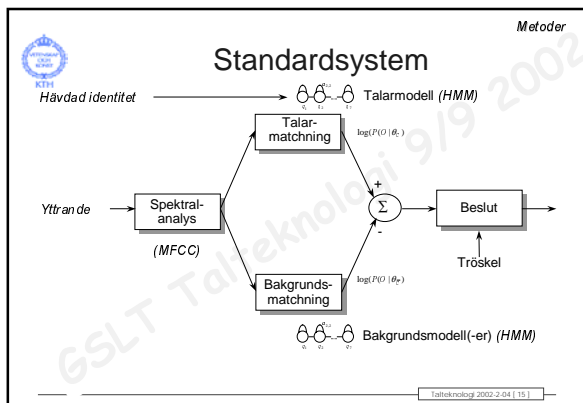
Om $\frac{P(\text{Klienten talar så här})}{P(\text{Någon annan talar så här})} > R$ $\frac{P(O|\theta_c)}{P(O|\theta_c^-)} \geq R$

Så godkänn, annars underkänn

O: yttrande
 θ_c : klient C:s modell

Tröskeln R kan anpassas för att ge önskad felbalans, minsta totalfel eller lägsta felkostnad

Taltekniologi 2002-2-04 | 14 |



Metoder

Bakgrundsmodeller

Två varianter på bakgrundsmodeller (antiklienter)

- Metod 1: "Världsmodell" En modell, tränas på stort antal talare Klientoberoende

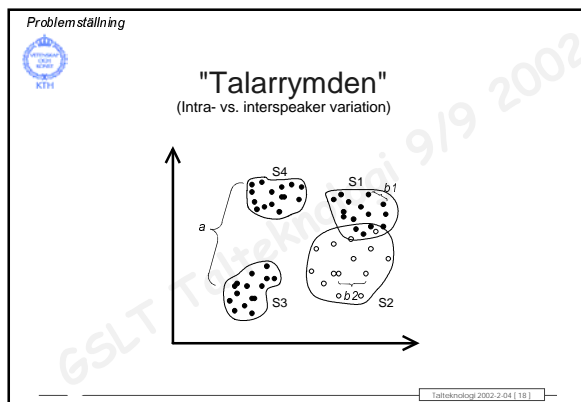
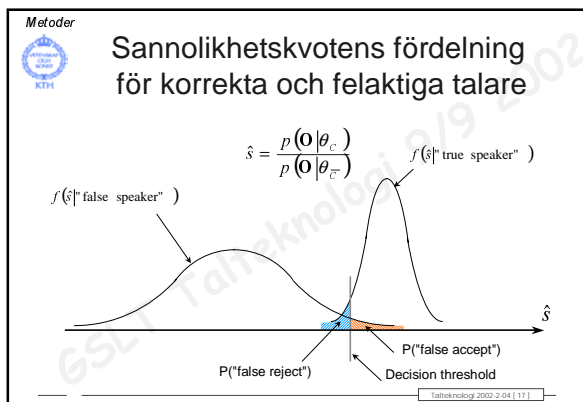
$$P(O|\theta_c^-) \approx P(O|\theta_M)$$

- Metod 2: "Kohort-modell" Flera delmodeller tränade på mindre talargrupper "nära" resp. klient Klientspecifik

$$P(O|\theta_c^-) \approx \frac{1}{|W_M|} \sum_{i \in W_M} P(O|\theta_i) \approx \frac{1}{N} \sum_{i \in W_M} P(O|\theta_i)$$

N "closest"

Taltekniologi 2002-2-04 | 16 |

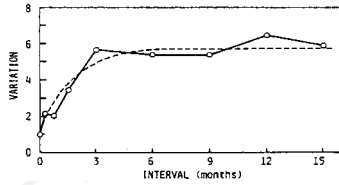


Problemställning



Rösten varierar över tiden

- Spridning inom en talare



Akustisk variation bland identiska yttranden som en funktion av längden av intervallet för inspelningarna. Medelvärde för nio manliga talare. (Furui, 1986)

Problemställning

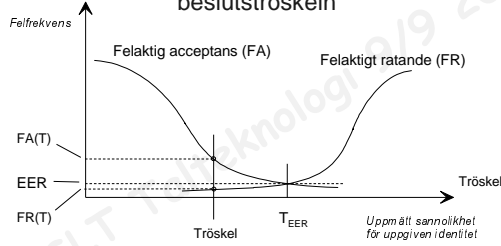


Telefonnätets inverkan

- Olika telefoner (mikrofoner)
- Transmission
 - Varierande ledningar och utrustning
 - Digital kodning
 - Brus
- Liten kontroll över talaren och miljön som talaren ringer ifrån
- Svårt att separera talarspecifika parametrar från miljöspecifika parametrar!



