ITC-irst

Center for Scientific and Technological Research

38050 Povo (Trento), ITALY

ROBUST F0 ESTIMATION BASED ON A MULTICHANNEL PERIODICITY FUNCTION FOR DISTANT-TALKING SPEECH

F. Flego, M. Omologo

AISV 2005 - Salerno

30^{th} November - 2^{nd} December

Outline

- Acoustic scene analysis for ambient intelligence
  - Principal dimensions of distant-talking analysis and ASR
  - Limitations of the common approach
  - Distributed Microphone Network
  - CHIL room for seminars and meetings (CHIL European Project)
  - Current IRST activities
- Robust F0 estimation
  - Single channel algorithms: WAUTOC, YIN
  - Multichannel algorithms: WAUTOC, YIN
  - Multichannel Periodicity Function (MPF)
- Preliminary results
  - Multichannel corpus derived by the Keele database
- Conclusions and future perspectives

Principal dimensions of distant-talking analysis and ASR

- Type and location of microphones
- Talker position, direction, and head rotation
- Talker speech clarity, SNR
- Possible talker position changes
- Environmental noise and acoustics
  - Echoes and reverberation
  - Noise level, noise source positions
  - Diffuse or coherent noise (stationary vs unstationary)
  - From babble noise to competitive talkers

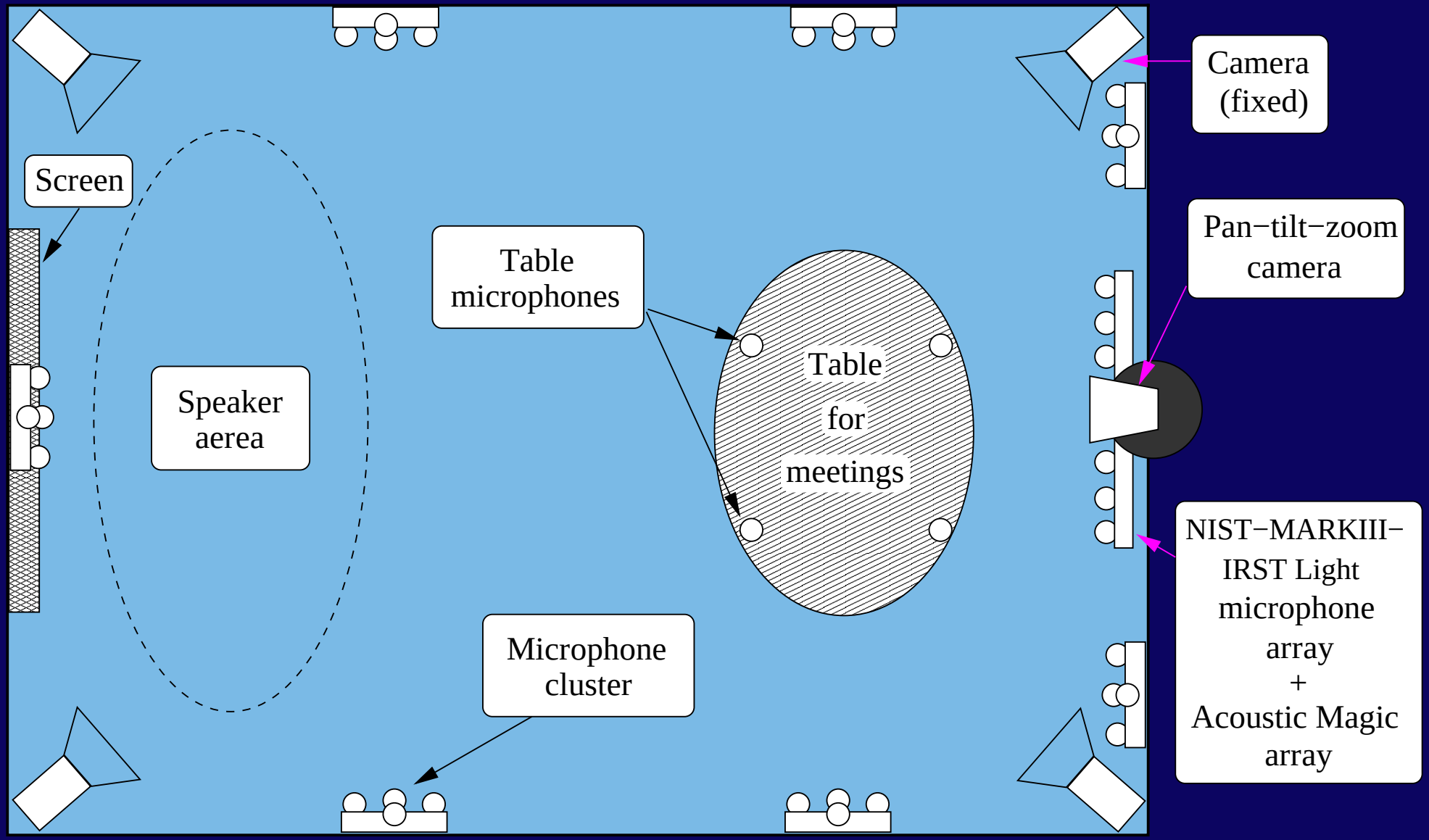
Limitations of the common approach

- Need of a rather controlled acoustic environment
  - Limit reverberation and noise
- Assumption to have one speaker (vs competitive speakers)
- Head orientation and speech clarity dependency
- Room coverage limitations
  - Microphone arrays placement constrained by room geometry
  - Placed sensors may result intrusive
  - Need of expensive microphones/technology

