Speaker Verification over Cellular Networks

Ran Gazit

Persay

PerSay
A Comverse Company

Standard Speech Corpora

• **TIMIT**	**clean**
• **YOHO**	**clean**
• **KING**	**clean**
• **SIVA**	**PSTN**
• **PolyVar**	**PSTN**
• **POLYCOST**	**ISDN**
• **Switchboard I-II**	**PSTN**
• **OGI Speaker Recognition**	**PSTN**

Campbell and Reynolds, ICASSP 99
Melin, COST 250, 2000

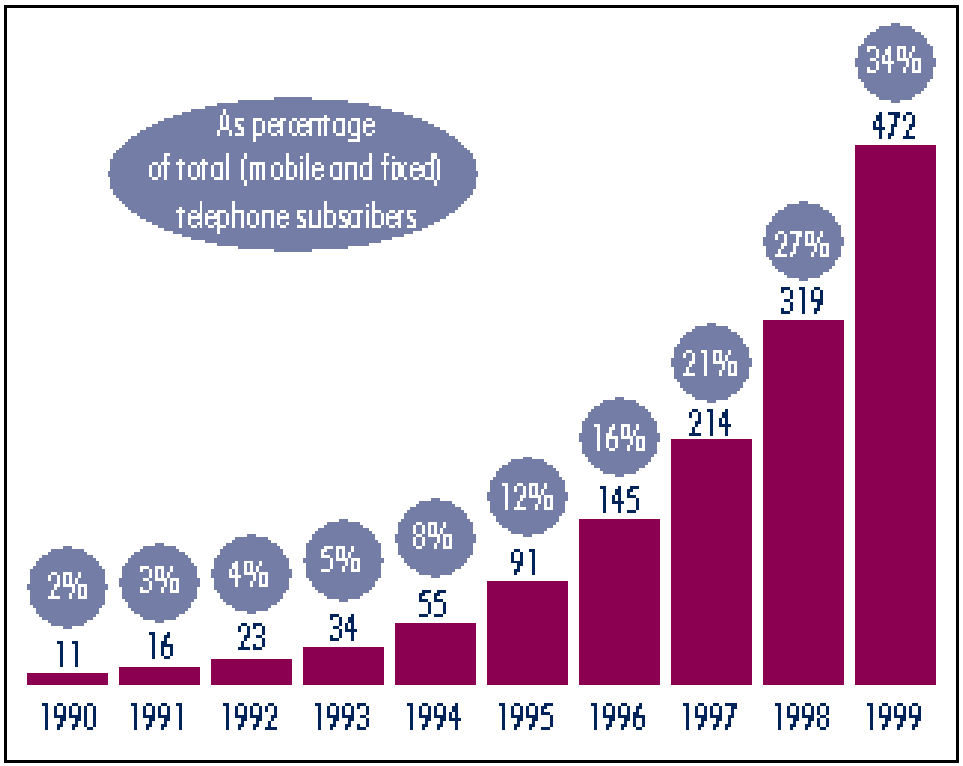
PerSay
A Comverse Company

Cellular Penetration Rate

"It seems highly likely that the number of mobile cellular subscribers will surpass conventional fixed lines during the first decade…"

ITU World Telecommunication Development Report, 1999

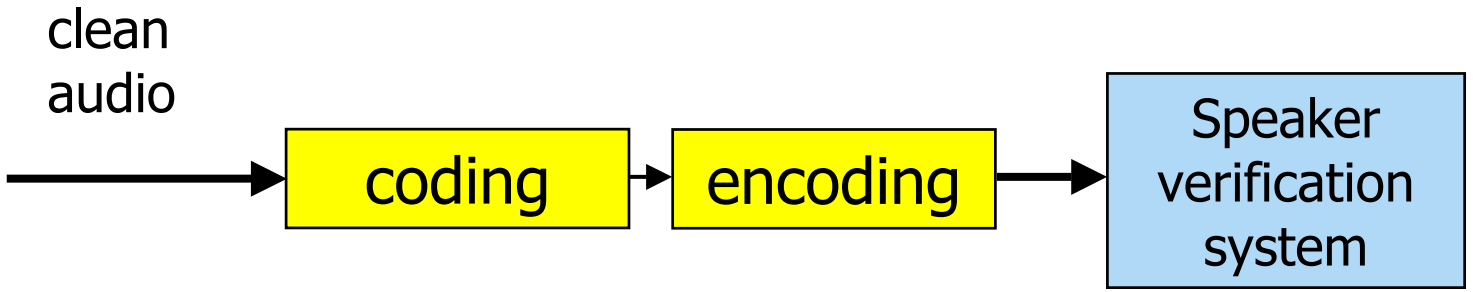
WORLDWIDE MOBILE CELLULAR SUBSCRIBERS (MILLIONS)



Source: ITU, 2000.

