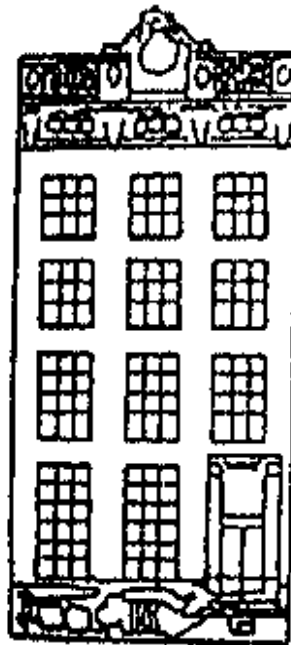# CAN STANDARD ANALYSIS TOOLS BE USED ON DECOMPRESSED SPEECH?

R.J.J.H. van Son
Institute of Phonetic Sciences/ACLC
University of Amsterdam
Herengracht 338, 1016CG Amsterdam
Rob.van.Son@hum.uva.nl

Introduction

Large Speech Corpora aim at
- Natural Interactions
- Field Recordings by Volunteers
- Large Amounts of it *(Months)*
- Internet Distribution

Solutions
- Minidisc Recorders
- Compressed Storage
- Compressed Distribution

Methods

SPEECH (*IFAcorpus*):

- 125 Segmented sentences, read and retold
- 4 male and 4 female speakers
- Recorded on 2 microphones to CD-audio