Felkaraktistiken beror på beslutströskeln



EER: Equal Error Rate, vanligt mått på systemets prestanda
HTER: Half Total Error Rate, $(FA(t) + FR(t))/2$



Prestandamått

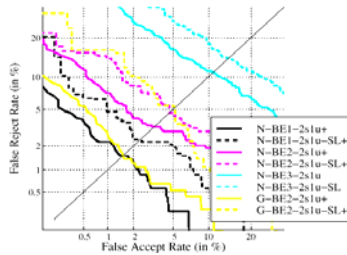
Utvärdering

- False Rejection rate (FR)
 - $FR = (\text{Antal false reject}) / (\text{Antal korrekta försök})$
- False Acceptance rate (FA)
 - $FA = (\text{Antal false accept}) / (\text{Antal intrångsförsök})$
- Half Total Error Rate (HTER)
 - $HTER = (FR + FA) / 2$
- Equal Error Rate (EER)
 - $EER = FR = FA$ vid en i efterhand vald tröskel
 - Välddefinierad punkt, men kan ej väljas i praktiken
- Detection Error Trade-off (DET)
 - Visar FR och FA vid olika beslutströsklar
 - Motsvarar "Receiver Operating Characteristics" (ROC)



Detection Error Trade-off (DET)

Utvärdering



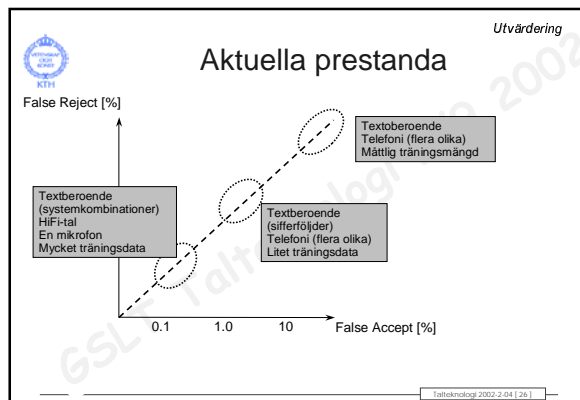
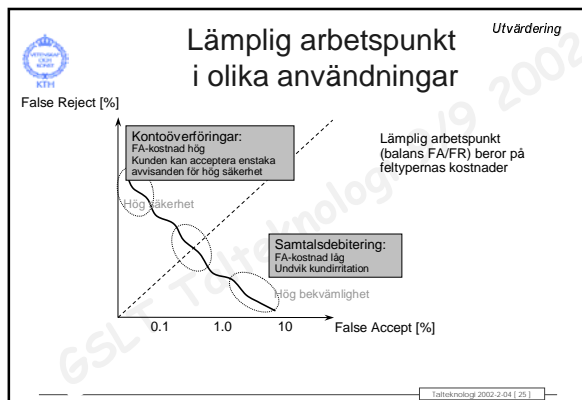
Minimera felkostnaden

Metoder

Felkostnadsminimerande tröskel $\hat{R} = \frac{P(\bar{C}) K_{FA}}{P(C) K_{FR}}$

Den optimala tröskeln minimerar den totala kostnaden för felbeslut $(FA + FR)$. Den är beroende av hur ofta korrekta inpasseringsförsök $P(C)$ och intrångsförsök $P(\bar{C})$ sker samt kostnadsrelation mellan False Accept (K_{FA}) och False Reject (K_{FR})

Om kostnaderna är lika och bedragare är lika vanliga som sanna kunder blir tröskeln = 1, dvs acceptera identiteten om sannolikheten för rätt kund är större än sannolikheten för "vem som helst". Rimligt.




- Utvärdering
- ## Säkerhetsaspekter
- Uppgivna prestanda mätta med slumpvisa röster som bedragare
 - Vilka prestanda vid verkliga bedrägeriförsök?
 - Härmning? Inspelningar? "Personlig" talsyntes?
 - Kombination med annan metod höjer säkerheten
 - Kan skydda vid stöld/rån av kort + PIN-kod
 - Avskräckande effekt
 - Inspelningar kan avlysnas i efterhand
 - Minskar troligen antalet intrångsförsök
- Tallteknologi 2002-2-04 | 27 |

- Utvärdering
- ## Bedragaraspekter
- Att fundera på:
 - Känner han till lösenordet?
 - Har han inspelningar?
 - Kan han köpa information?
 - Härmar han?
 - Familjemedlem, tvilling??
 - Hur mycket kan de skada?
 - Farligast?
 - När någon säljer kontonummer, PIN-koder och lösenord på "postorder"?
 - Professionell brottsling som kan bugga telefoner, göra inspelningar, adaptiv talsyntes?
- Tallteknologi 2002-2-04 | 28 |

- Användning
- ## Teknisk och naturlig härmning
- Riskscenarier
 - Inspelat verifieringsyttrande av måltalaren
 - Inspelade verifieringsord spelas upp i rätt sekvens
 - Inspelade verifieringsord av talare med liknande röst
 - Transformering av yttrande till måltalaren
 - Individuellt anpassad syntes
 - Naturlig härmning (spec. tvilling, familjemedlem, imitator)
 - Dagens system är sårbara för teknisk härmning
- Tallteknologi 2002-2-04 | 29 |

