

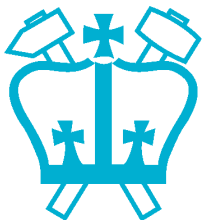
Acoustic Analysis of Babel Recording Conditions

Dan Ellis
Columbia & ICSI
Team SWORDFISH

dpwe@ee.columbia.edu

<http://labrosa.ee.columbia.edu/>

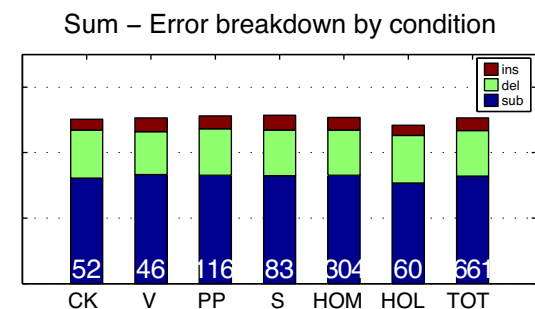
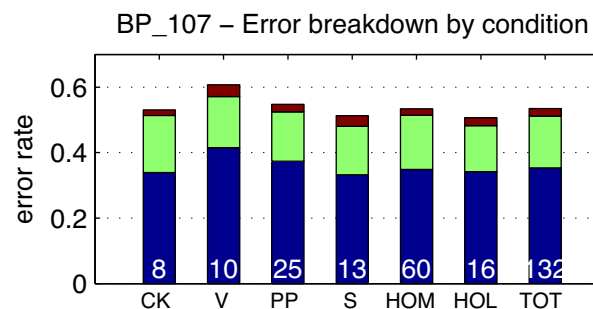
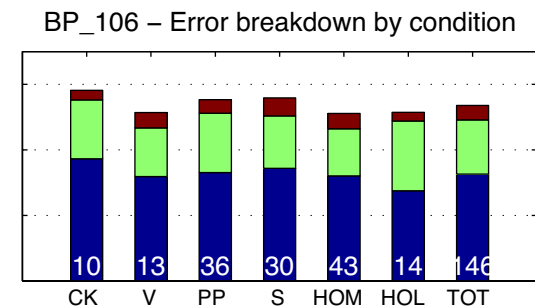
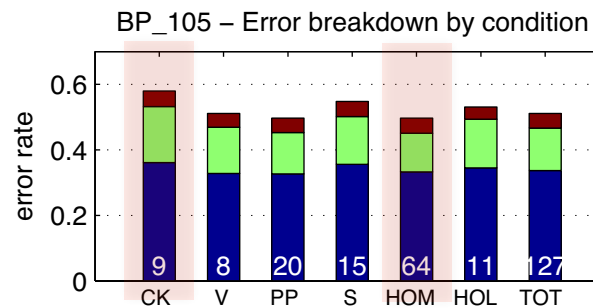
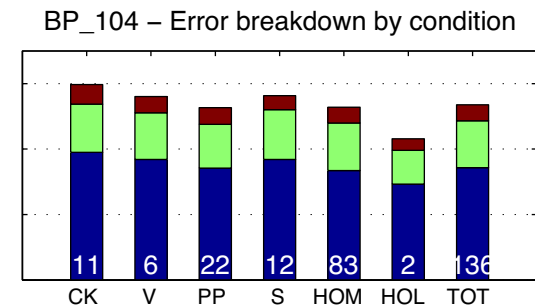
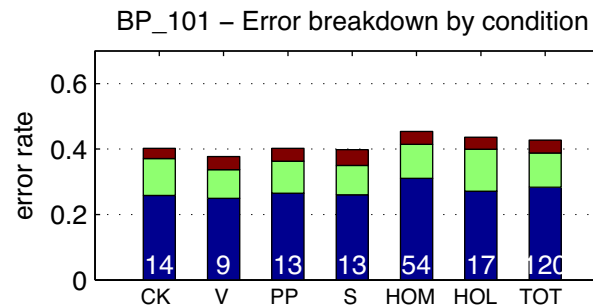
1. Environment Types
2. Acoustic Condition Histograms
3. Microphone Condition
4. Conclusions



COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

I. 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

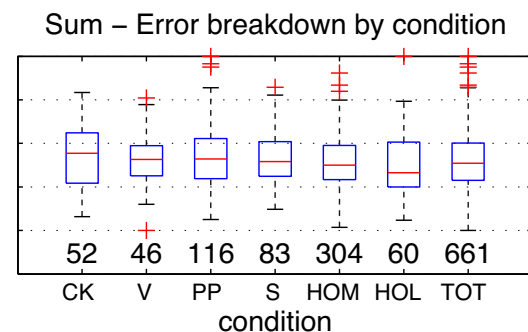
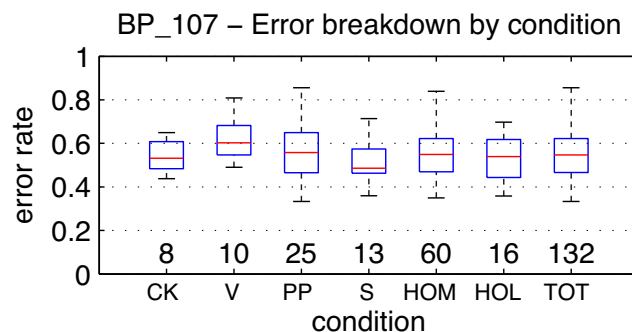
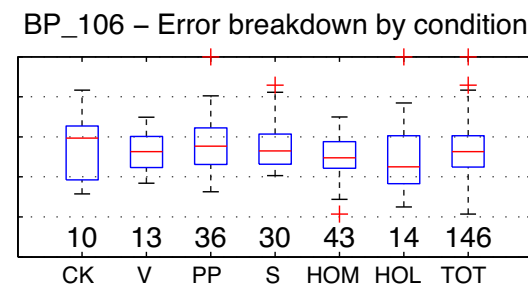
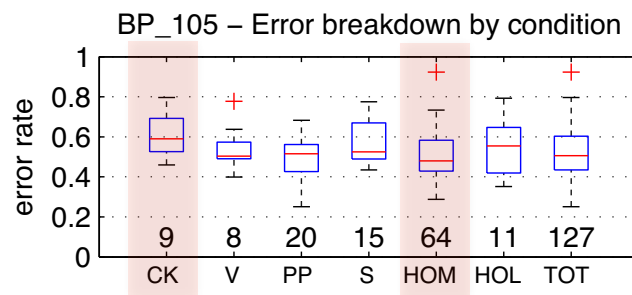
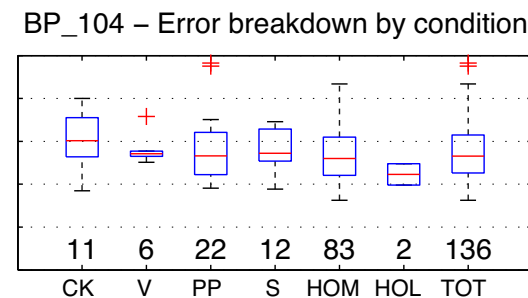
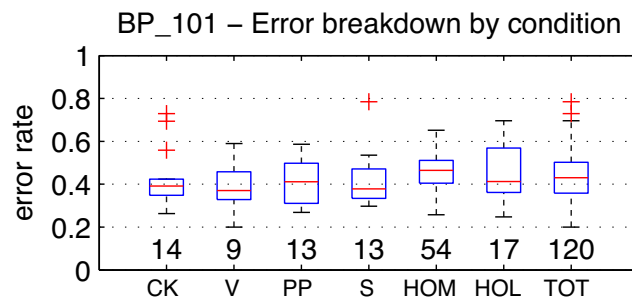