PerSay
A Comverse Company

Previous Work

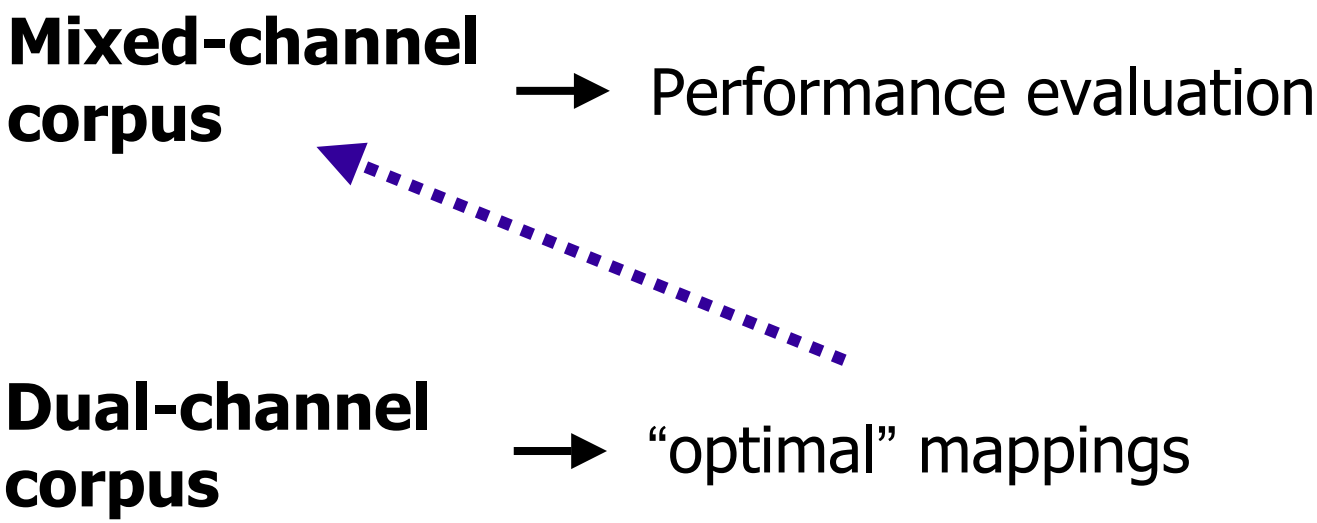


Kuitert and Boves, Eurospeech 97 (SESP - GSM)

Quatieri et. al., Eurospeech 99 (Swbrd - GSM, G.729, G.723.1)

Besacier et. al., ICASSP 2000 (TIMIT - GSM)

Method



PerSay
A Comverse Company

Mixed-Channel Corpus

Target speakers	43F / 42M 10 1 min. Land-line 10 1 min. Cellular (various)
Background speakers	60F / 60M 1 min Land-line 1 min Cellular (various)