- Användning
- ## Teknisk och naturlig härmning
- Motåtgärder
 - Undvik fasta lösenord (kan spelas in)
 - Undvik vanliga ord i lösenordsmeningen
 - Vissa personer bör använda annan metod
 - Kombinera metoder
 - Individuella lösenord
 - Detektera manipulerat tal
 - Kraftfullare motåtgärd för högre skyddat värde
 - Bytet ska inte vara värt besväret
- Tallteknologi 2002-2-04 | 30 |

Utvärdering

 **Att kombinera metoder**

- Hur kan röstverifiering komplementera PIN-kod?
- Säkerhet med PIN-kod:
 - Antag FR = 0.2% Korrekt användare slarv/glömska (1 på 500)
 - FA = 0.01% Om förbrytaren ej känner till koden, lyckas 1 på 10 000
 - FA = 100% Om förbrytaren kan koden, kommer han garanterat in
- Utför röstverifiering på de som klarar PIN-koden
- Får ej nämnvärt höja FR => FR(röst) bör ≈ 0.2%
- Ur DET-diagrammet fås ny FA för FR = 0.2%
 - som exempel: linjär extrapolation av bästa kurva i DET-diagrammet ovan (EER ≈ 1.5%)
 - FA ≈ 20% (bättre än 100% utan röstverifiering) Hindrar 4 bedrägeriförsök av 5

Taltekniologi 2002-2-04 | 31 |


 **Djurparken**

"Klassning" av en användare efter hur bra systemet fungerar för denne


- Får - "Snälla" användare med låg förväxlingsrisk
- Getter - "Opålitliga", hög variabilitet ger stor förväxlingsrisk
- Lamm - Känsliga för bedragare, lätta att imitera
- Vargar - Potentiella bedragare

Tilltalande men felaktig beskrivning, som lägger problemet hos användarna, inte i systemets begränsningar


Taltekniologi 2002-2-04 | 32 |

 **Användaraspekter**

- Så litet träning som möjligt, helst ingenting
 - Talarens variabilitet kan ej mätas vid träning
- Talarverifiering ska förenkla, inte försvåra användningen, helst ske omärkligt
- Dörrvakt eller varningssignal?
- Vilken balans FA / FR?
 - Beror på säkerhetskrav och kostnader
 - Korrekta användare bör ej störas





Taltekniologi 2002-2-04 | 33 |

 **Projekt på CTT + TMH**



- **TVIT** - TalarVerifiering i Telenätet (*Telia - 1995-1998*)
- **CAVE, PICASSO** -Användningsförsök i bank, telekom (*EU-projekt 1995-1997, 1998-2000*)
- **Verivox** - Styrning av användarbeteende (*EU-proj 1997*)
- **PER** - Inpasseringssystem vid TMH (*CTT*)
- **CTT-Bank** -telefonbank med talgränssnitt (*CTT*)
- Utförd och pågående forskning
 - Träningsmetoder
 - Härmning (mänsklig och syntetisk)
 - Textberoende system
 - Kombination med PIN-kod
 - Experiment med Artificiella Neurala Nät (*ANN*) (*Exjobb*)


Taltekniologi 2002-2-04 | 34 |

 **CTT-projektet PER** Dagsläget 
(Prototype Entrance Receptionist)

- Detekterar närvaro av personer vid TMHs entré och tilltalar dessa
- Identifierar anställda mha talarverifiering och öppnar grinden
 - Säg ditt namn och en slumpad siffersekvens
- Hälsar (in- och utgående)
- Ska så småningom tala med och hjälpa besökare

Taltekniologi 2002-2-04 | 35 |

 **CTT-Bank** Dagsläget 
röststyrd telefonbank

- Deltagare
 - CTT, Handelsbanken, Ericsson, Trio, Hjälpmedelsinstitutet
- Tjänster
 - Överföring mellan (virtuella) konton
 - Uppläsning av senaste transaktioner
 - Saldofråga
- Användarstudie juni-juli 2000
 - 20 försökspersoner
 - Användarförtroende: 3.5 (skala 1 - 5)
 - Skulle du vilja använda CTT-Bank för dina egna pengar?
 - 67% Ja, 33% Nej
 - Talarverifiering FA - 5%, FR - 4%
 - Taligenkänning WER - 5% (siffror)
 - Lärdomar: Dialog, Felhantering, Teknik
- Dialogexempel 

Taltekniologi 2002-2-04 | 36 |



CAVE - Caller Verification in Banking and Telecommunication PICASSO - Pioneering Caller Authentication for Secure Service Operation

- Syften
 - Testa försökssystem i praktisk användning: bank och telefoni
 - Användarkrav, marknadsaspekter, tillämpad forskning
 - Kombinerad taligenkänning och talarverifiering
- Deltagare
 - Nederländerna, Frankrike, England, Schweiz, Sverige (KTH + Telia (CAVE))
- Försök
 - Banktransaktion på telefon, Telefonering med telefonkort, Aktiehantering, Nummerupplysning för synskadade
- Resultat
 - Bra forskningsresultat. Positiv användarattityd i försöken. Patent, Video



Sammanfattning

- Användbar idag i begränsande tillämpningar
- Kan höja säkerheten i kombination med andra metoder
- Positiv användarattityd i praktiska försök
- Användaraspekter måste tas i beaktande
- Mer forskning => ökad säkerhet => bredare användning