BP Kaldi
on dev

Variations

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

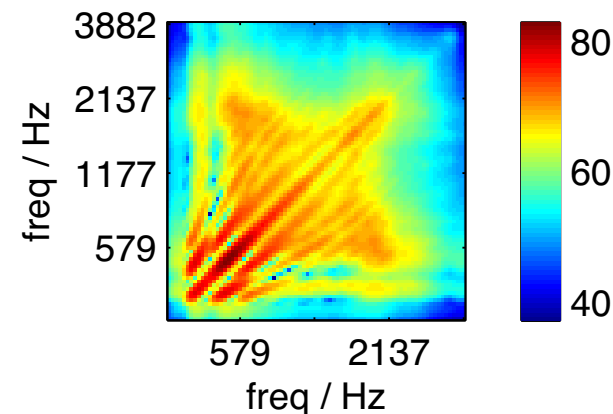
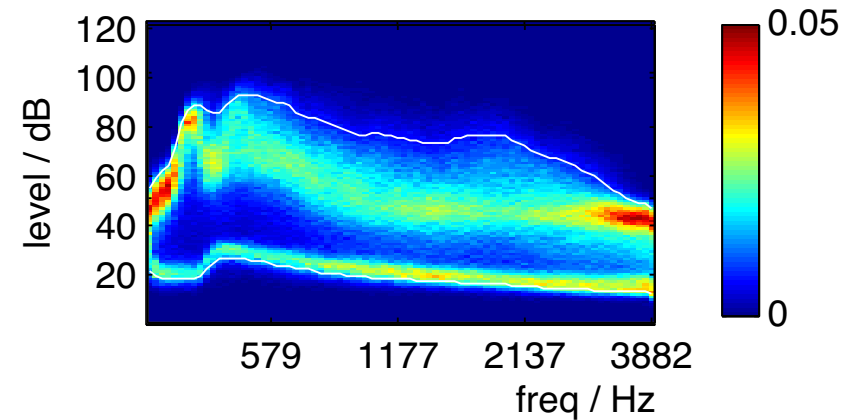
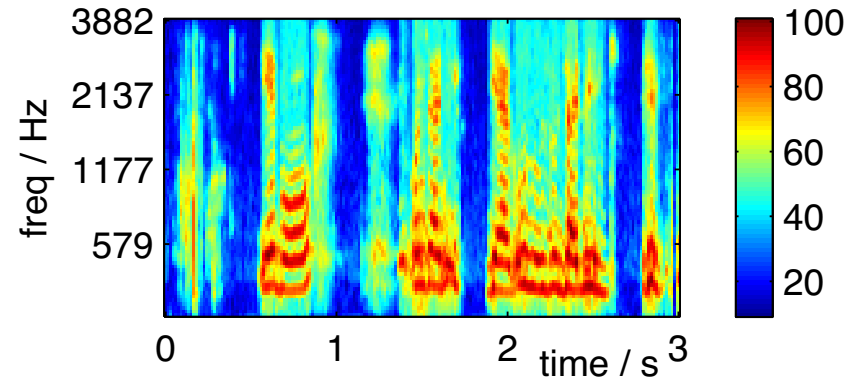
- But...
 - within-condition variation exceeds between-condition



Characterizing Acoustics

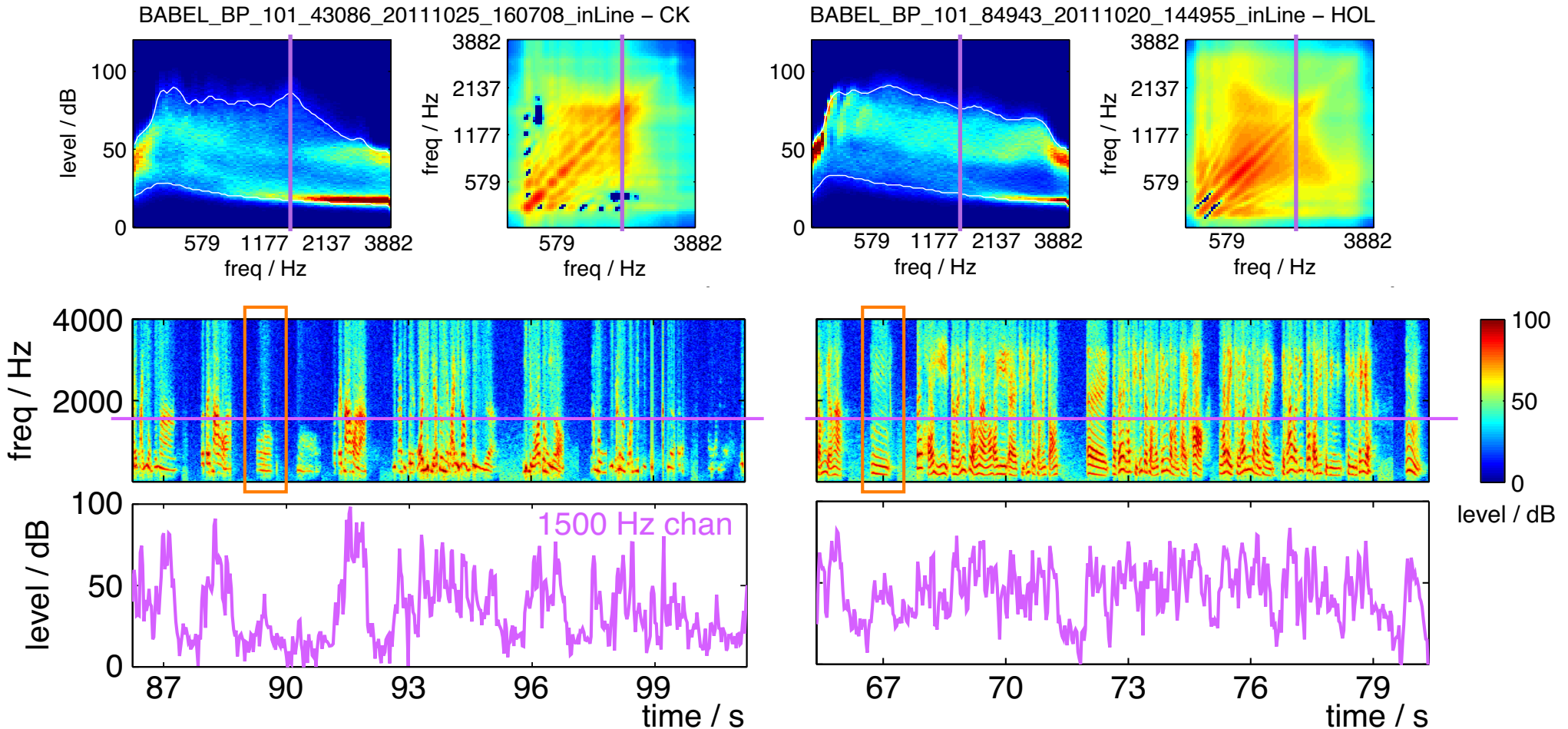
- Mel Spectrogram
- Histogram of dB energy in each Mel band
- Energy covariance