Distributed Microphone Network

- Ubiquitous sensoring
  - Wall-mounted T-shaped (low-cost) microphone clusters
  - Microphone arrays
  - Table microphones, etc
- Main objective
  - Analysis and interpretation of the scenario
  - Integration with visual information provided by cameras
- Possible applicative contexts:
  - Security and surveillance
  - Videoconferencing
  - Smart home (television control, elderly and disabled assistance, etc)
  - Meeting rooms, lecture rooms (CHIL project)

CHIL room for seminars and meetings

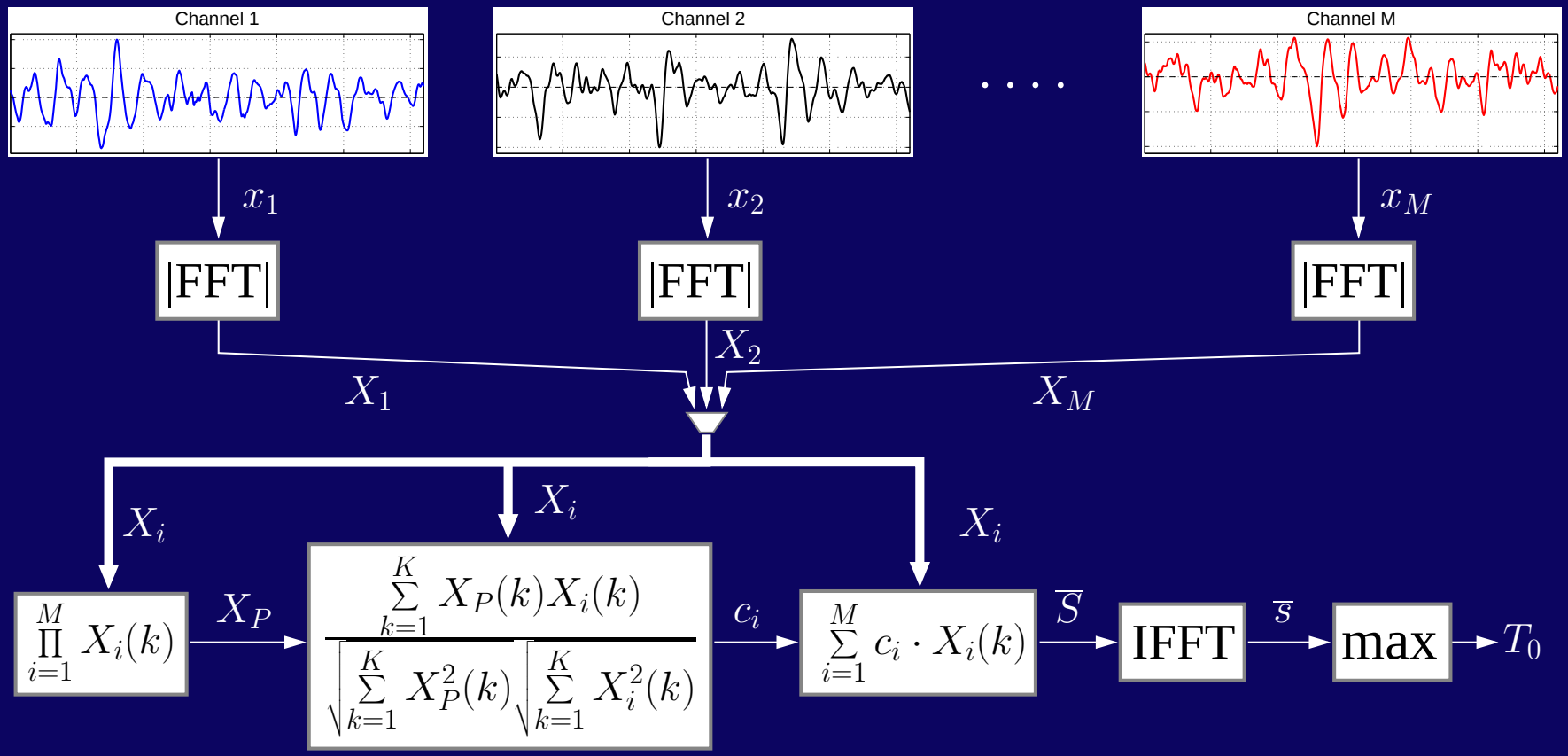


Current IRST activities

- Speech activity detection
- Acoustic Event Detection (AED)
- Speaker(s) localization (tracking position and orientation)
- Distant-talking ASR
- Audio-video integration, person identification, etc