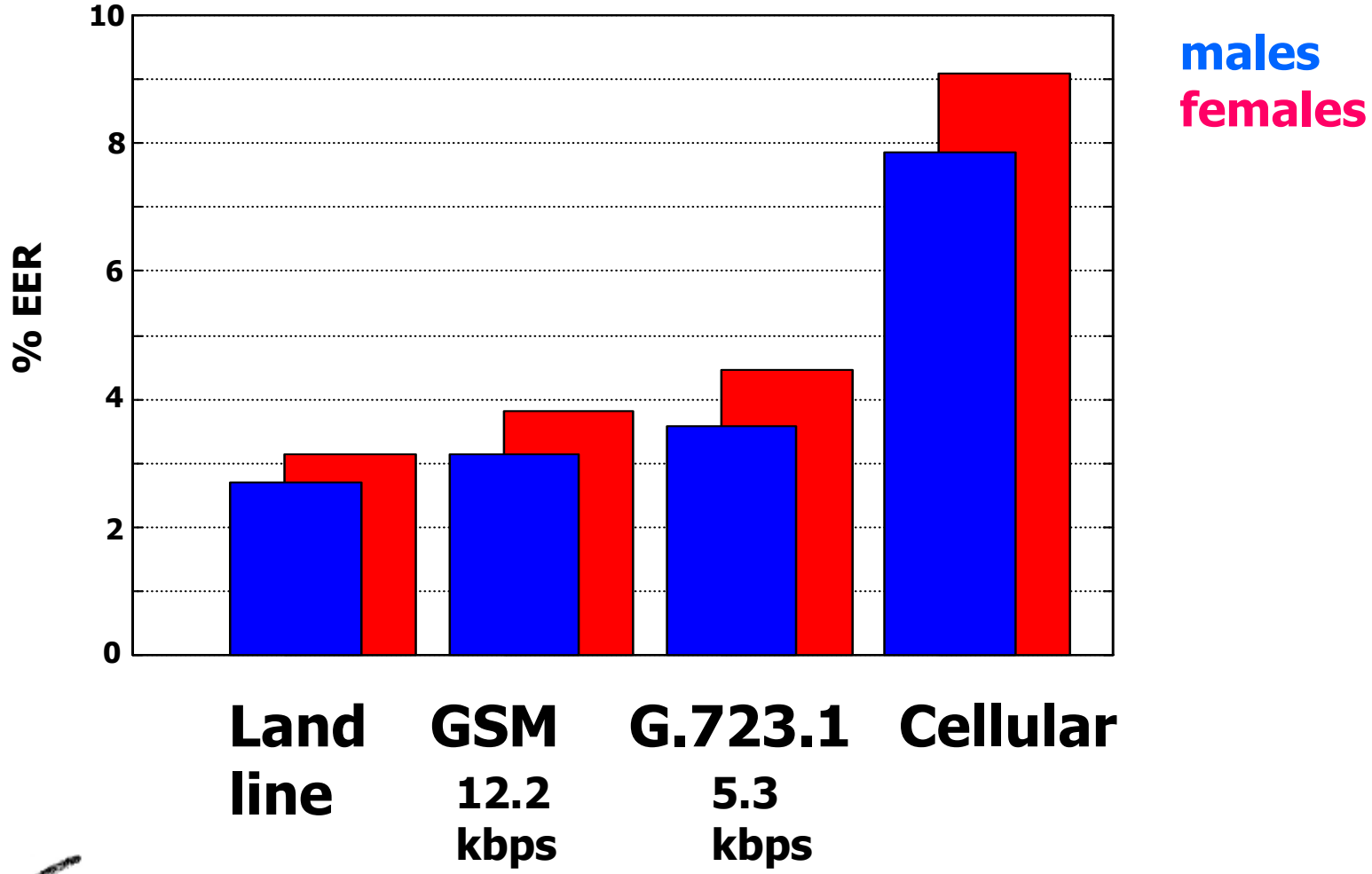
PerSay
A Comverse Company

Experiment Setting

Working point	3 min. train 30 sec. test
System	25 ms frames, 12.5 ms shift Energy-based VOX 12 LPC-Ceps, CMS 12 delta-Ceps 30-mixture GMM Gender,channel-dependent BM