BABEL_OP1_206_65882_20121201_174526_outLine_PP



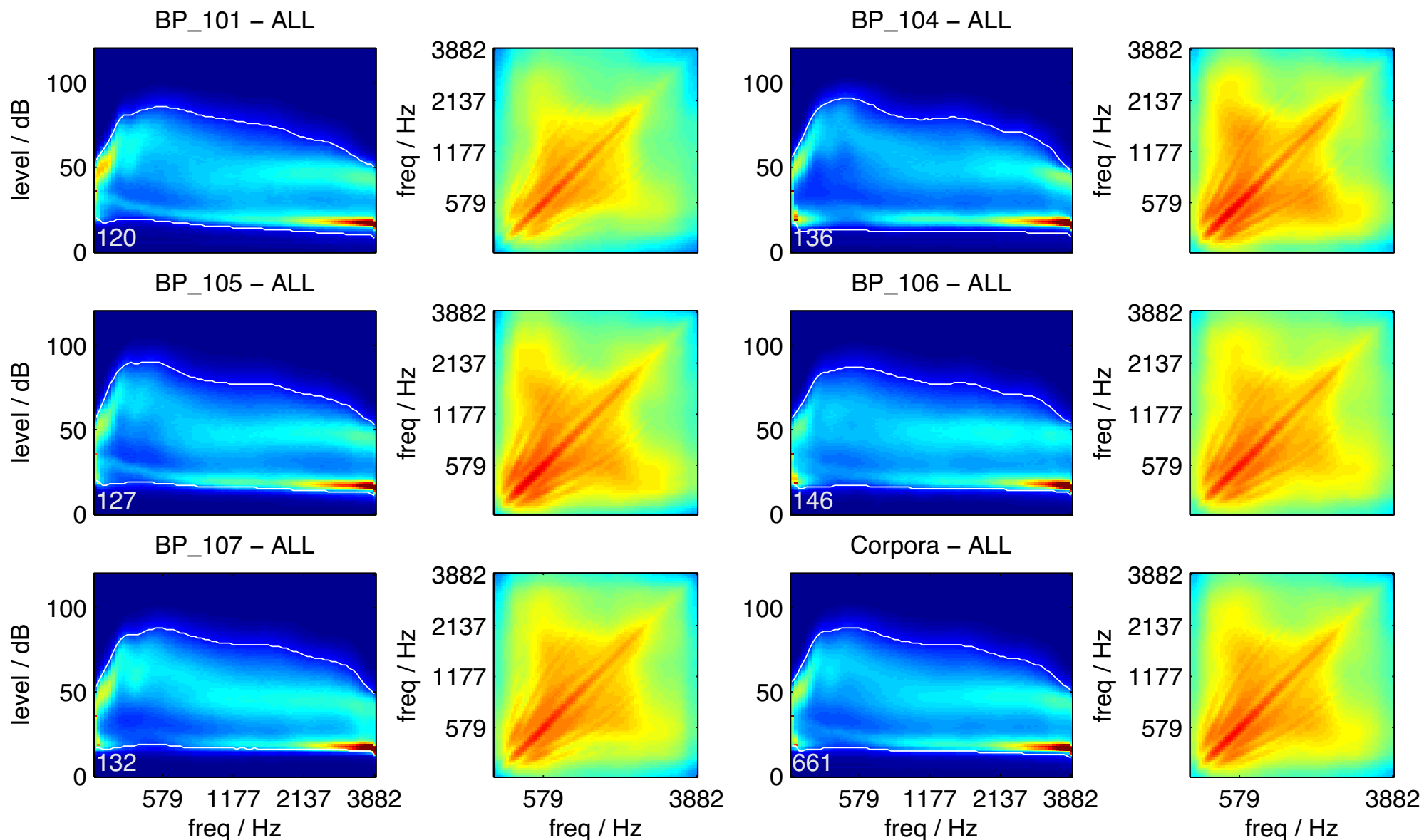
Example: BP_101 Car_Kit

- Gating by subbands



Top-Level View - BP Languages

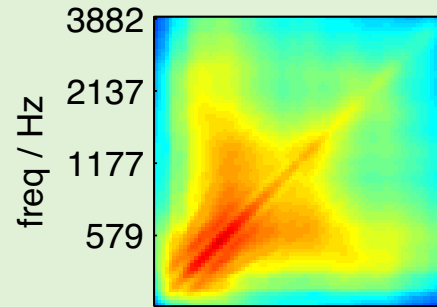
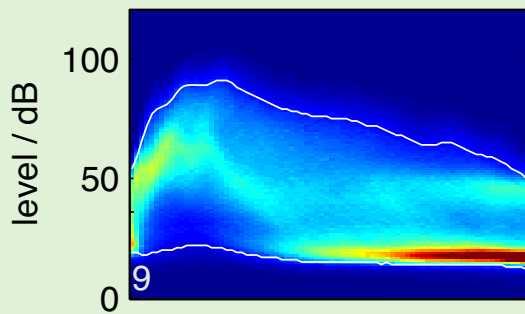
- All BP_1xx dev set utterances pooled



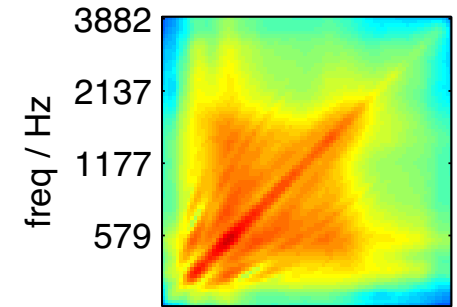
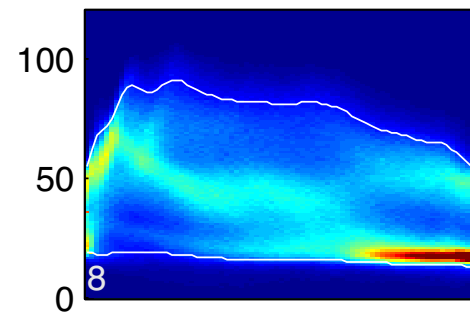
Drilling Down: BP_105 (Turkish)