- Multi-microphone pitch analysis

MPF - Multichannel Periodicity Function

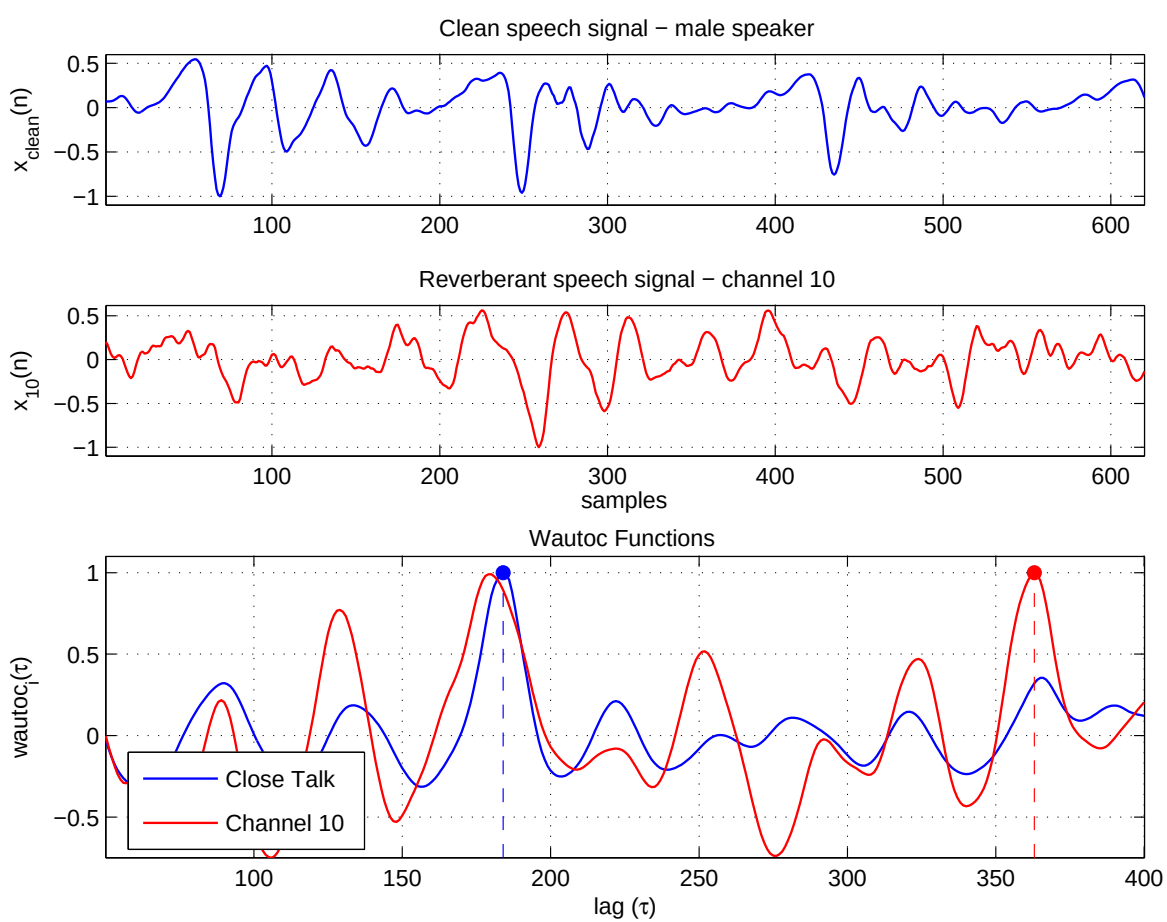


- $x_i = i$-th channel signal
- $X_i = i$-th channel |FFT|
- X_P = reference spectrum
- $c_i = i$-th channel weight
- $\overline{s}$ = MPF
- T_0 = estimated period

Weighted Autocorrelation (single channel)

$$wautoc_i(\tau) = \frac{\sum_{n=0}^{N-\tau-1} x_i(n)x_i(n+\tau)}{\sum_{n=0}^{N-\tau-1} |x_i(n) - x_i(n+\tau)| + \epsilon}$$

