

1. Inleiding

- Detecteren van ongewenste video's op sociale media
- Case: haatdetectie via audio analyse
- Keyword spotter (kws) voor scheldwoorden
- Classificeer video d.m.v. output keyword spotter
- Meerdere scheldwoorden per video
- IWT beurs 120007 (AMiCA)



2. Data

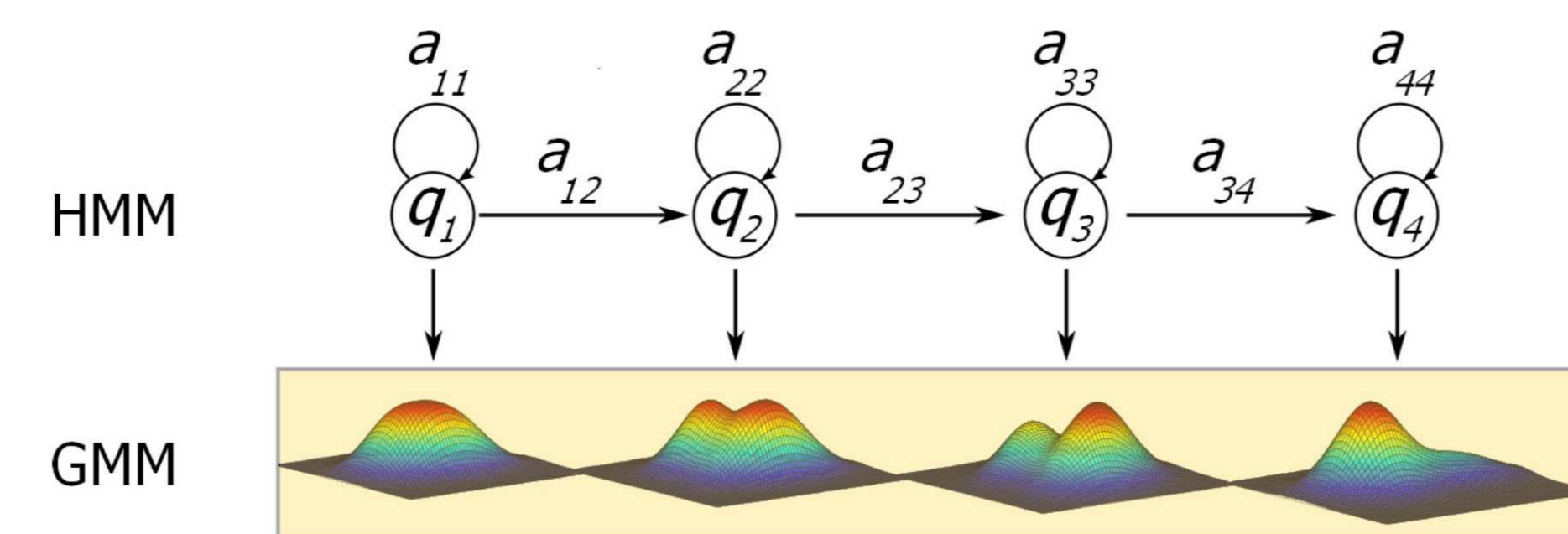
- Filmpjes van YouTube
- Engels wegens overvloed aan beschikbaar materiaal
- Meerdere sprekers
- Veel factoren die performantie spotter verlagen: achtergrondlawaai, overlappende spraak, zwaar accent, slechte verstaanbaarheid, grote afstand tot microfoon of microfoonruis
- Haatvideos onderverdeeld in 3 klassen:
 - Klasse 0: Alle videos
 - Klasse 1: Enkel video's waarin geen van bovenstaande factoren sterk aanwezig
 - Klasse 2: Enkel video's waarin geen van bovenstaande factoren sterk of mild aanwezig