- DEV utterances sorted by envType

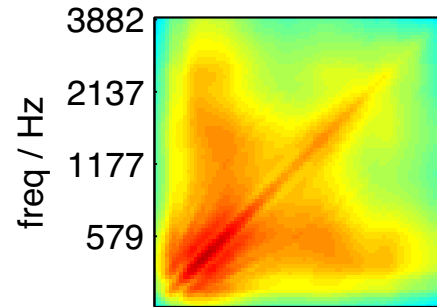
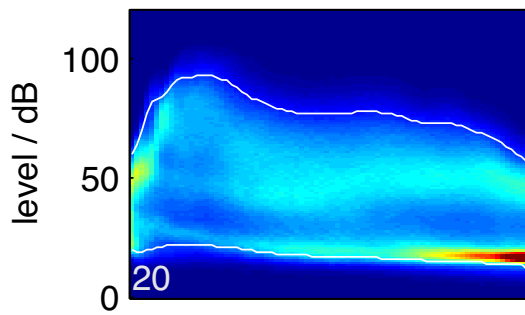
bp_105 - CAR_KIT



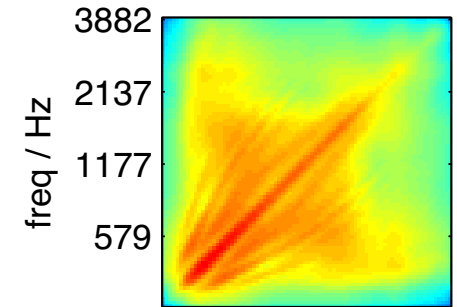
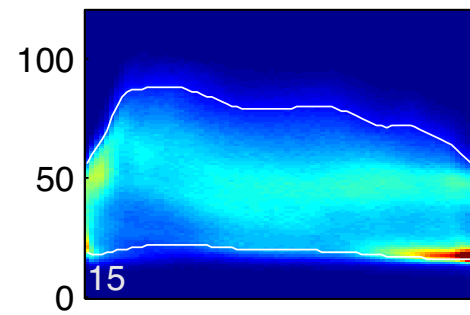
bp_105 - VEHICLE



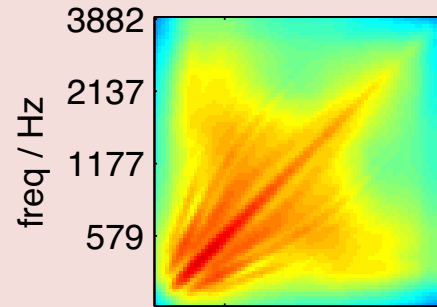
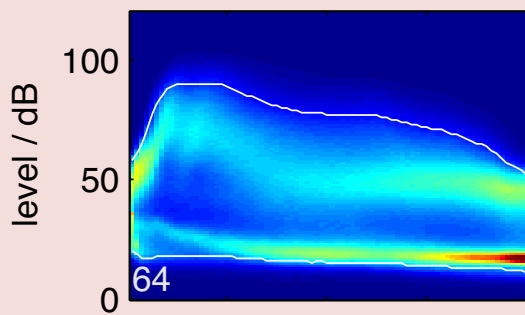
bp_105 - PUBLIC_PLACE



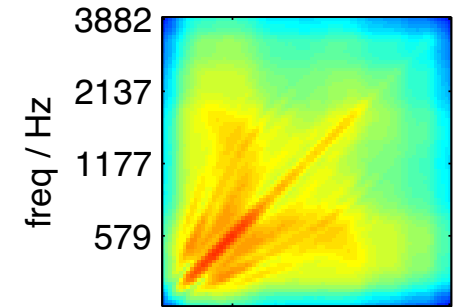
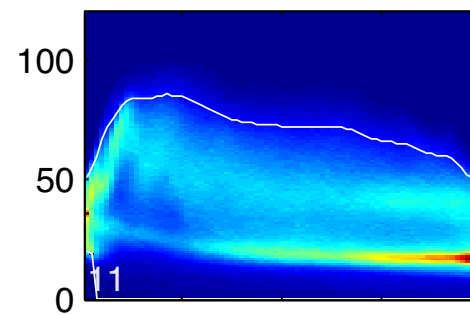
bp_105 - STREET



bp_105 - HOME_OFFICE_MOBILE



bp_105 - HOME_OFFICE_LANDLINE



579 1177 2137 3882
freq / Hz

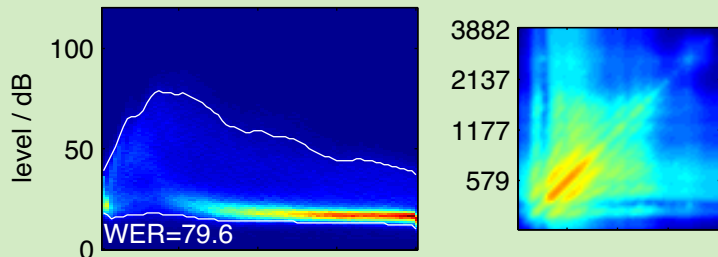
579 1177 2137 3882
freq / Hz

579 1177 2137 3882
freq / Hz

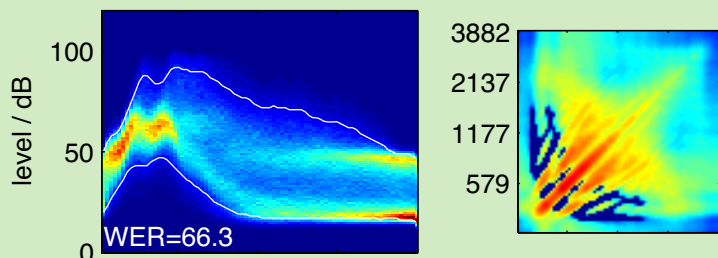
579 1177 2137 3882
freq / Hz

BP_105 Car_Kit vs. HO_Mob

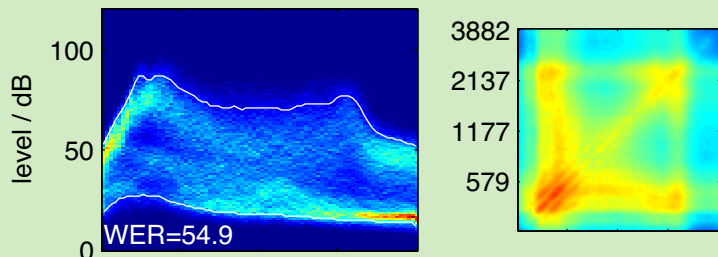
BABEL_BP_105_31256_20120531_015506_inLine - CK



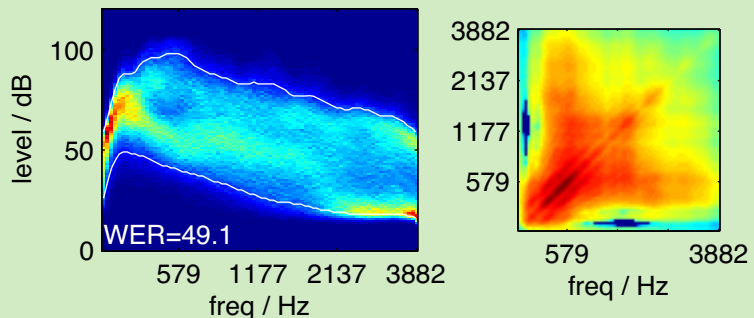
BABEL_BP_105_42212_20120706_194059_inLine - CK