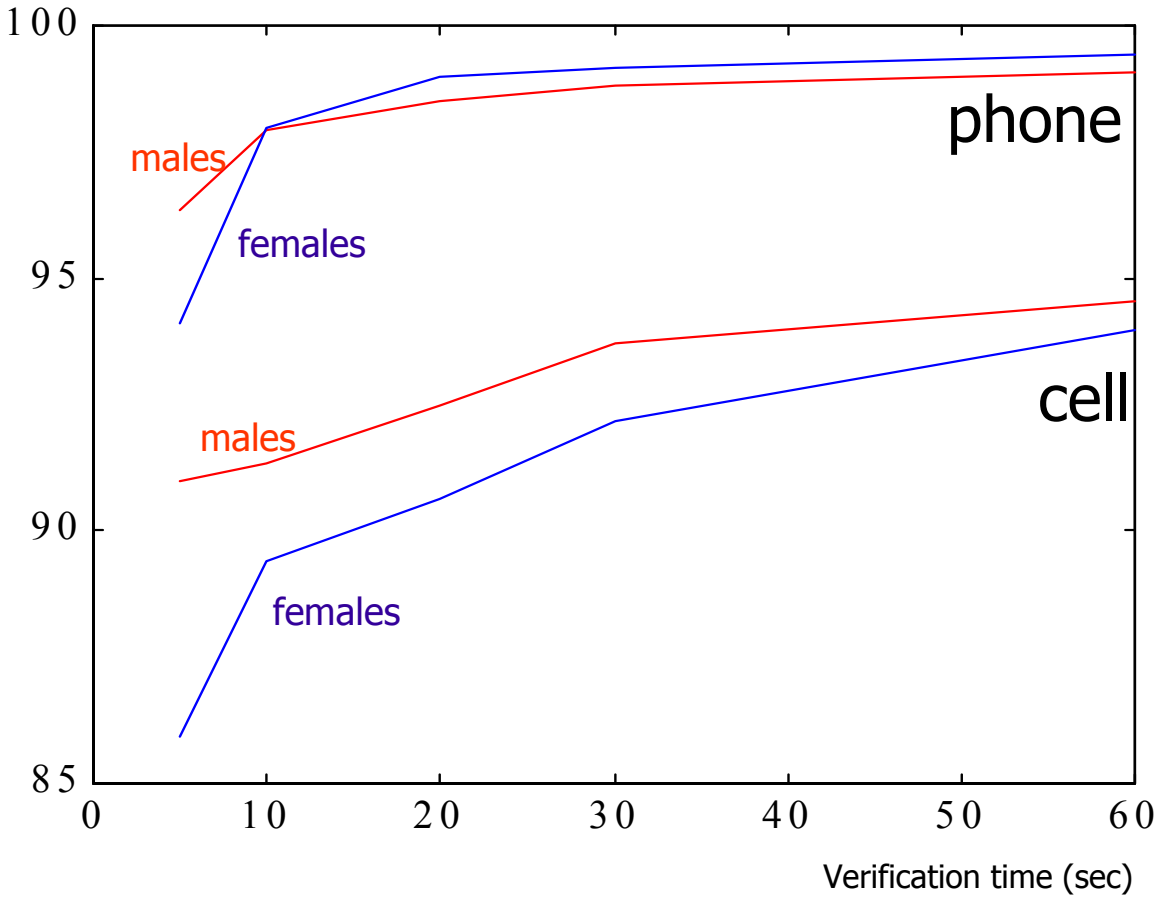
PerSay
A Comverse Company

Matched-Conditions EER



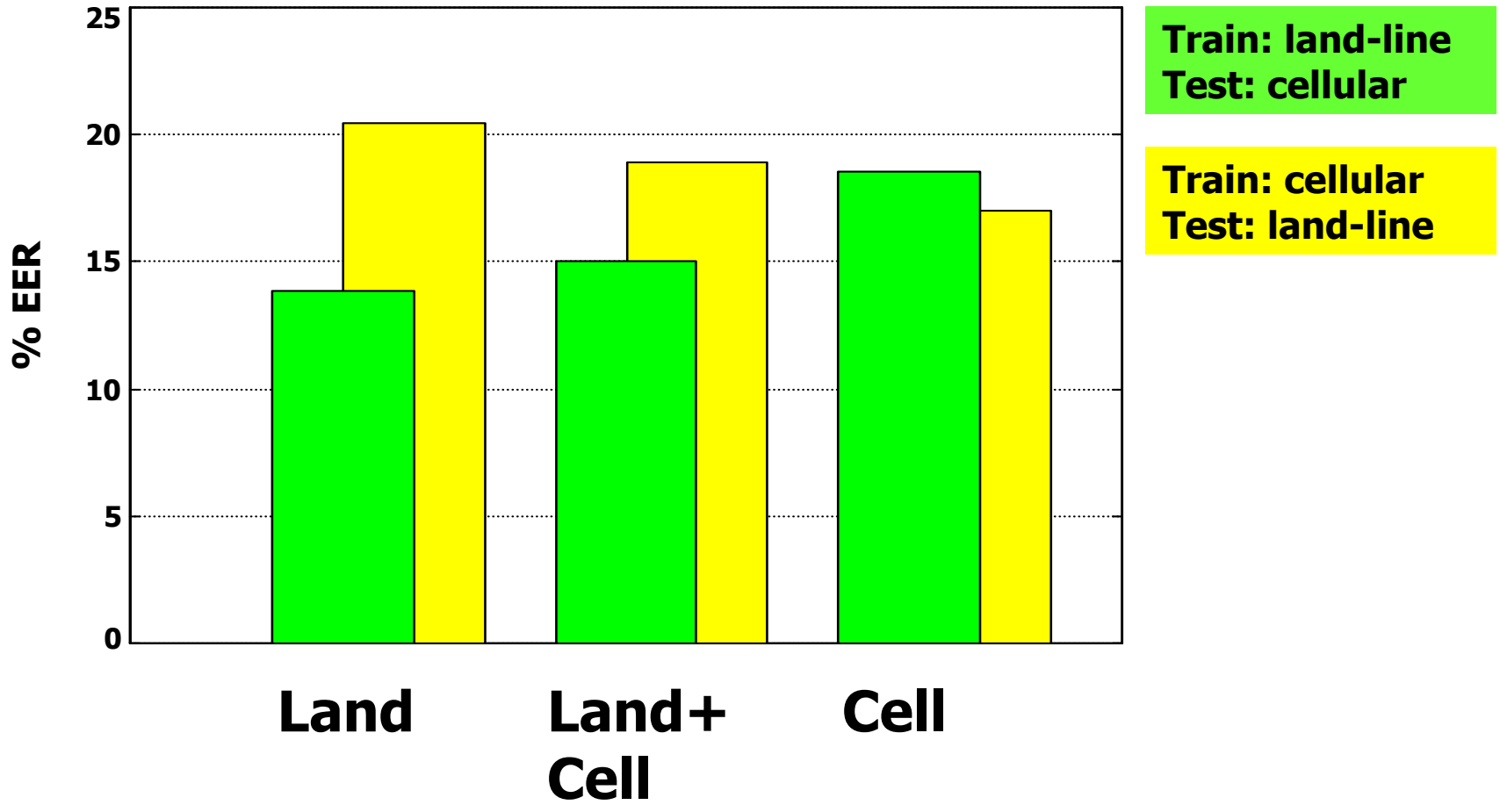
PerSay
A Comverse Company

Confidence Level



PerSay
A Comverse Company

Mismatched-Conditions EER



PerSay