Analysis using *praat 4.0.16*:

- Pitch (*Simple*: Auto Correlation)
- Formants 1-3 (*Burg* algorithm)
- Spectral Center of Gravity (first spectral moment)

TEST CONDITIONS:

Microphone change: From HF condenser (Sennheiser MKH 105) to head-mounted dynamic (Shure SM10A)

Sony Minidisc: *ATRAC3* on Walkman MZ-R909

Ogg Vorbis (40 kbs): *1.0rc3*, 45 kbs effective (factor 15.5)

Ogg Vorbis (80 kbs): *1.0rc3*, 85 kbs effective (factor 8.3)

MP3 (192 kbs): *LAME 3.92*, 204 kbs effective (factor 3.5)
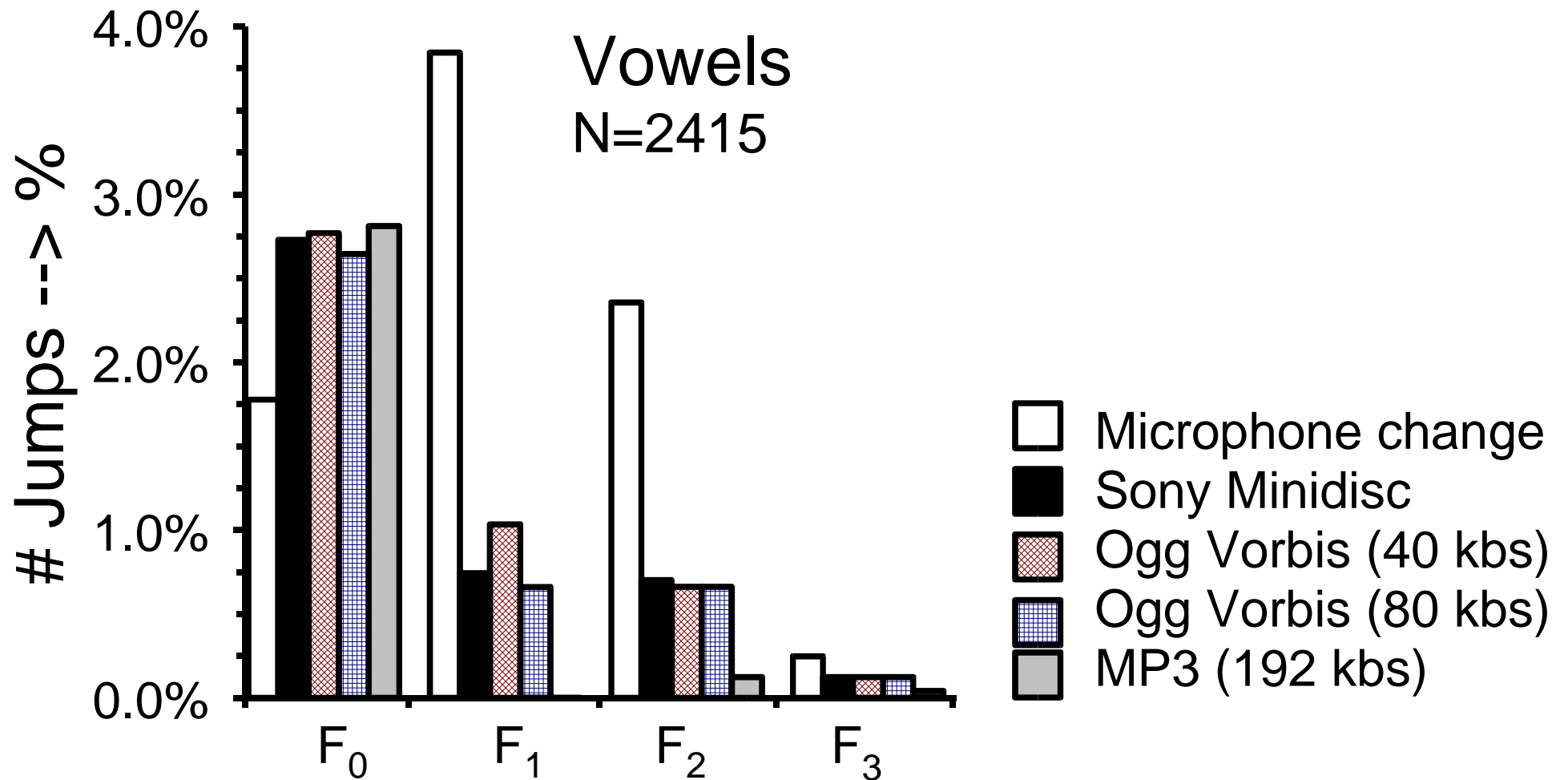
All compressed recordings aligned to within **0.5 ms** of original