# Vids	geen haat	haat		
		klasse 0	klasse 1	klasse 2
	129	318	179	74
Gem. duur	4:19	3:29	3:39	4:00

3. Model

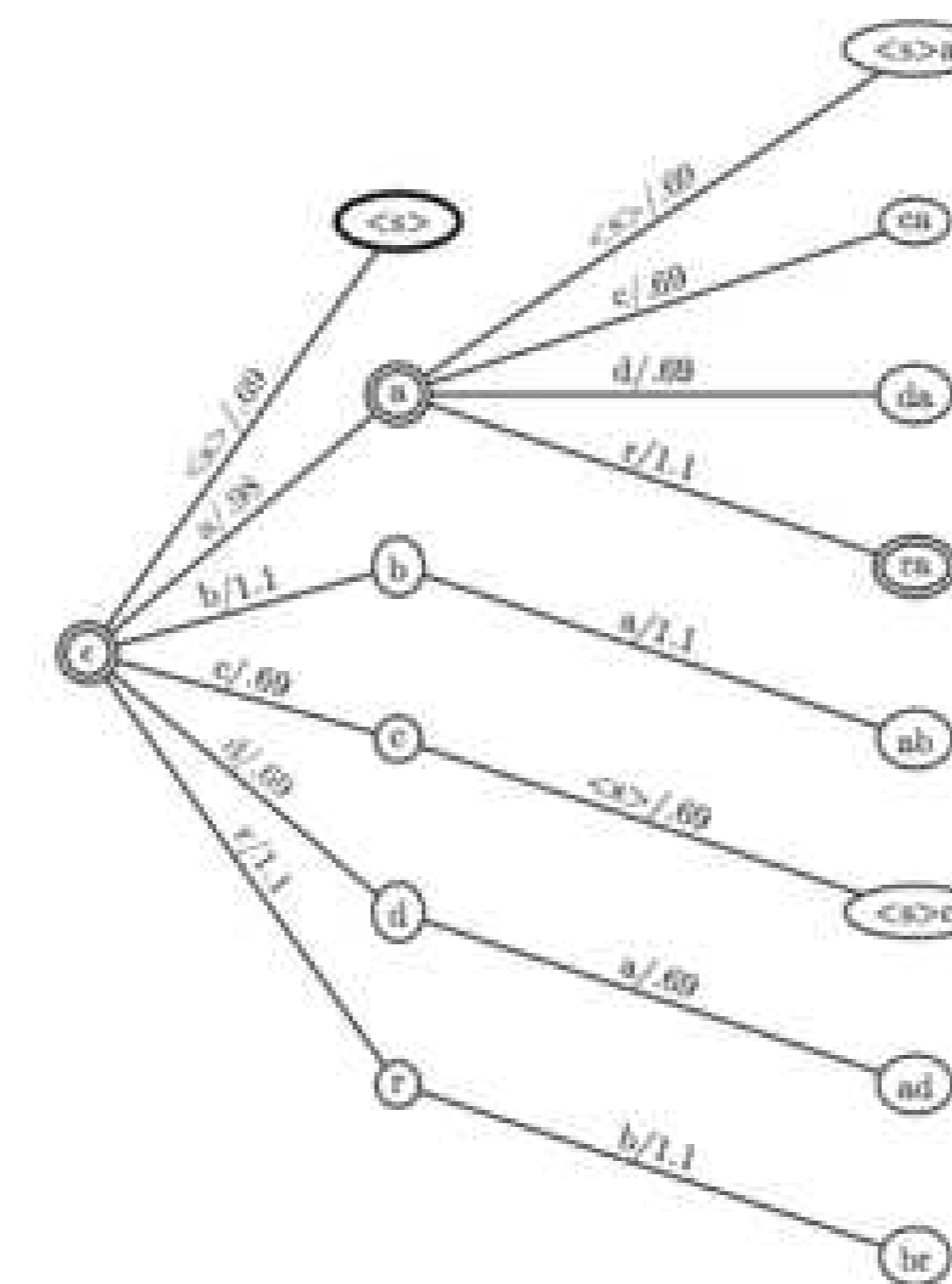
Akoestisch model

- GMM-HMM
- Getraind op spraak

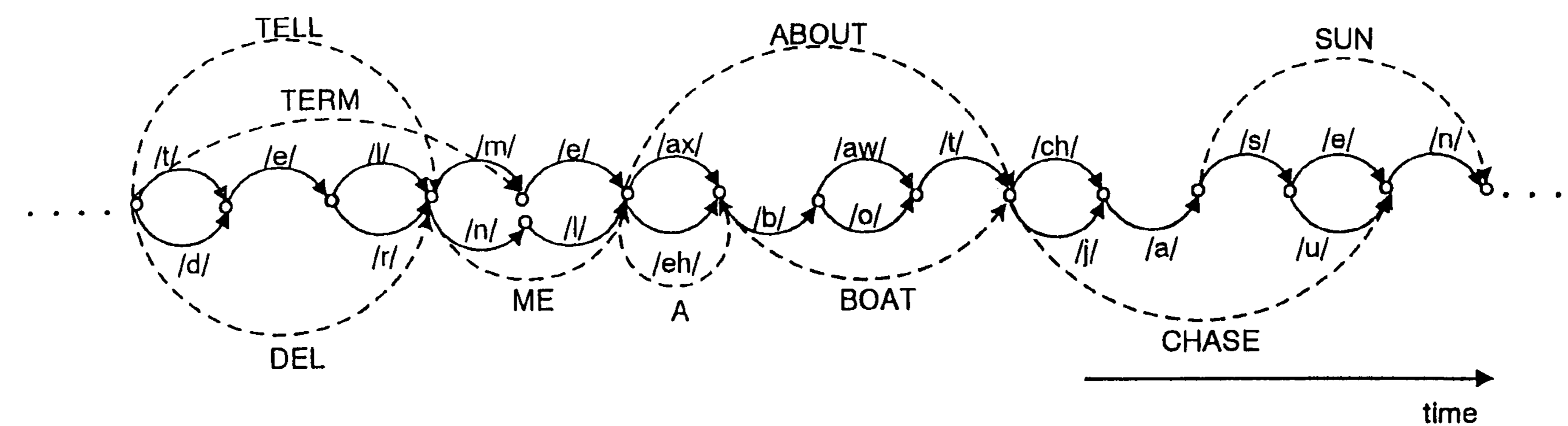


Fonetisch taalmodel

- N-gram
- Getraind op (fonetische) tekst



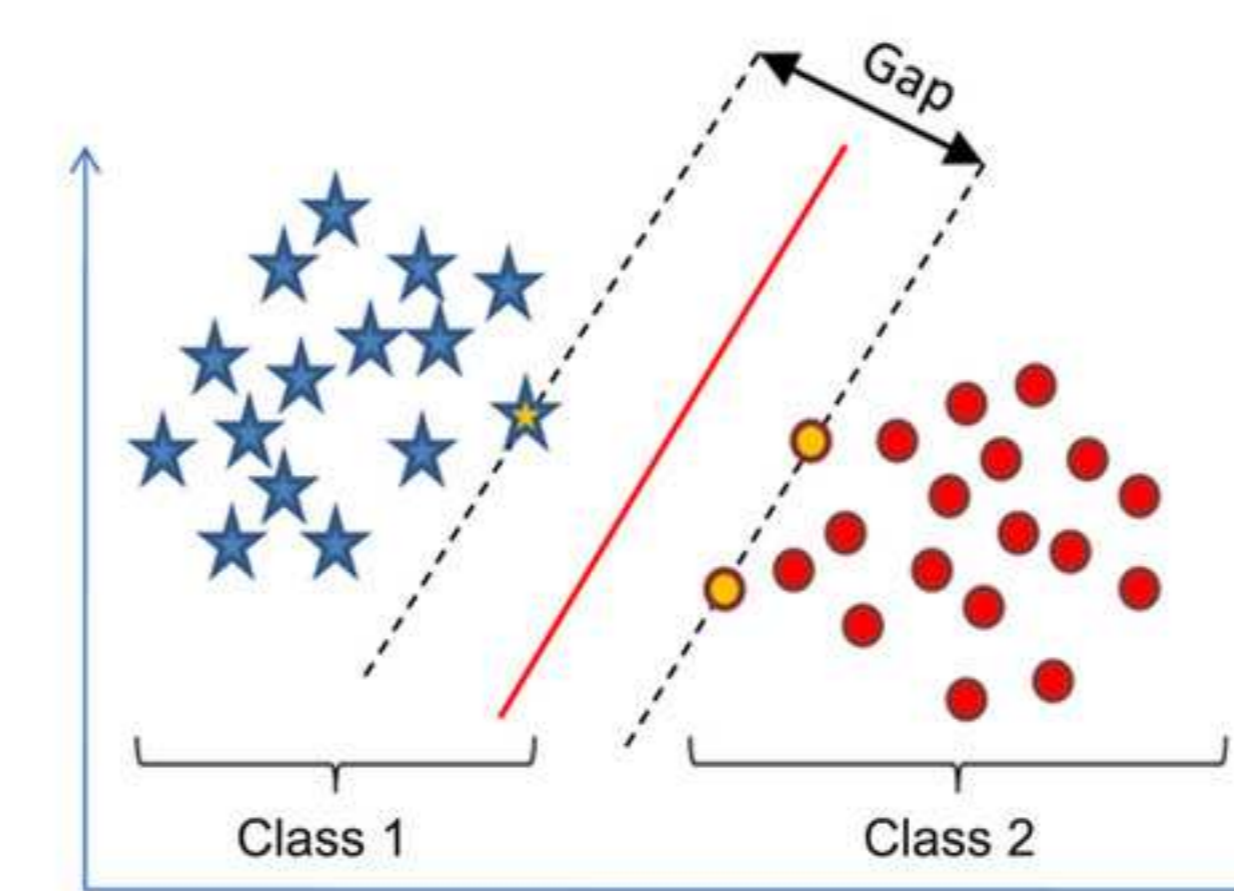
4. Spotting via lattice



- Traditionele spraakherkenner: Meest waarschijnlijke opeenvolging van woorden
- Keyword spotter: Waarschijnlijkheid voorkomen sleutelwoord in *lattice* (en threshold)

5. Classificatie

- Lijst van $N_{kw} = 74$ sleutelwoorden
- Voor elke video, tel hits elk sleutelwoord
- Per video N_{kw} -dimensionale vector
- Leer een Support Vector Machine (SVM) met haat en niet haat videos