A Comverse Company

Dual-Channel Corpus

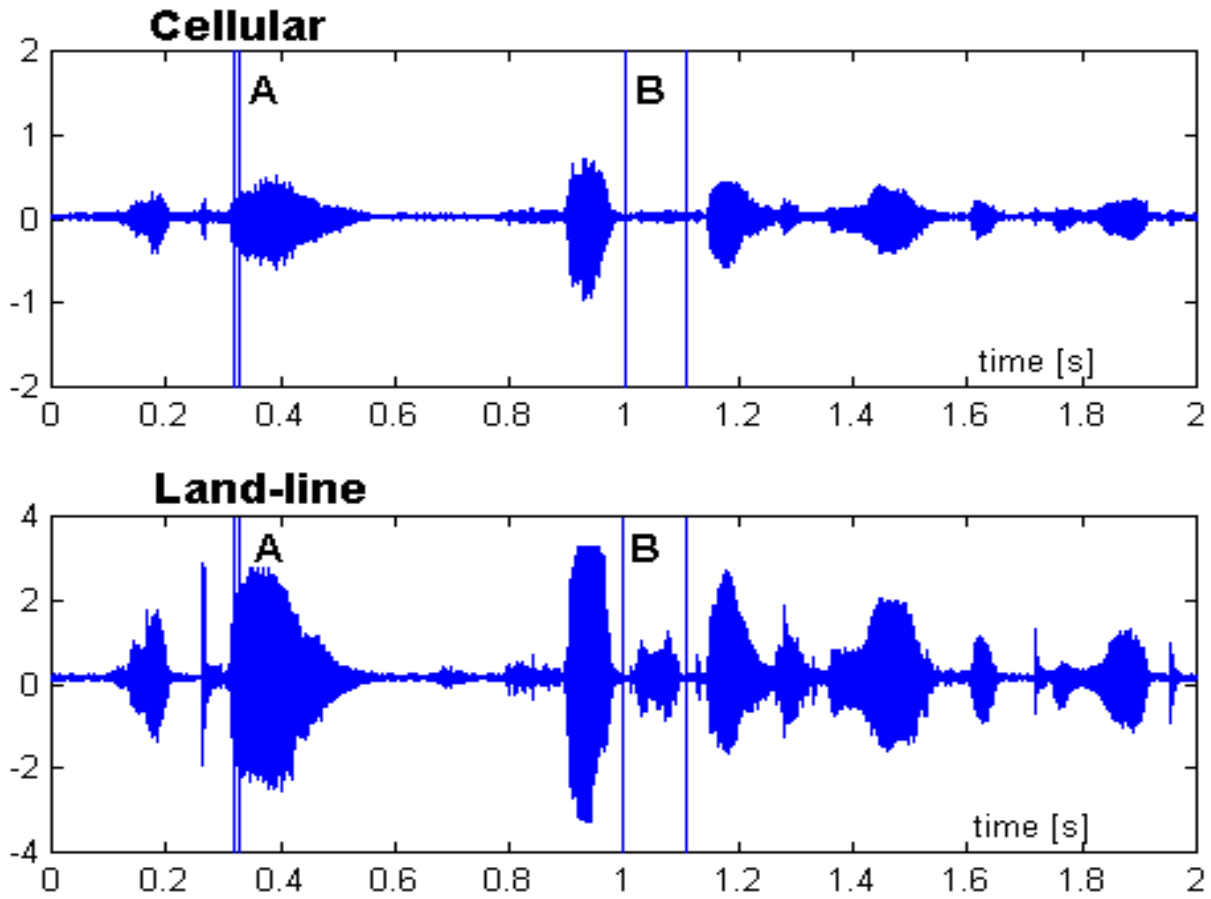
Target speakers

10F / 10M

1 min. Land-line & Cellular simultaneously

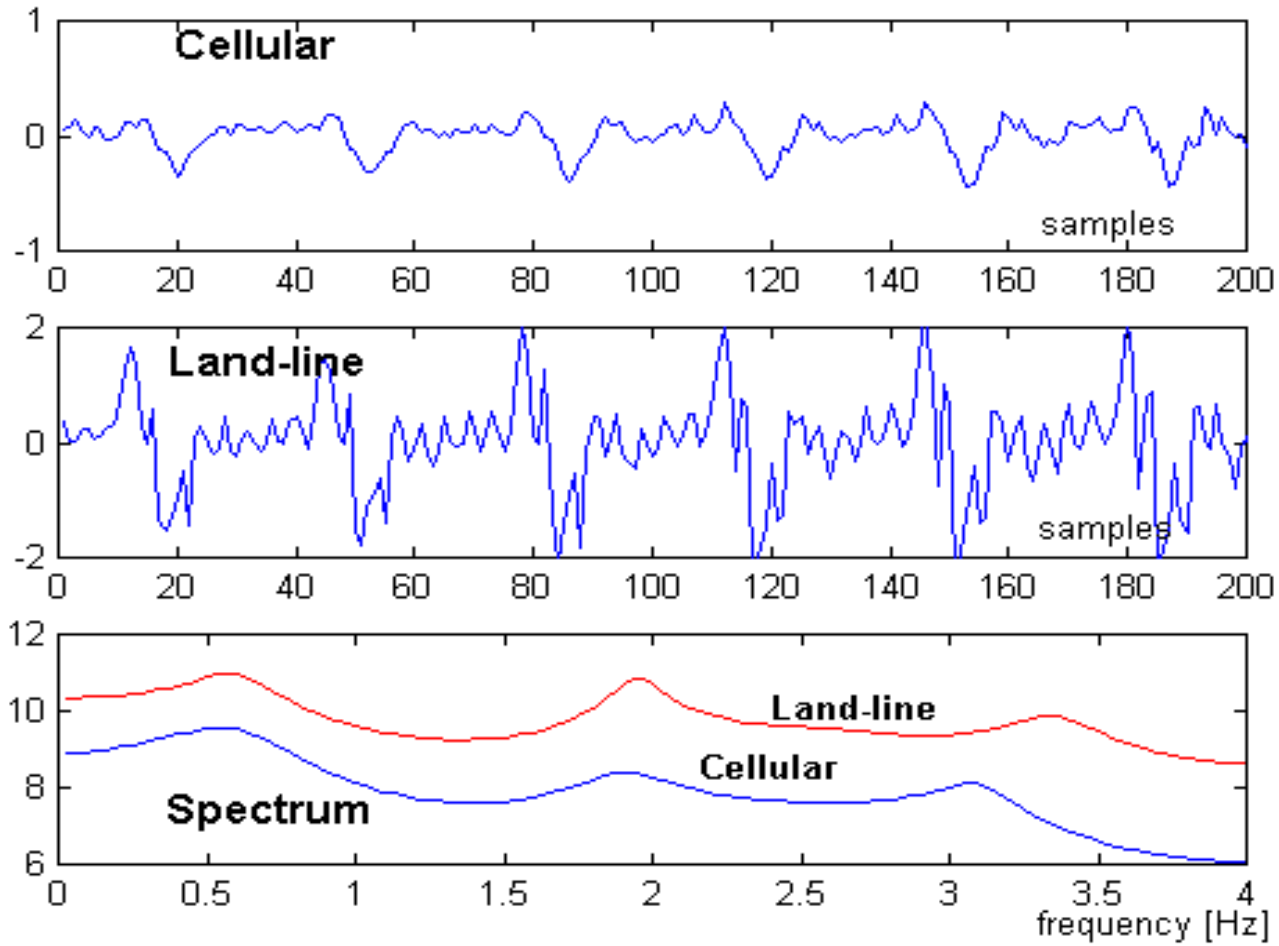
PerSay
A Comverse Company

Dual-media Speech Signal



PerSay