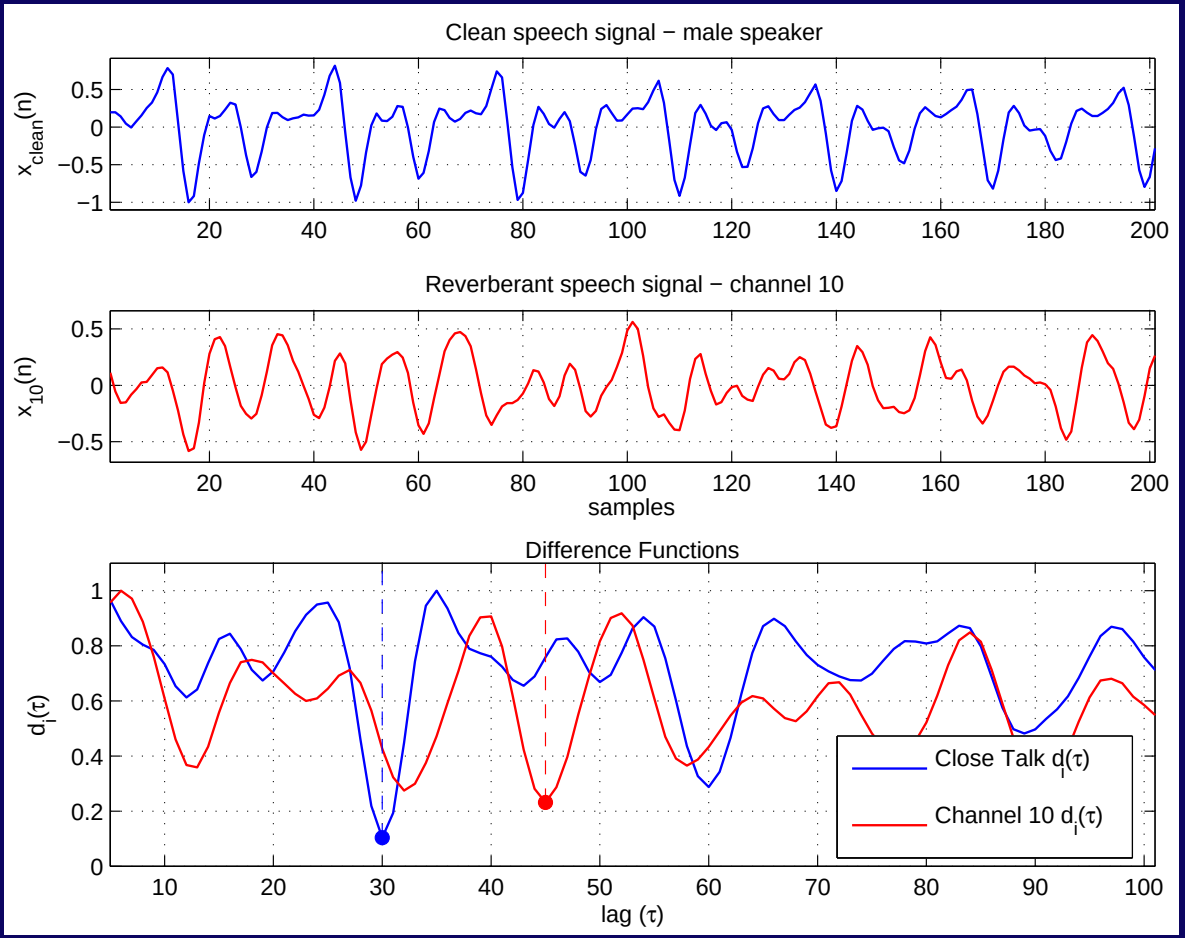
T. Shimamura
H. Kobayashi



YIN - Cumulative Mean Normalized Difference function (single channel)

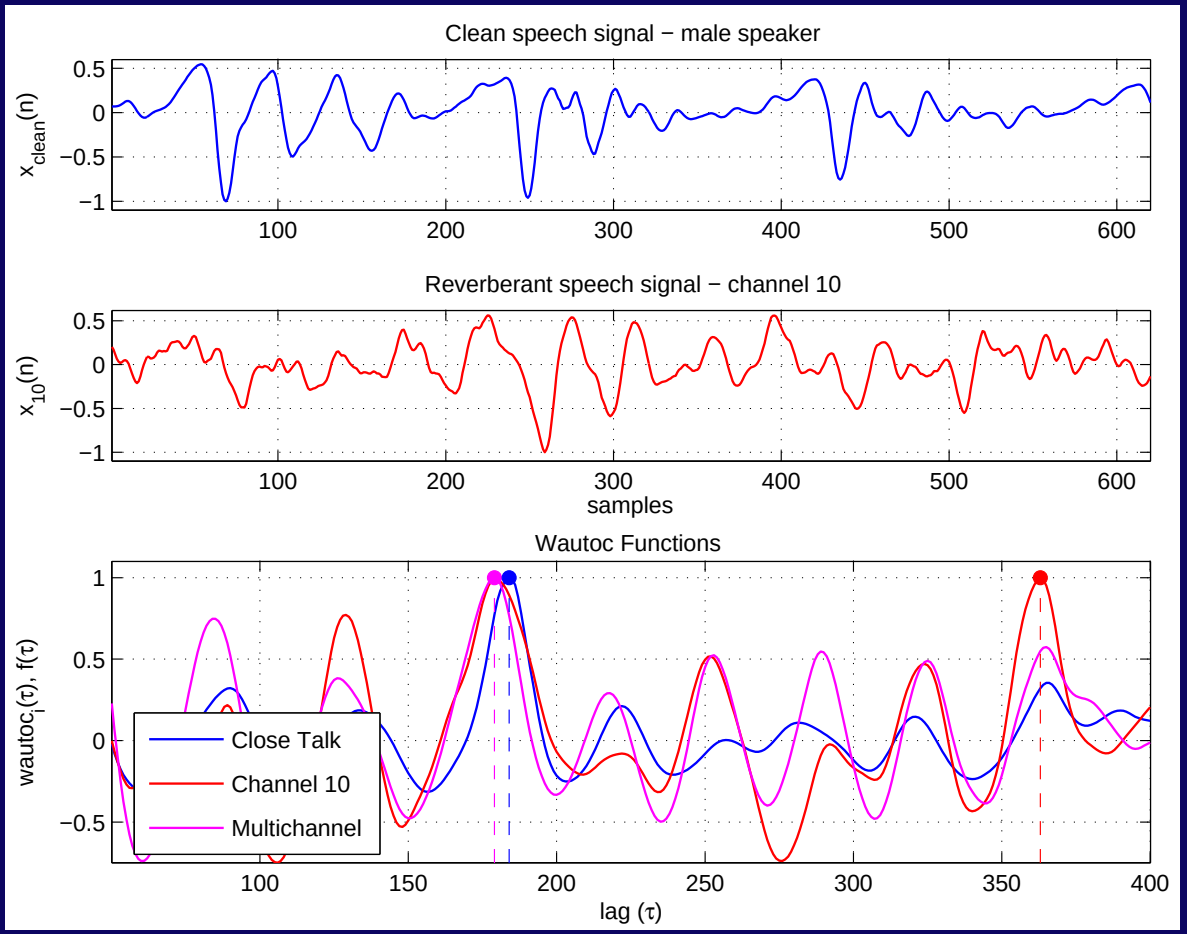
$$d_i(\tau) = \sum_n [x_i(n) - x_i(n+\tau)]^2, \quad d'_i(\tau) = \begin{cases} 1, & \text{if } \tau = 0, \\ d_i(\tau)/[(1/\tau)\sum_{j=1}^{\tau} d_i(j)] \end{cases}$$