Jump Errors

- Pitch can pick wrong (sub-)harmonic

- Formants can be mislabeled

- Results in large, "*jump*", errors that have to be handled

- Excluding differences larger than 9 semitones catches most of these jumps

Large Jumps in F_0-F_3
(# differences > 9 semitones)

Systematic Differences

Bit-rate 80 kbs and higher

- Pitch < 0.04 semitones
- Formants < 0.04 semitones
- CoG < 0.15 semitones

Bit-rate 40 kbs

- $F_2/F_3 \sim$ 0.1 semitones
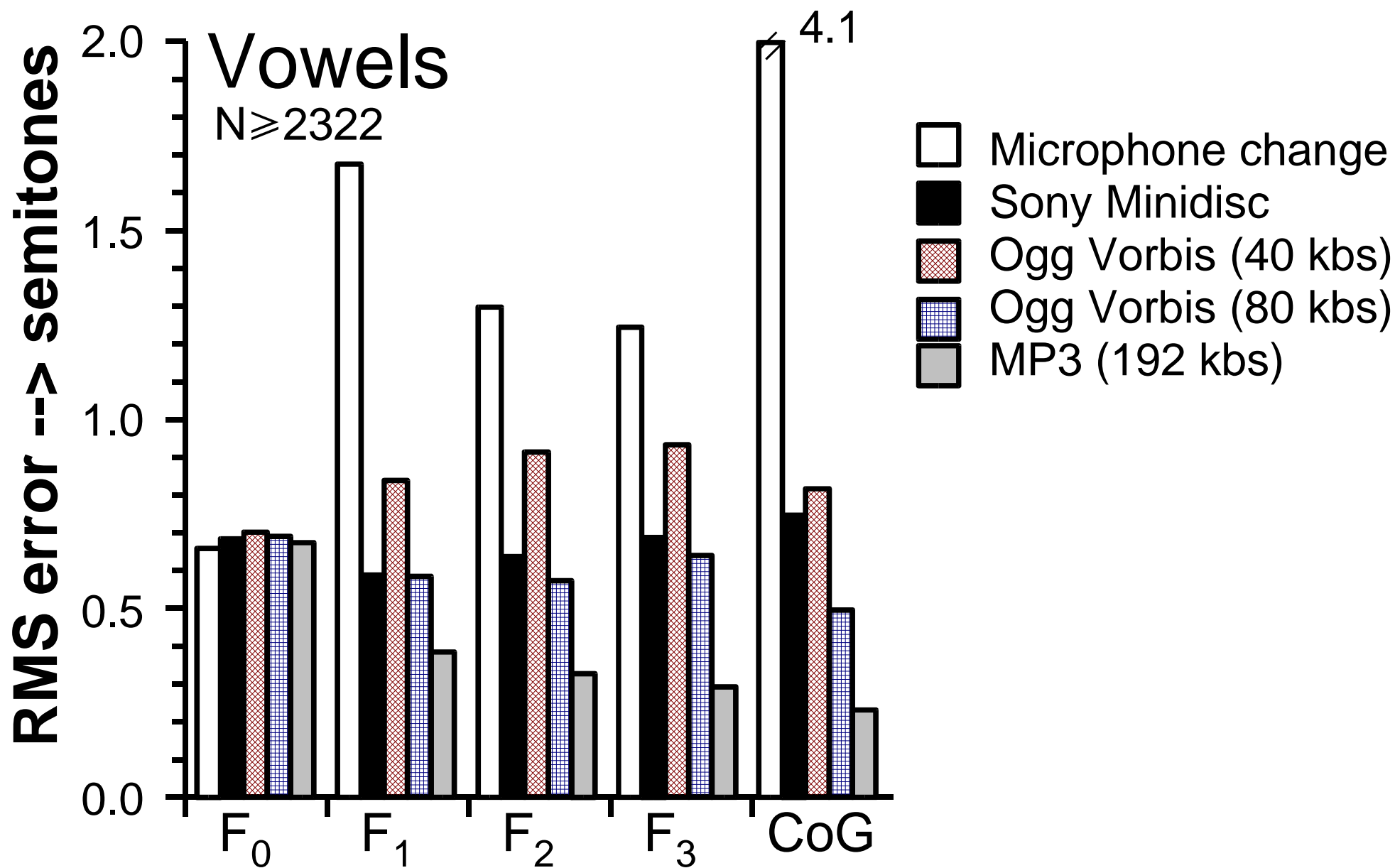- CoG < 0.5 semitones
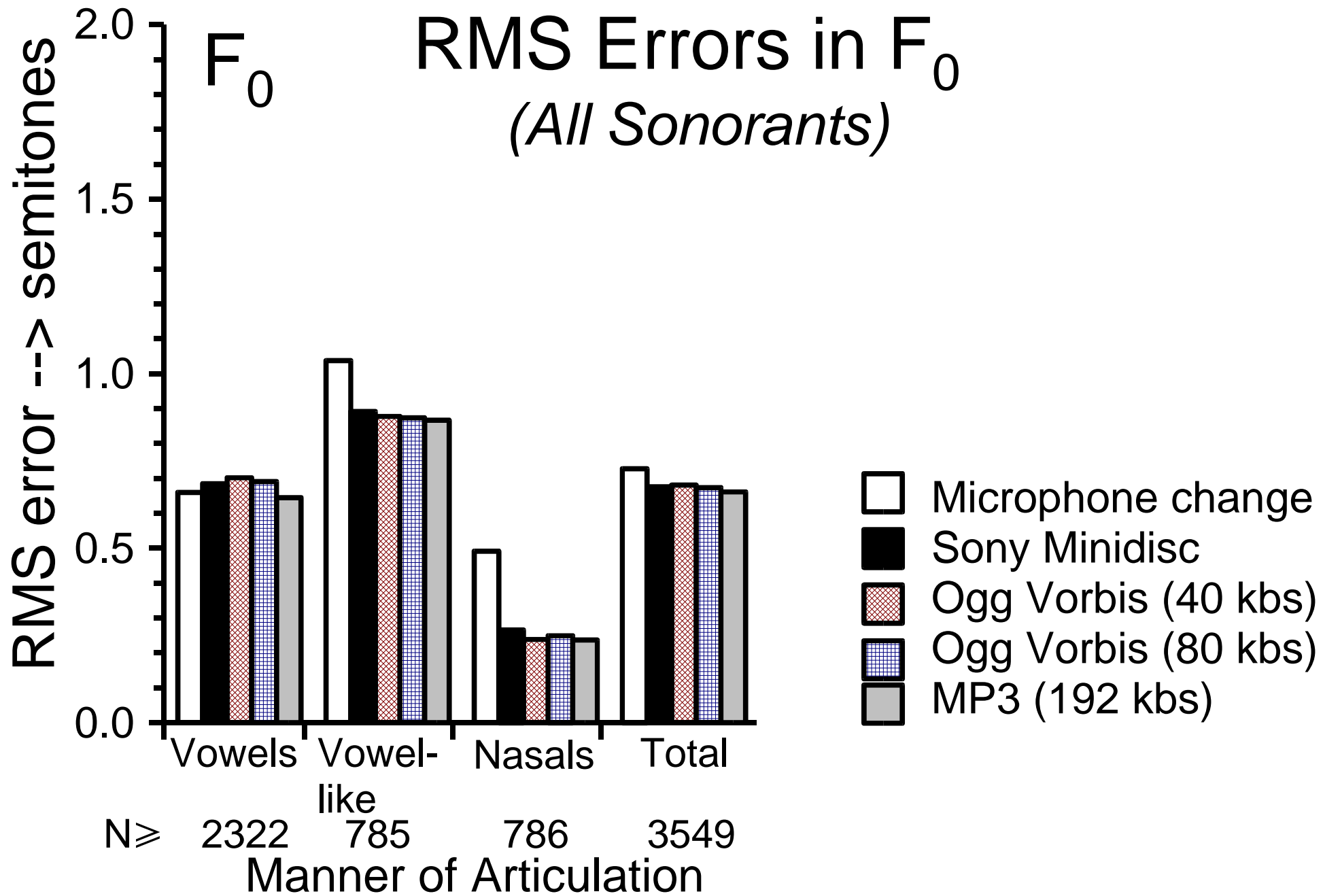
Microphone switch