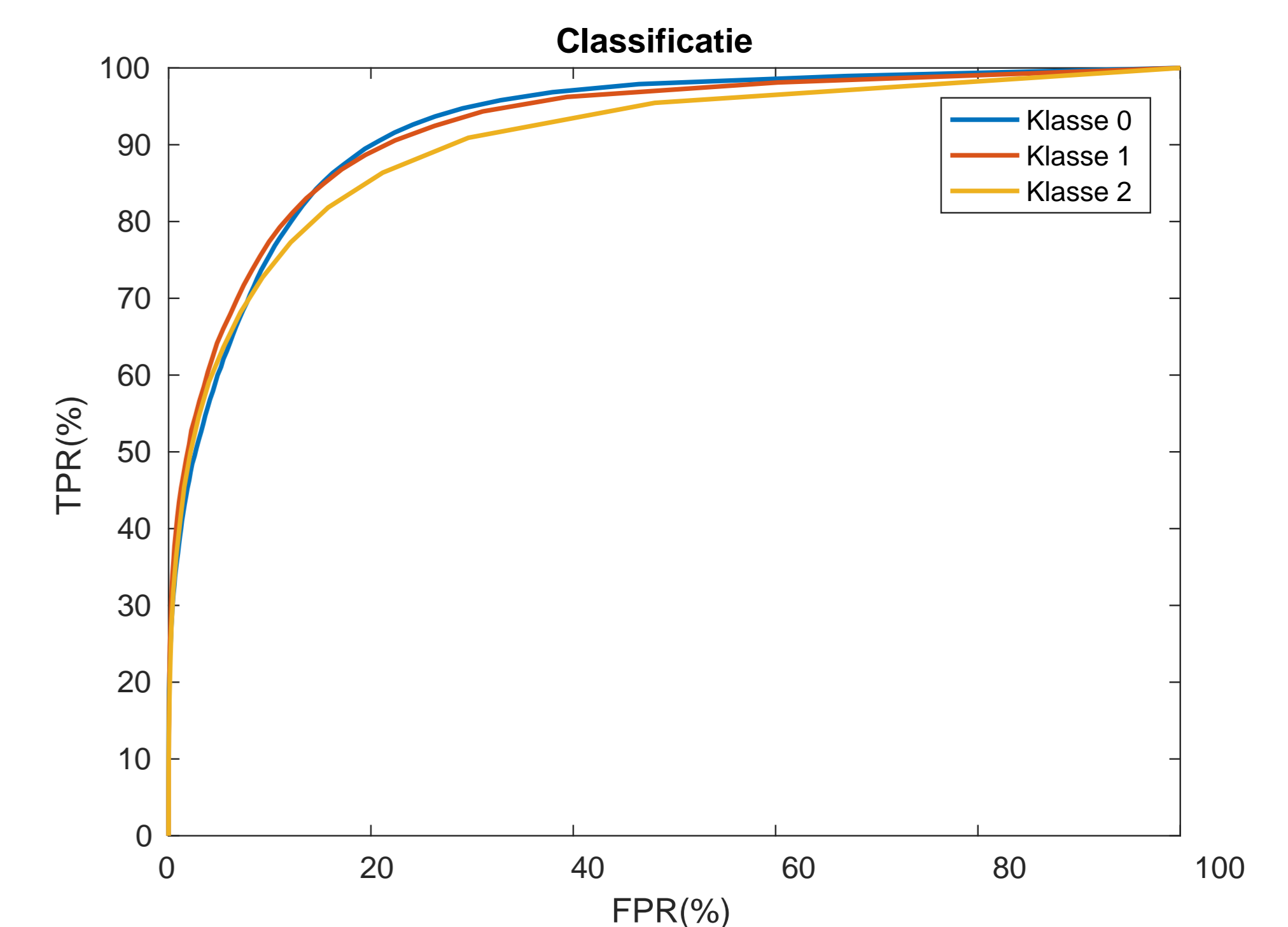
6. Resultaten spotter

- Manuele annotaties voor 10 videos
- Gemiddelde *recall* en *precision*:

	Recall(%)	Precision(%)
kws 1	61.64	0.86
kws 2	52.08	1.03
kws 3	39.48	1.35
kws 4	31.29	2.16

7. Resultaten classificatie

- Training set: 70% van data
- Test set: 30% van data
- 1000 willekeurig en onafhankelijke sessies met verschillende training/test sets
- Beste resultaten voor kws 2:



8. Conclusie

- Prestatie van spotter laag
- Veel storingsfactoren bemoeilijken spotting
- Prestatie classificatie goed
- Gemiddeld 32 scheldwoorden per video (mediaan=20.5)
- Toekomst: speech enhancement als preprocessing