A. de Cheveigné
H. Kawahara



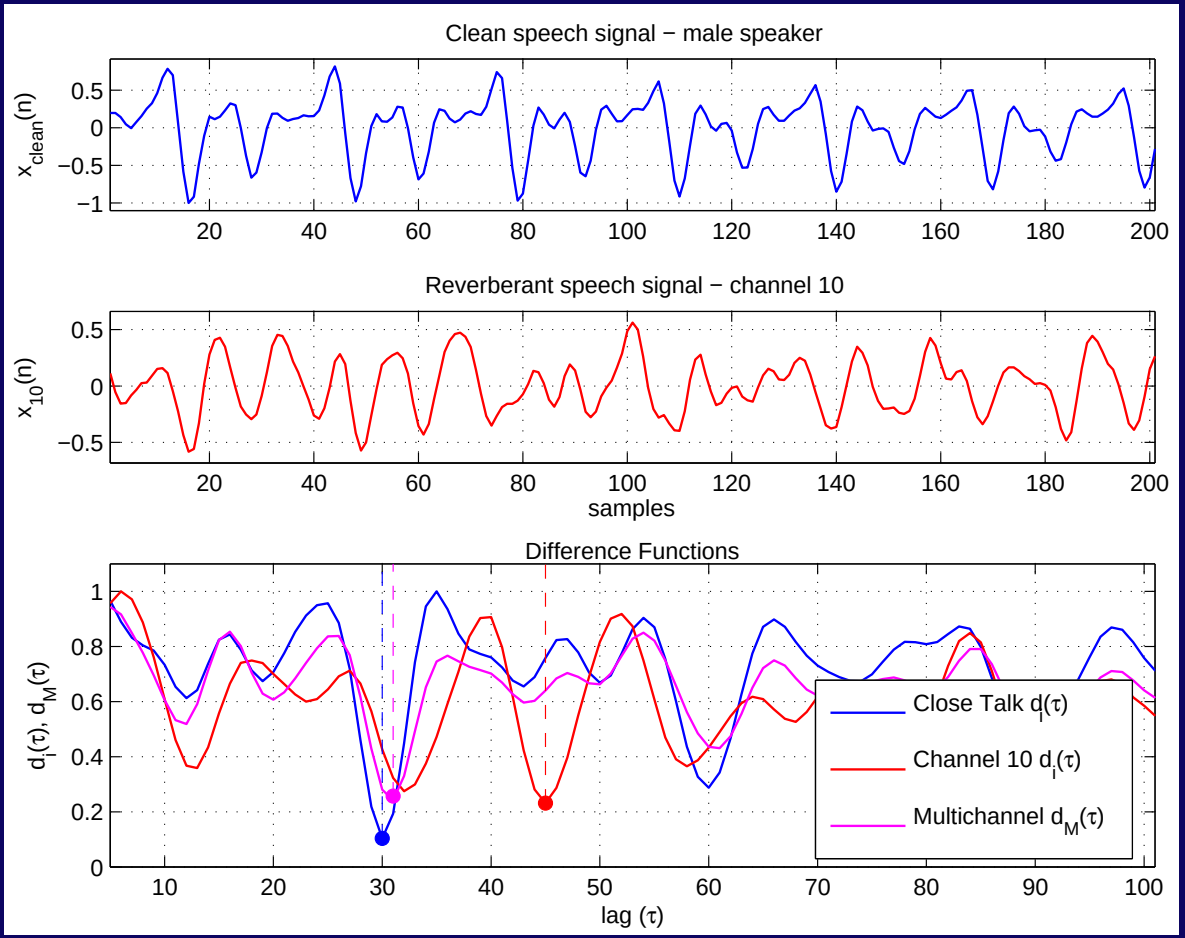
Weighted Autocorrelation (multichannel extension)

$$wautoc_i(\tau) = \frac{\sum_{n=0}^{N-\tau-1} x_i(n)x_i(n+\tau)}{\sum_{n=0}^{N-\tau-1} |x_i(n) - x_i(n+\tau)| + \epsilon} \quad \blacktriangleright\blacktriangleright\blacktriangleright \quad f(\tau) = \sum_{i=1}^{M} wautoc_i(\tau)$$