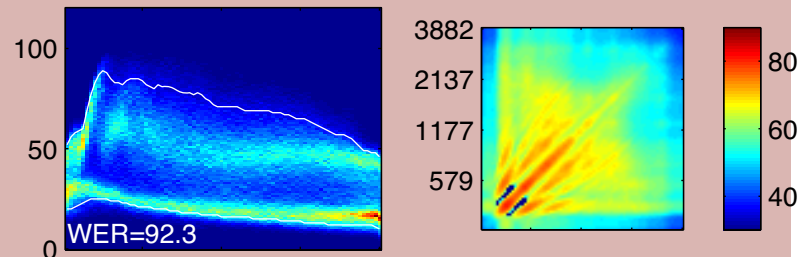
BABEL_BP_105_39774_20120623_021946_inLine - CK



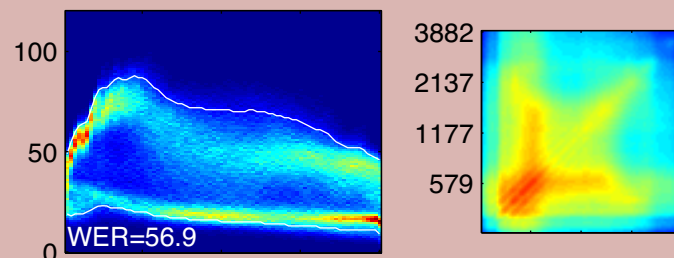
BABEL_BP_105_31345_20120515_214849_inLine - CK



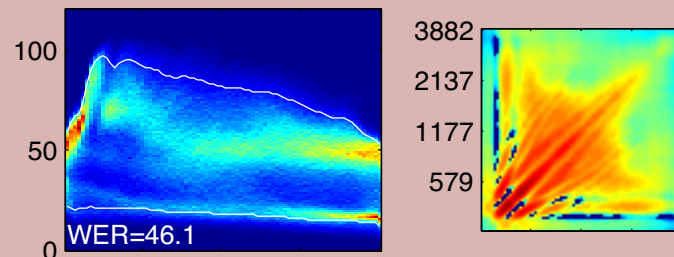
BABEL_BP_105_20213_20120123_011920_outLine - HOM



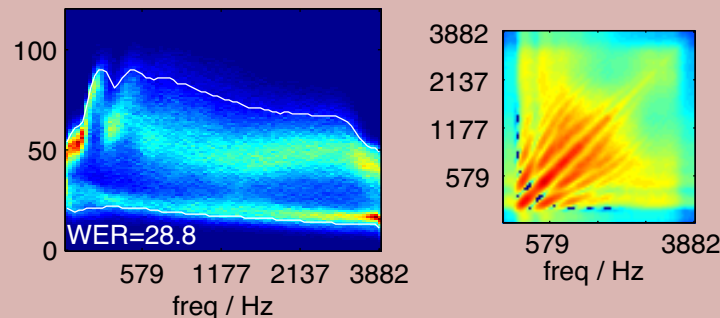
BABEL_BP_105_48536_20120208_212737_outLine - HOM



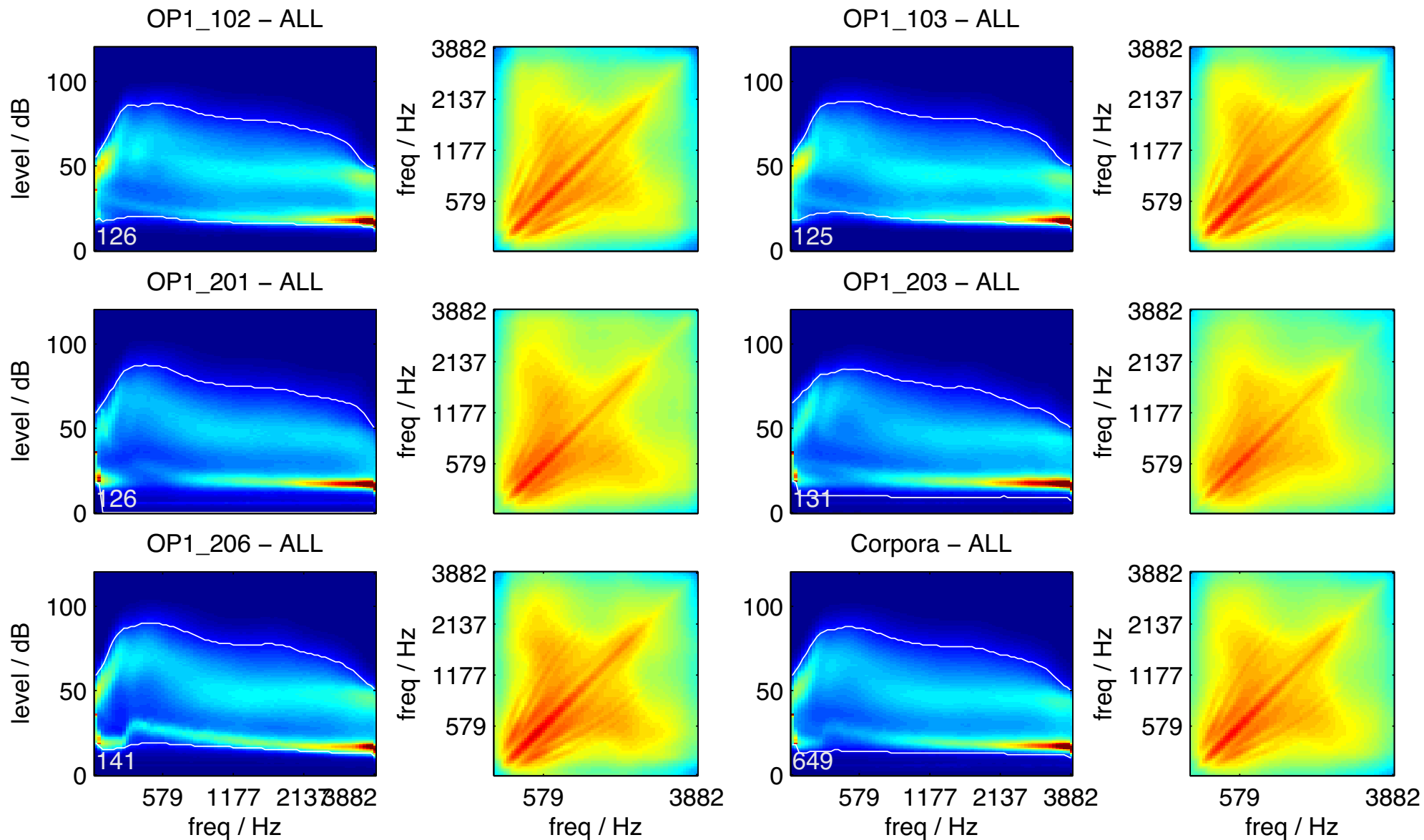
BABEL_BP_105_66883_20120207_051718_inLine - HOM