- Formants < 0.5 semitones
- CoG < 5 semitones (!)

Root-Mean-Square Errors

- Systematic Differences are Ignored in this Study

- **Standard Deviation** == Root-Mean-Square Error

- Discard Pitch and Formant (*not* CoG)
  Differences > **9** semitones
  (>10 standard deviations of the difference)

RMS Errors in Pitch, Formant & CoG

RMS Errors in F_0
(All Sonorants)

F_0

RMS error --> semitones

- Microphone change
- Sony Minidisc
- Ogg Vorbis (40 kbs)
- Ogg Vorbis (80 kbs)
- MP3 (192 kbs)

Vowels   Vowel-like   Nasals   Total

N≥   2322   785   786   3549

Manner of Articulation

RMS Errors in CoG
(all continuants)

RMS error --> semitones

Manner of Articulation

Legend:
- Microphone change
- Sony Minidisc
- Ogg Vorbis (40 kbs)
- Ogg Vorbis (80 kbs)
- MP3 (192 kbs)
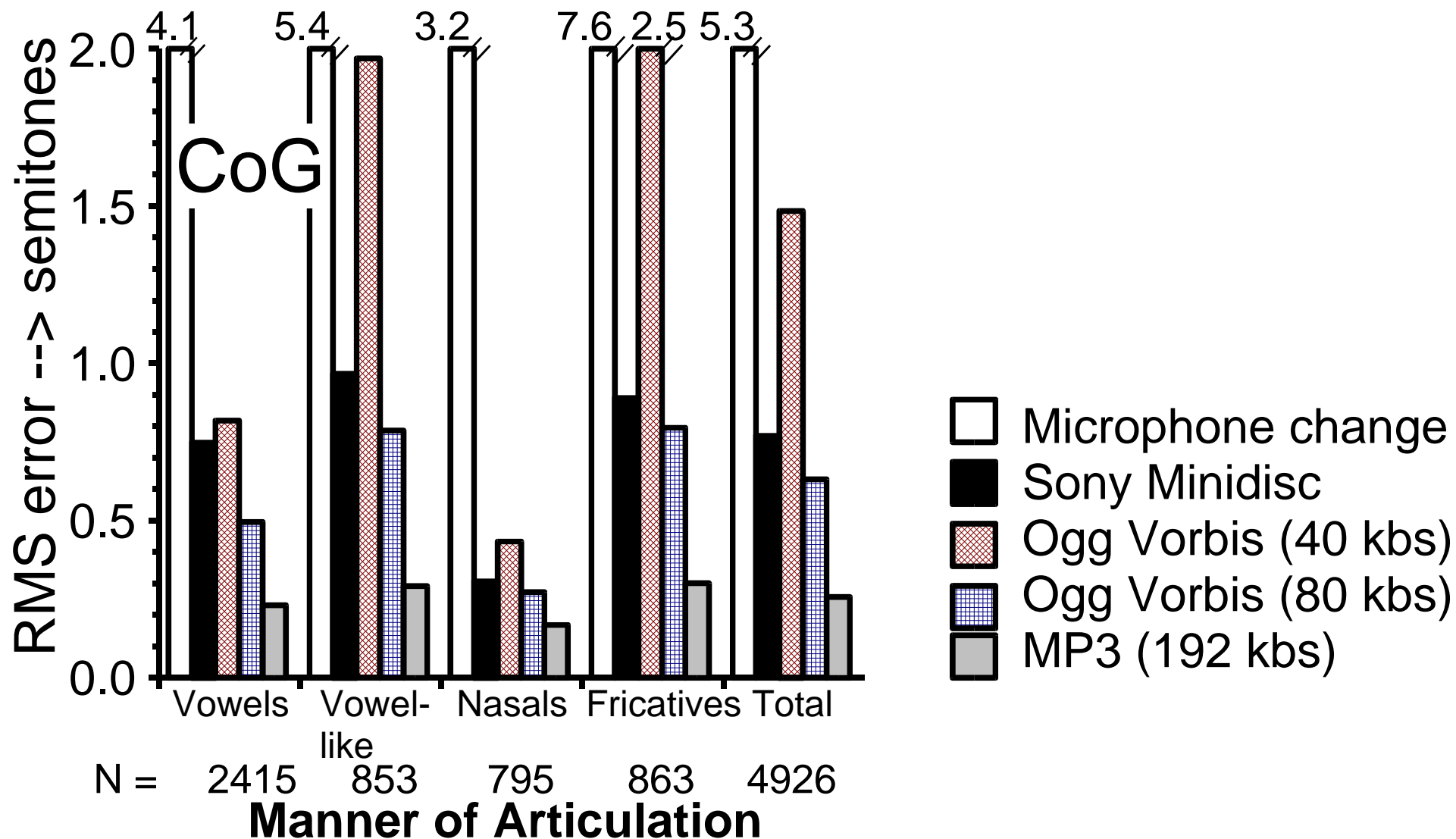
Cascaded Compression