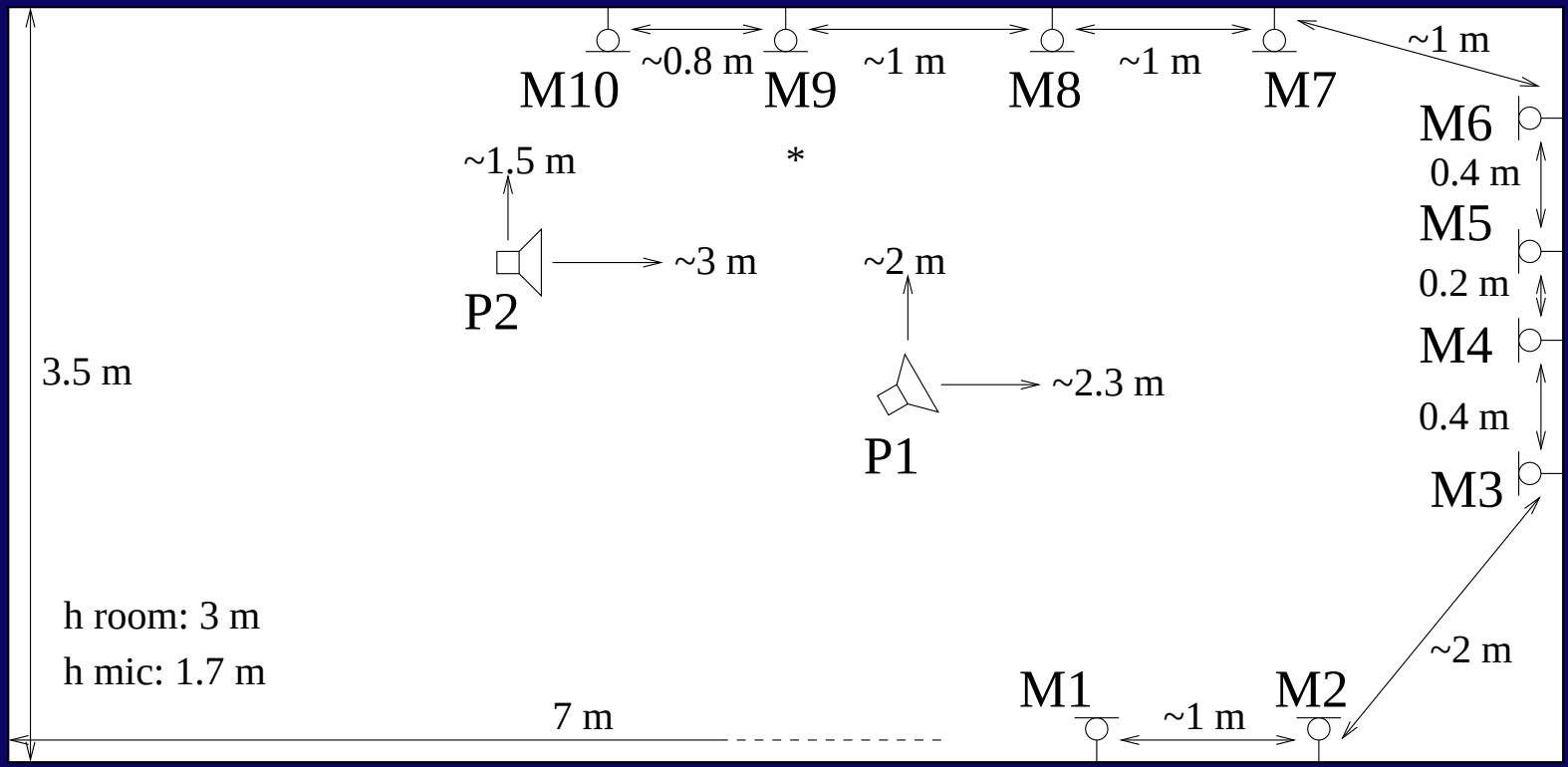
YIN (multichannel extension)

$$d_M(\tau) = \frac{1}{M}\sum_{i=1}^{M} d_i(\tau) \quad \blacktriangleright\blacktriangleright\blacktriangleright \quad d''_i(\tau) = \begin{cases} 1, & \text{if } \tau = 0, \\ d_M(\tau)/[(1/\tau)\sum_{j=1}^{\tau} d_M(j)] \end{cases}$$