BABEL_BP_105_87806_20120201_235442_outLine - HOM

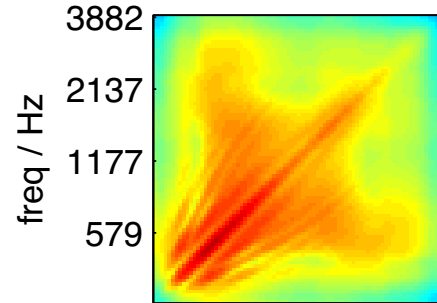
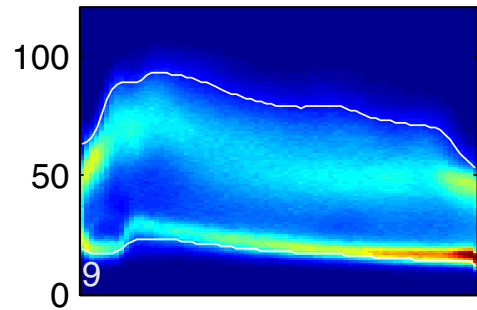


Top-Level View - OPI Languages

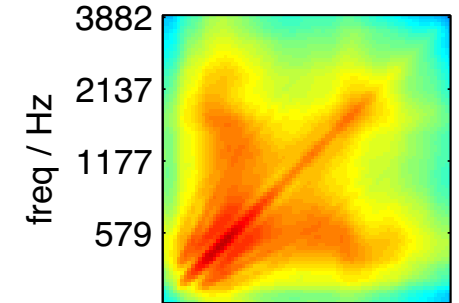
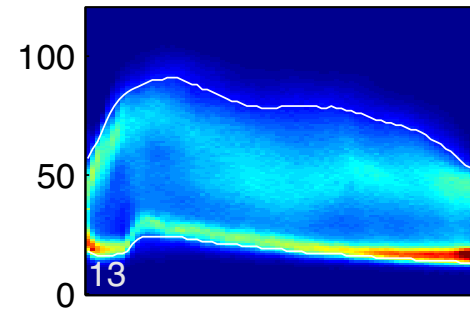


Drilling Down: OPI_206 (Zulu)

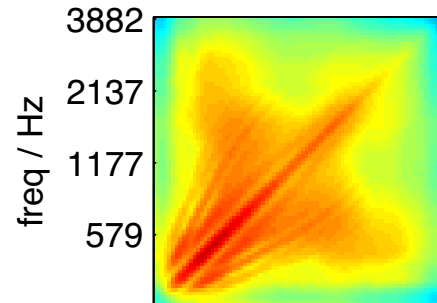
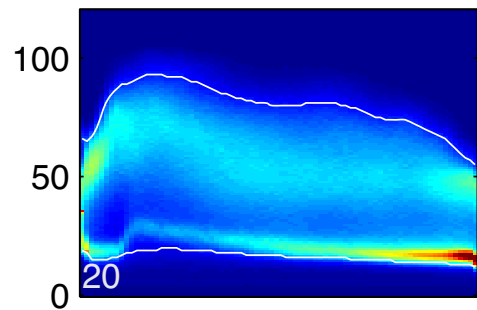
op1_206 - CAR_KIT



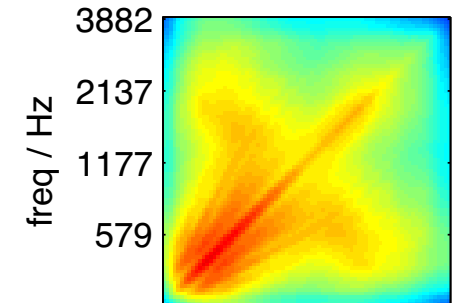
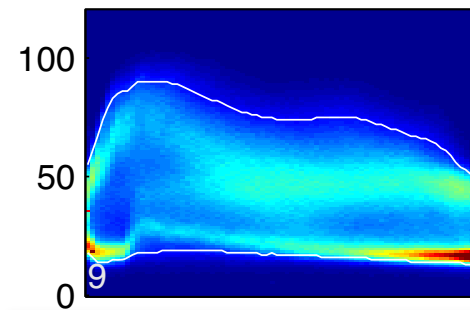
op1_206 - VEHICLE



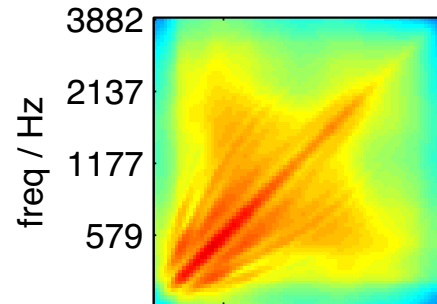
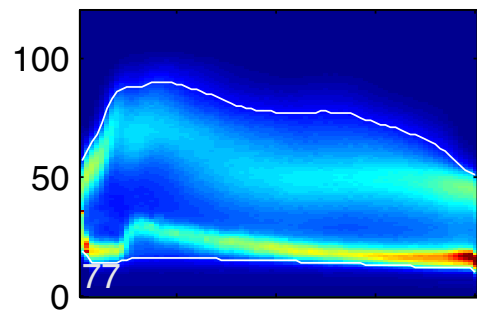
op1_206 - PUBLIC_PLACE



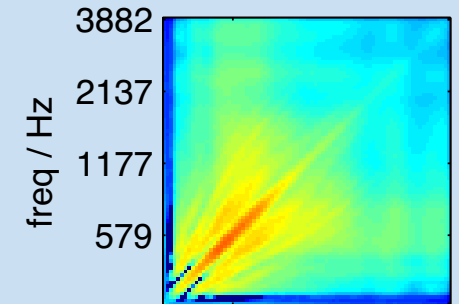
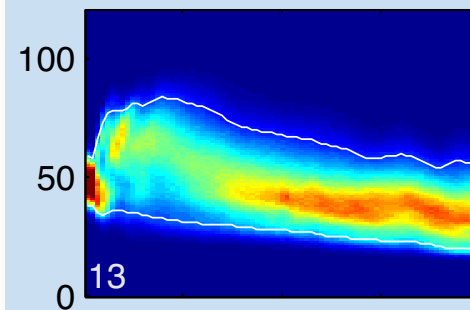
op1_206 - STREET



op1_206 - HOME_OFFICE_MOBILE

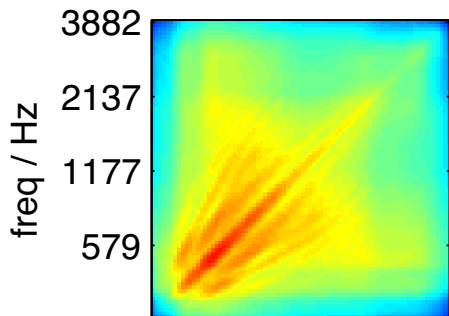
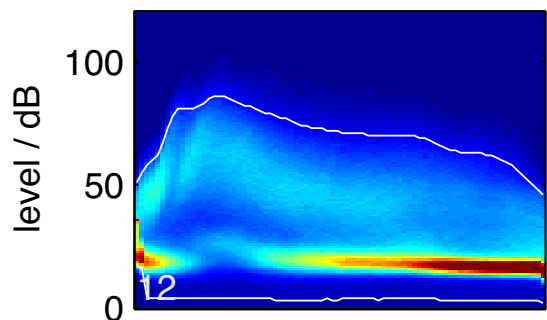