Field situation:
- Record on Minidisc
- Transmit/Store/Distribute with 80 kbs Compression
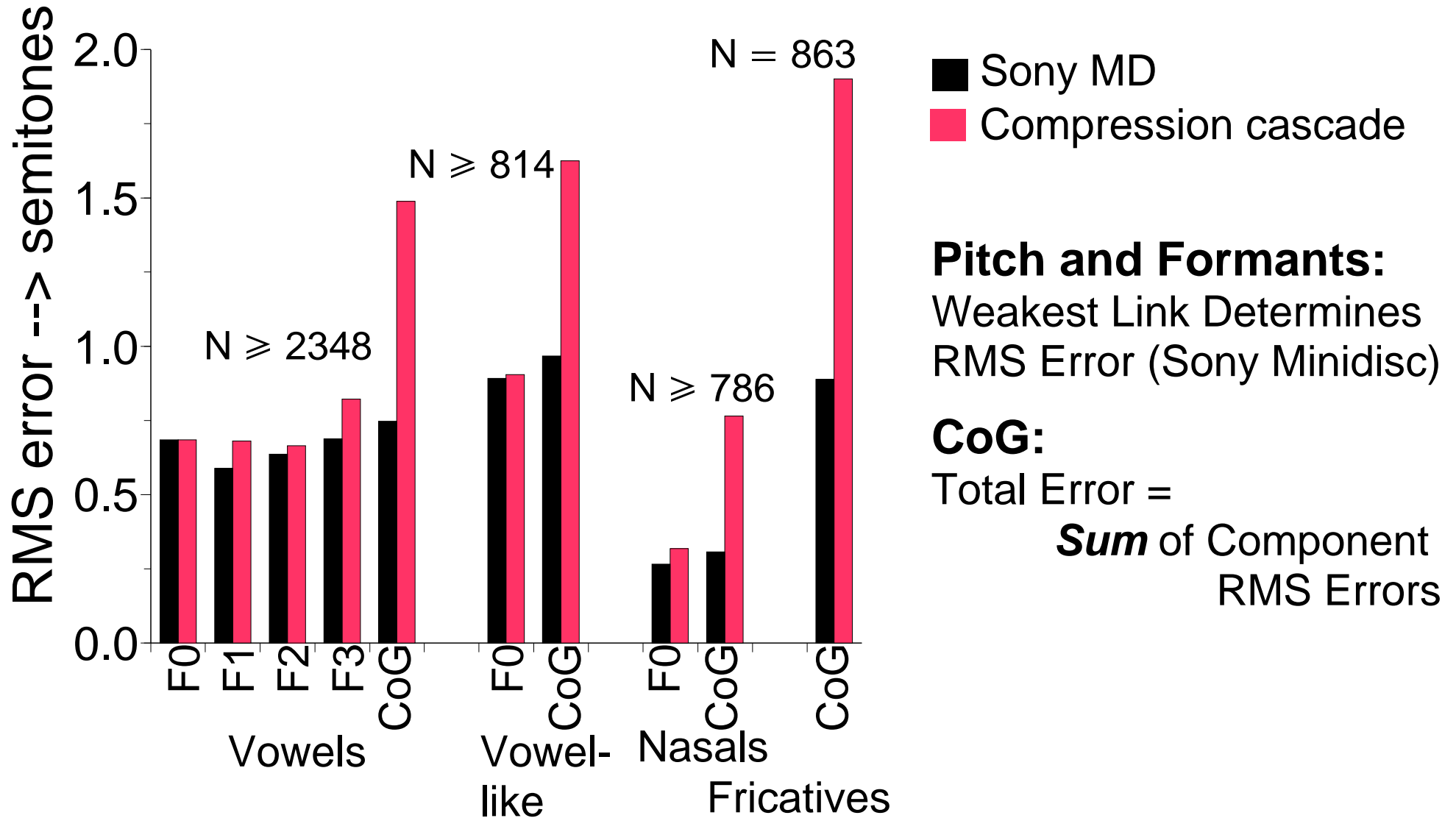- Archive with 192 kbs Compression

Simulated with:

CD-audio (Original)

-> Sony Minidisc

-> Ogg Vorbis 80 kbs

-> MP3 192 kbs

Cascaded Compression

Sony MD > Ogg Vorbis (80kbs) > MP3 (192kbs)

RMS error --> semitones

N ≥ 2348

N ≥ 814

N = 863

N ≥ 786

■ Sony MD
■ Compression cascade

Pitch and Formants:
Weakest Link Determines
RMS Error (Sony Minidisc)

CoG:
Total Error =
Sum of Component
RMS Errors

Vowels: F0, F1, F2, F3, CoG

Vowel-like: F0, CoG

Nasals: F0, CoG

Fricatives: CoG

Discussion and Conclusions

- Decompressed Speech can be used for *Pitch*, *Formant*, and Whole Spectrum (*CoG*) Analysis

- RMS error < 1 semitone (<6%)
  - Vowels < 0.7 semitone
  - Nasals < 0.3 semitone
  - Holds for Low bit-rates (40 kbs) for Pitch and Formants

- Repeated Compression *Combined* Error
  - Pitch & Formants: Weakest Link
  - CoG: *Sum* of Component RMS Errors
    Solution: (Partial) Translation of Formats, i.e., No Decompression

- CoG Strongly Affected by
  - Low bit-rates (40 kbs)
  - Repeated Compression
  - Microphone Choice