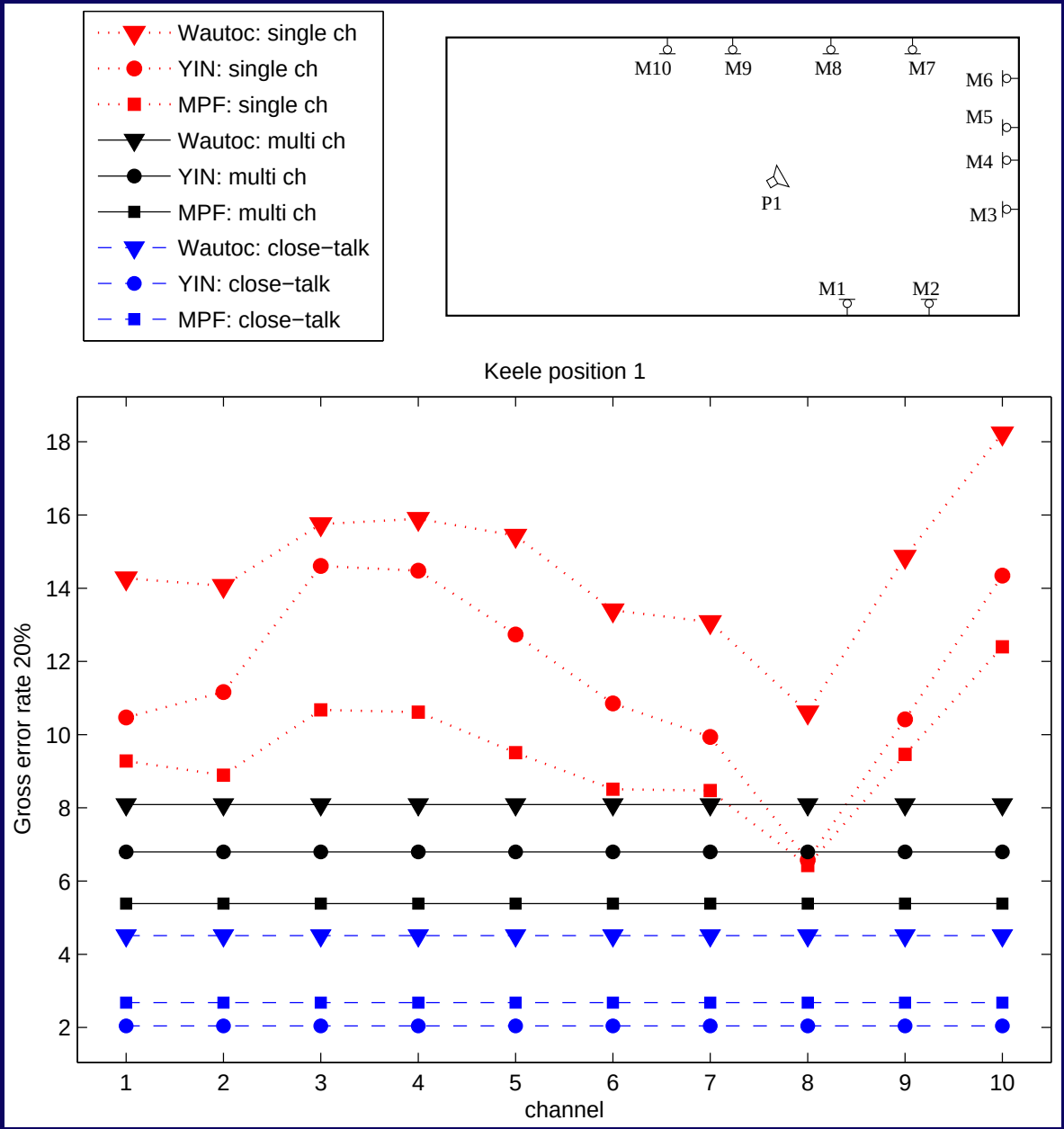


Multichannel corpus derived by the Keele database

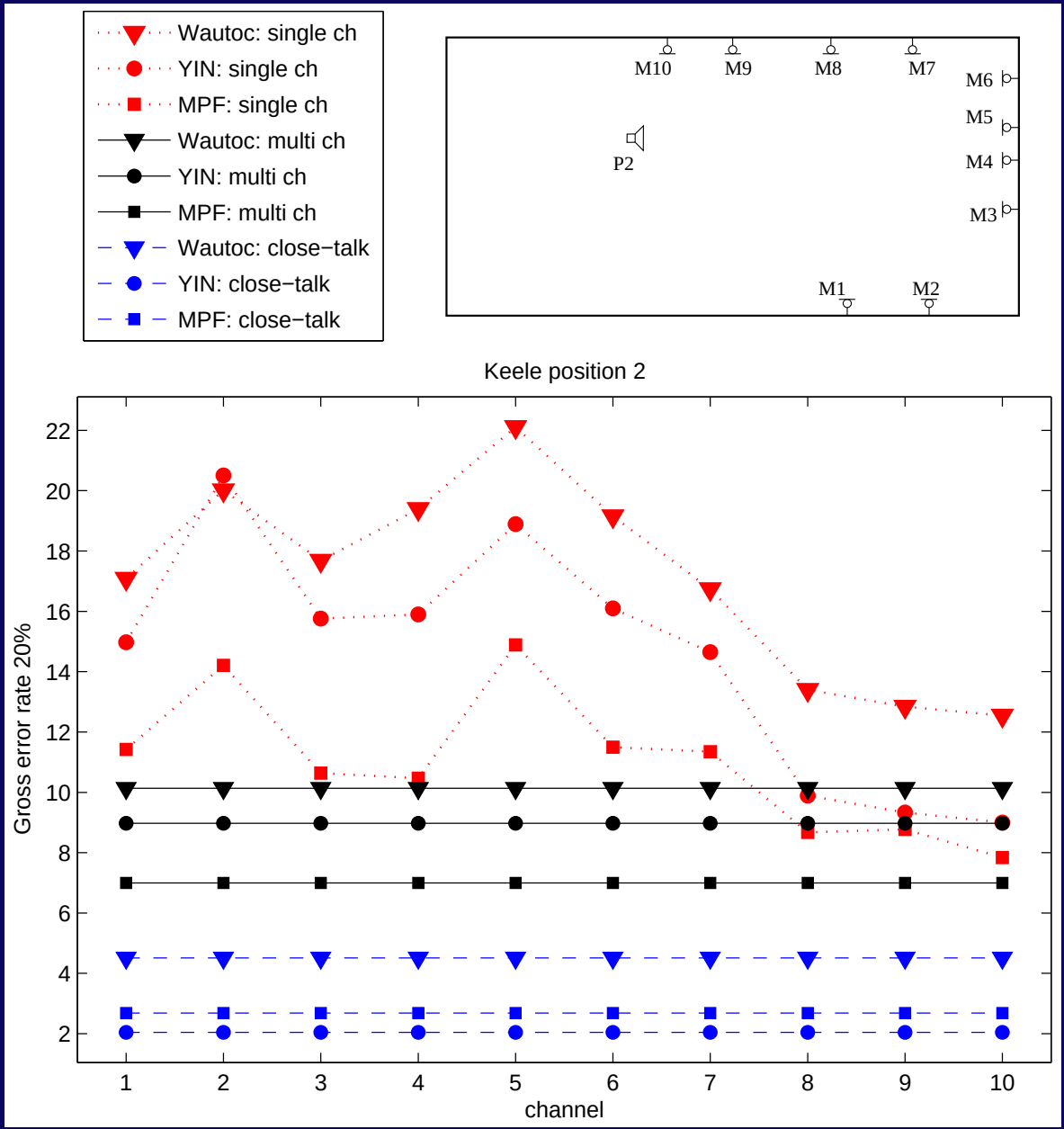
- Keele Database reproduced in position P1 and P2
- Recorded by 10 omnidirectional microphones
- Pitch references extracted from laringograph signal
- Reverberation time $T_{60} \simeq 0.35s$