A Comverse Company

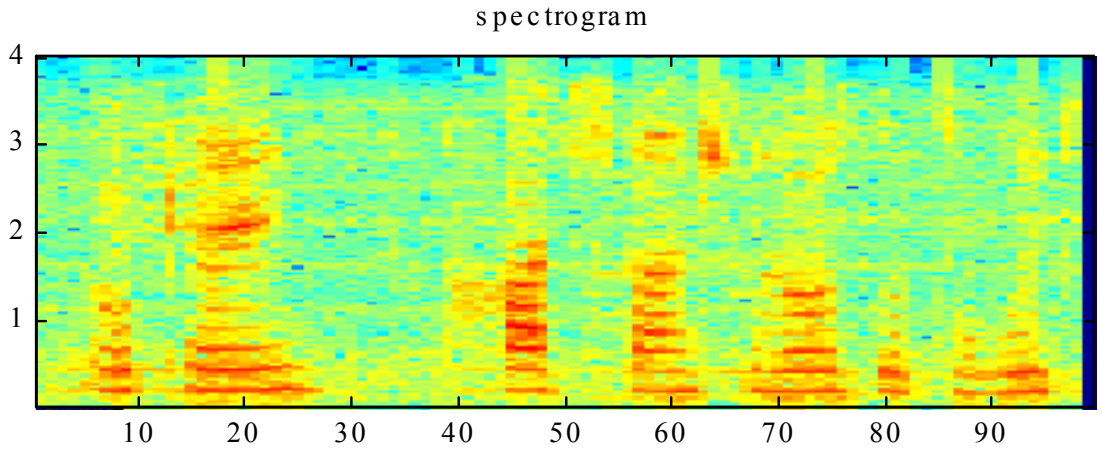
Dual-Media Speech Signal



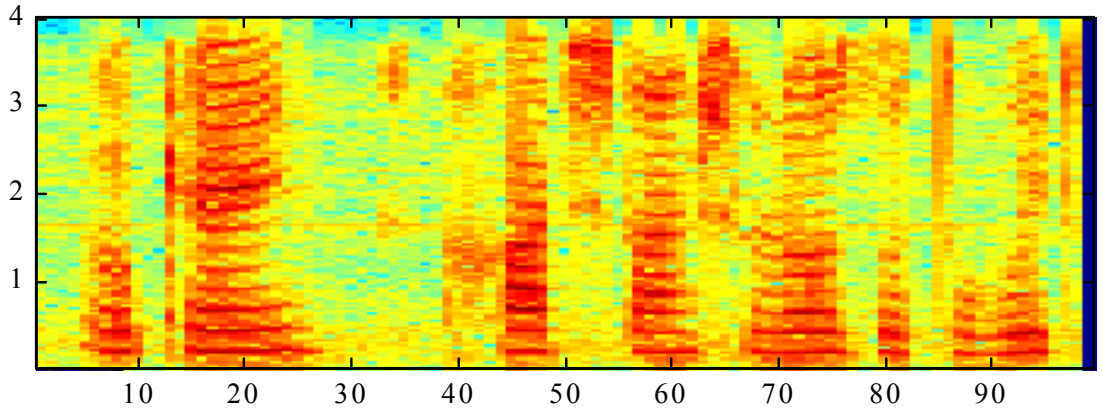
PerSay
A Comverse Company

Dual-media Spectrogram

cell

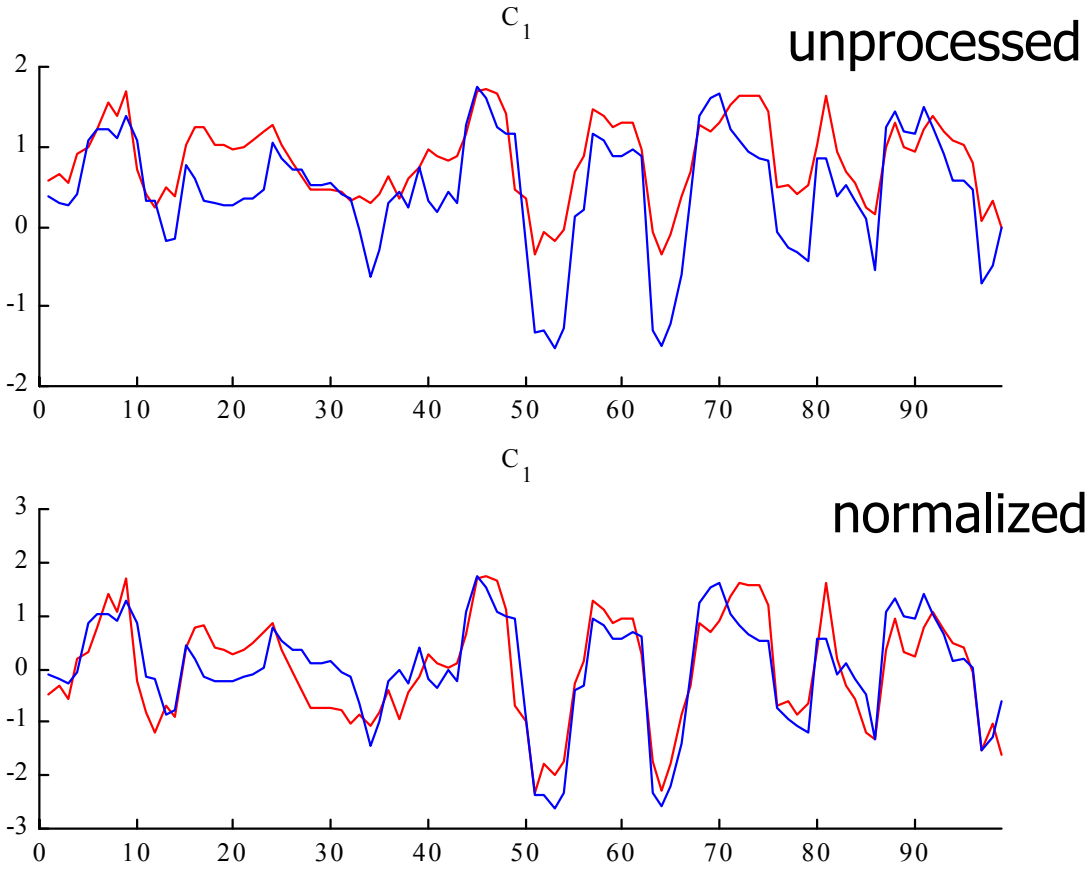


phone



PerSay
A Comverse Company

