Acoustic Analysis of Babel Recording Conditions

Dan Ellis
Columbia & ICSI
Team SWORDFISH

dpwe@ee.columbia.edu  http://labrosa.ee.columbia.edu/

1. Environment Types
2. Acoustic Condition Histograms
3. Microphone Condition
4. Conclusions
1. Environment Types

- Each Babel recording has `envType` metadata
  - STREET, PUBLIC_PLACE, HOME_OFFICE_MOBILE, VEHICLE, CAR_KIT, HOME_OFFICE_LANDLINE

- Substantial effect of condition & lang, e.g.:
  - BP_105
    CK vs. HOM
  - BP_108
    CK vs. HOL

![Error breakdown by condition](BP_101)

![Error breakdown by condition](BP_104)

![Error breakdown by condition](BP_105)

![Error breakdown by condition](BP_106)

![Error breakdown by condition](BP_107)

![Error breakdown by condition](Sum)

BP Kaldi on dev
Variations

- Per-condition averages hide spreads
  - still substantial differences between conditions

- But...
  - within-condition variation exceeds between-condition
Characterizing Acoustics

• Mel Spectrogram

• Histogram of dB energy in each Mel band

• Energy covariance
Example: BP_101 Car_Kit

- Gating by subbands
Top-Level View - BP Languages

- All BP_1xx dev set utterances pooled

BP_101 – ALL

BP_104 – ALL

BP_105 – ALL

BP_106 – ALL

BP_107 – ALL

Corpora – ALL
Drilling Down: BP_105 (Turkish)

- DEV utterances sorted by envType

(bp_105 – CAR_KIT)

(bp_105 – VEHICLE)

(bp_105 – PUBLIC_PLACE)

(bp_105 – STREET)

(bp_105 – HOME_OFFICE_MOBILE)

(bp_105 – HOME_OFFICE_LANDLINE)
BP_105 Car_Kit vs. HO_Mob

BABEL_BP_105_31256_20120531_015506_inLine – CK
WER=79.6

BABEL_BP_105_42212_20120706_194059_inLine – CK
WER=66.3

BABEL_BP_105_39774_20120623_021946_inLine – CK
WER=54.9

BABEL_BP_105_31345_20120515_214849_inLine – CK
WER=49.1

BABEL_BP_105_20213_20120123_011920_outLine – HOM
WER=92.3

BABEL_BP_105_48536_20120208_212737_outLine – HOM
WER=56.9

BABEL_BP_105_66883_20120207_051718_inLine – HOM
WER=46.1

BABEL_BP_105_87806_20120201_235442_outLine – HOM
WER=28.8
Top-Level View - OPI1 Languages

OP1_102 – ALL

OP1_201 – ALL

OP1_206 – ALL

Corpora – ALL
Drilling Down: OPI_201 (Creole)

- **op1_201 – CAR_KIT**: Frequency range from 579 to 3882 Hz. The level dB is shown across the frequency range.

- **op1_201 – PUBLIC_PLACE**: Frequency range from 579 to 3882 Hz. The level dB is shown across the frequency range.

- **op1_201 – VEHICLE**: Frequency range from 579 to 3882 Hz. The level dB is shown across the frequency range.

- **op1_201 – STREET**: Frequency range from 579 to 3882 Hz. The level dB is shown across the frequency range.

- **op1_201 – HOME_OFFICE_MOBILE**: Frequency range from 579 to 3882 Hz. The level dB is shown across the frequency range.

- **op1_201 – MICROPHONE**: Frequency range from 579 to 3882 Hz. The level dB is shown across the frequency range.

Babel Audio Analysis - Ellis
2xx Wideband Mic Channel

- 48 kHz sampling rate
- Single noise floor
- Extremely broad-band
Blind SNR Estimation

- **NIST STNR** - histogram-based
  - confused by double noise floor

- **SNRvad** - VAD-based
  - tricked by noise gating

![NIST STNR vs SNRvad for BABEL OP1 206 train](image)
Summary

• Acoustic environment affects performance
  ▪ impact varies with different languages

• Cellphone noise gating gives complex histograms
  ▪ worse in CAR_KIT
  ▪ confuses simple SNR estimation

• Mic channel is extremely high quality
  ▪ limited data - hard to exploit

• Opportunities
  ▪ improved normalization?
  ▪ clustering of channel characteristics