op1_206 - MICROPHONE

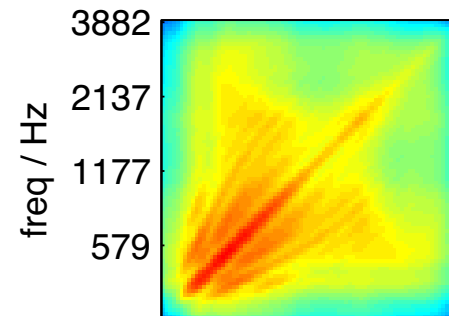
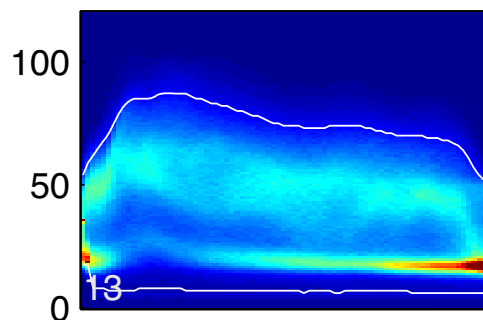


Drilling Down: OPI_201 (Creole)

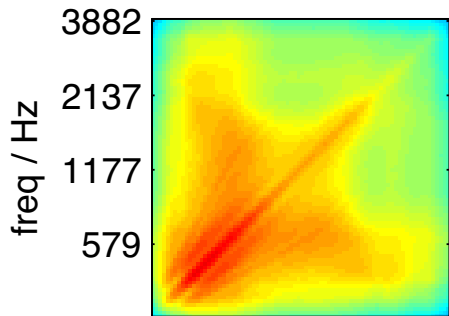
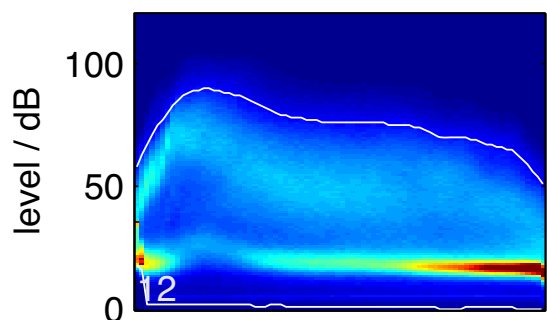
op1_201 - CAR_KIT



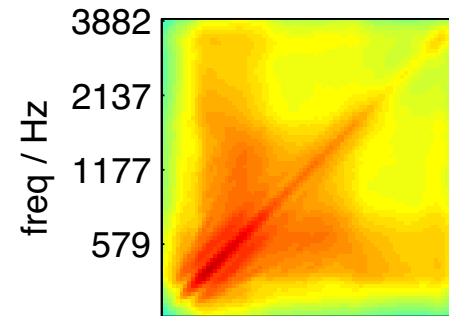
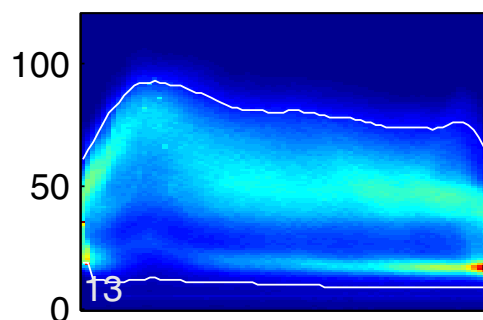
op1_201 - VEHICLE



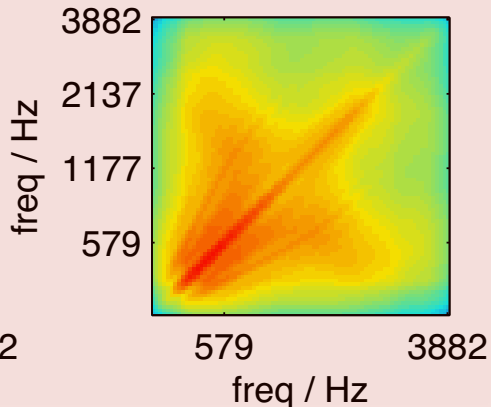
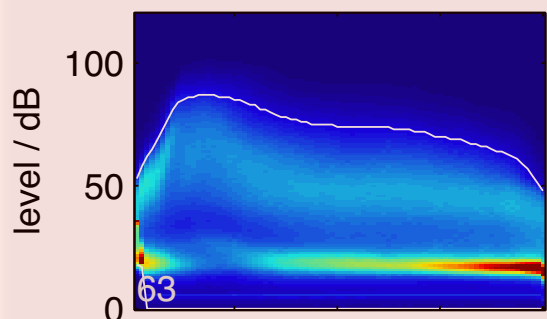
op1_201 - PUBLIC_PLACE



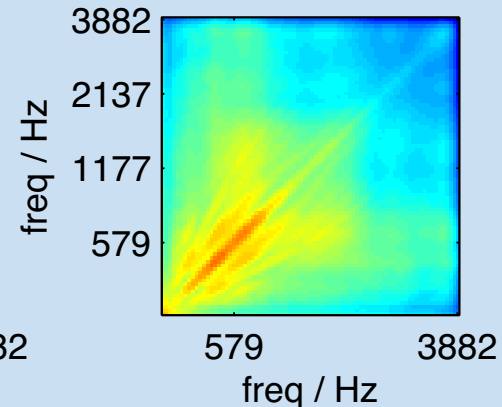
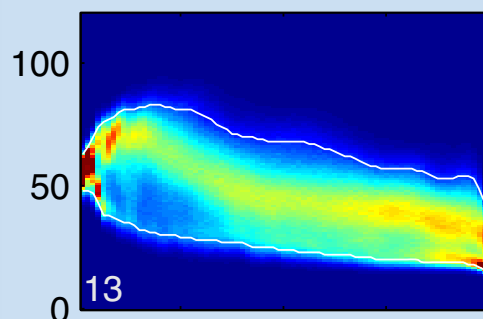
op1_201 - STREET