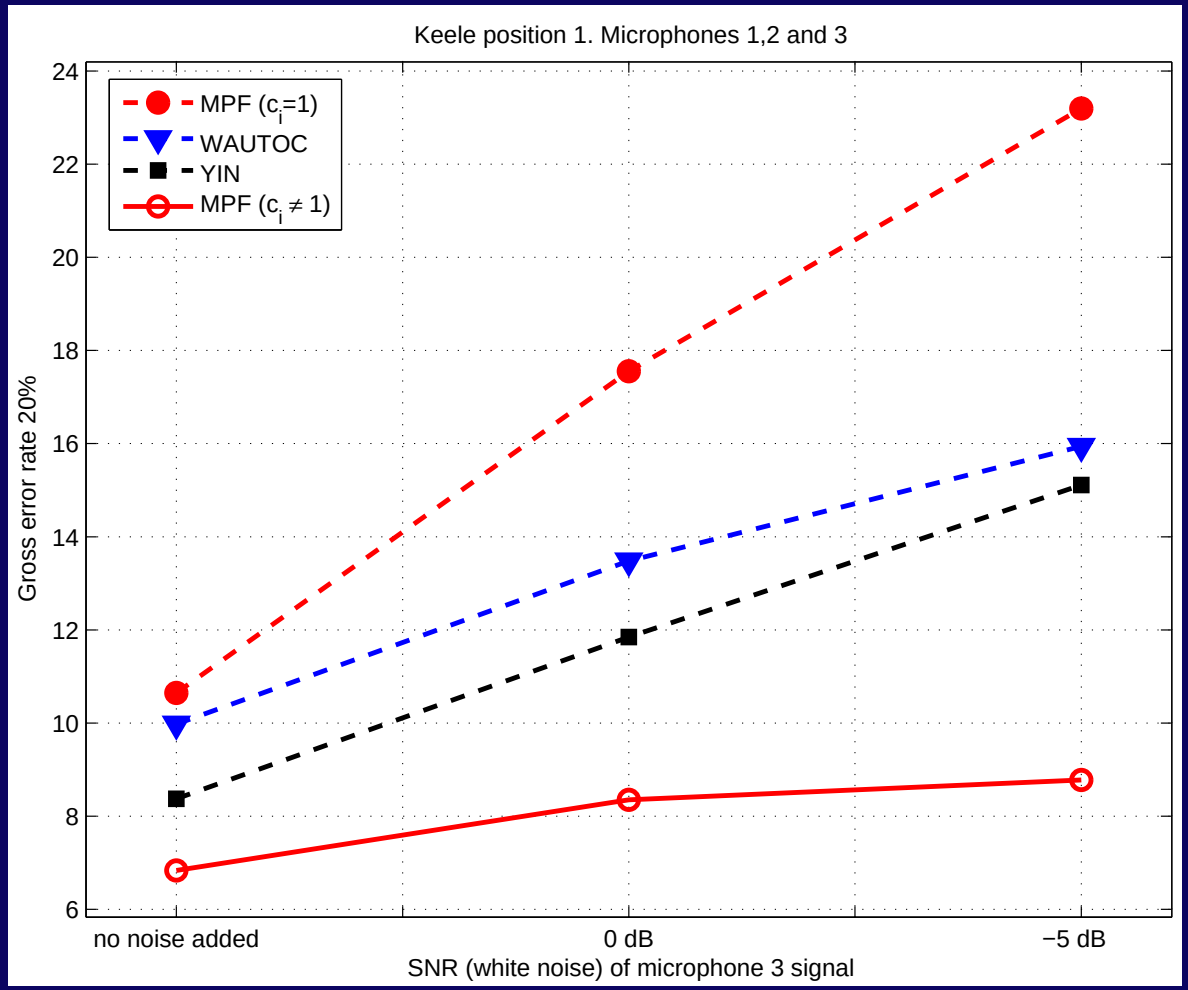
Wautoc, YIN, MPF (Gross Error Rate 20% - position P1)



Wautoc, YIN, MPF (Gross Error Rate 20% - position P2)



MPF channel reliability estimation (mics 1,2 and 3 - position P1)



- White noise added to channel 3 at different SNR
- MPF algorithm selects best channels (1 and 2) for F0 estimation

Conclusions

- Phase distortion introduced by reverberation is best handled by a frequency-domain approach (MPF)
- FFT based algorithms can easily perform real-time processing
- Channel reliability estimation permits to process concurrent activities independently

Future perspectives

- Test MPF performance on meeting and seminar data recorded in the IRST CHIL room
- Integration of F0 information to distant talking ASR (LPC pitch synchronous analysis)

Thank you very much