Dual-media Cepstrum Track

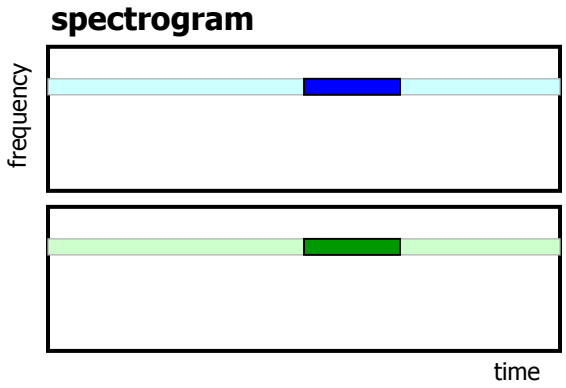


PerSay
A Comverse Company

Dual-Channel Mappings

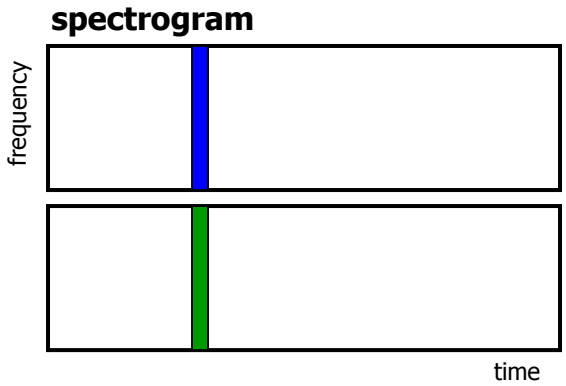
Goal:

- Suppress channel variability
- Preserve phonetic variability



Methods:

- Wiener-like FIR filters
- Spectral basis vectors

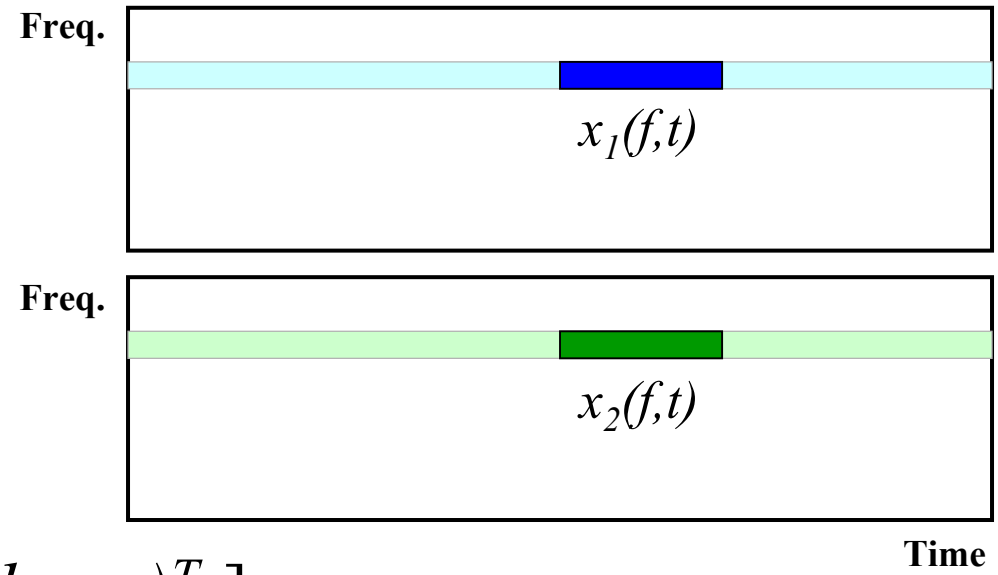


Hermansky et. al., ICASSP 95
Malayath et. al., DSP journal, 2000

PerSay
A Comverse Company

Data-Driven Temporal Filters

$$Y = X^T h$$



$$d = x_1 - x_2$$

$$\Sigma_{ch} = \mathrm{E}\left[(d - \mu_d)(d - \mu_d)^T \right]$$

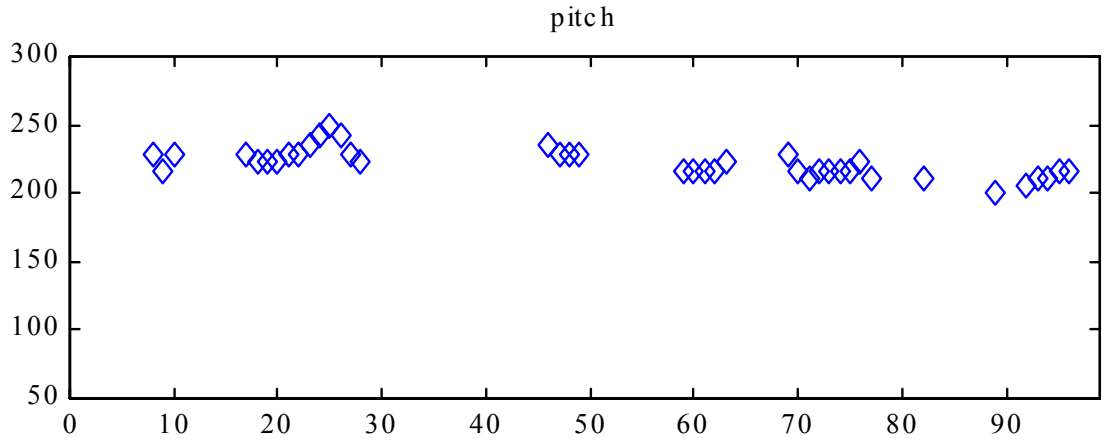
$$\Sigma_p = \mathrm{E}\left[(\mu_p - \mu)(\mu_p - \mu)^T \right]$$

$$\left[\Sigma_{ch}^{-1} \Sigma_p \right] e = \lambda e$$

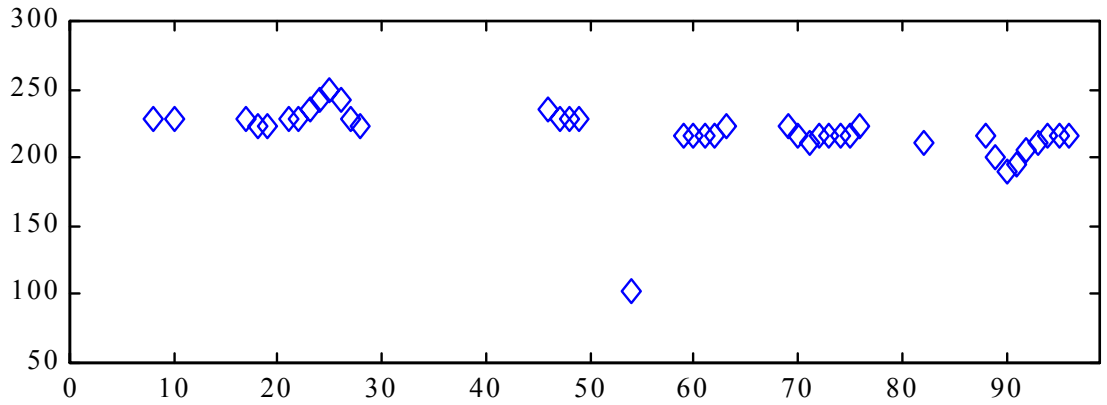
PerSay
A Comverse Company

Dual-media Pitch Track

cell



phone



PerSay
A Comverse Company

Cellular Effects

- Spectral and temporal distortions
- Noisy environment
- Speaking style
- Cutoffs
- Miniature microphone

PerSay
A Comverse Company