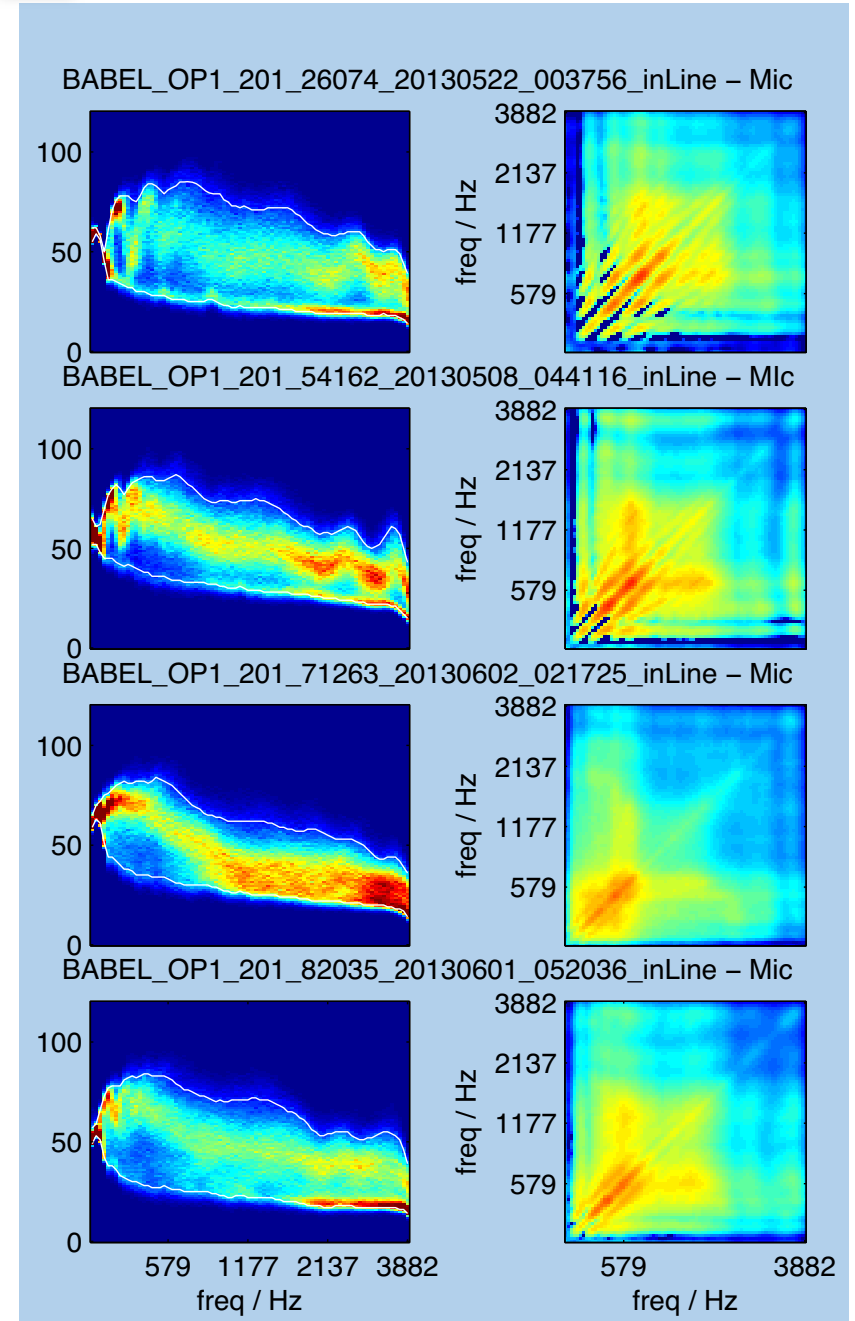
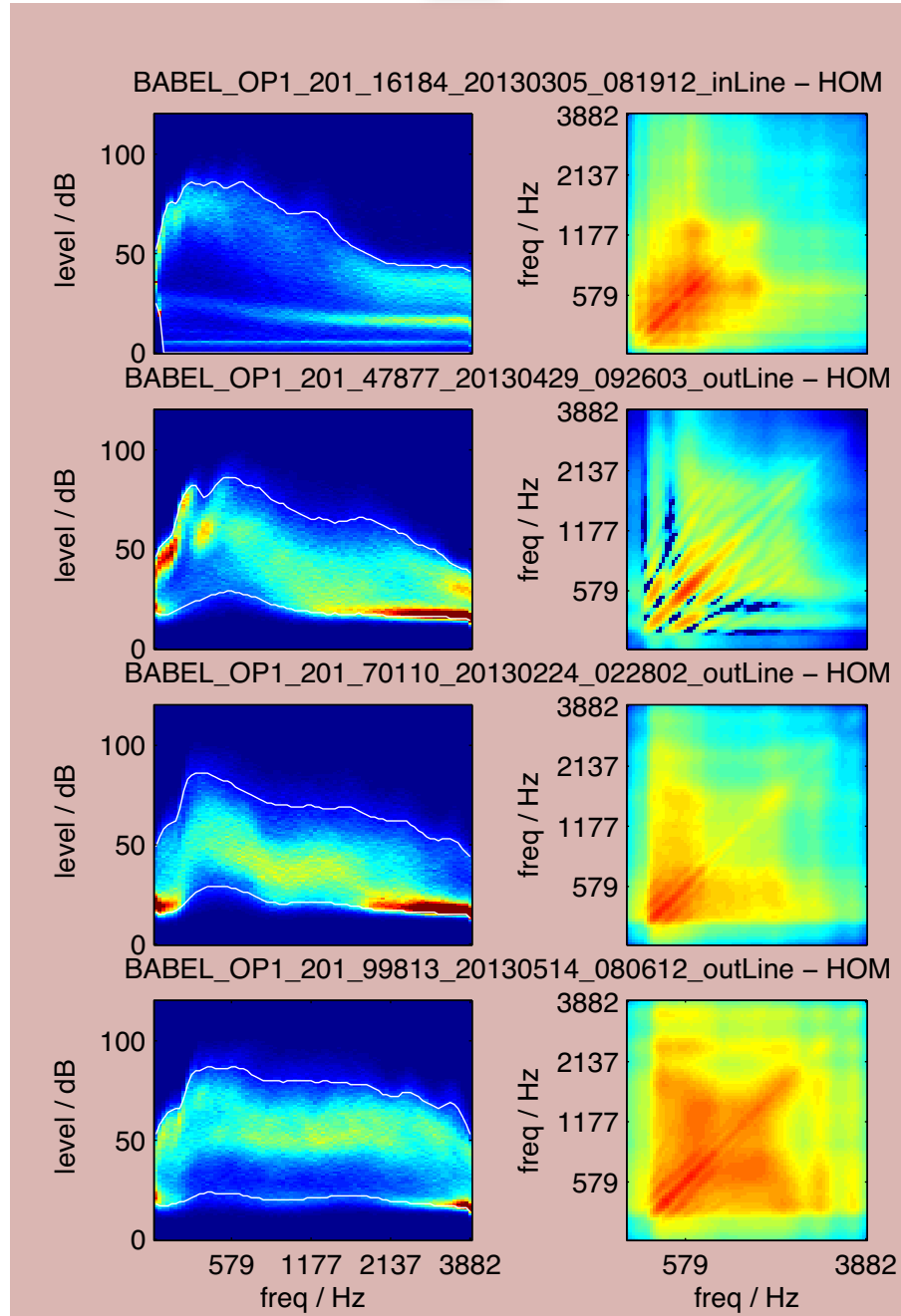
op1_201 - HOME_OFFICE_MOBILE



op1_201 - MICROPHONE

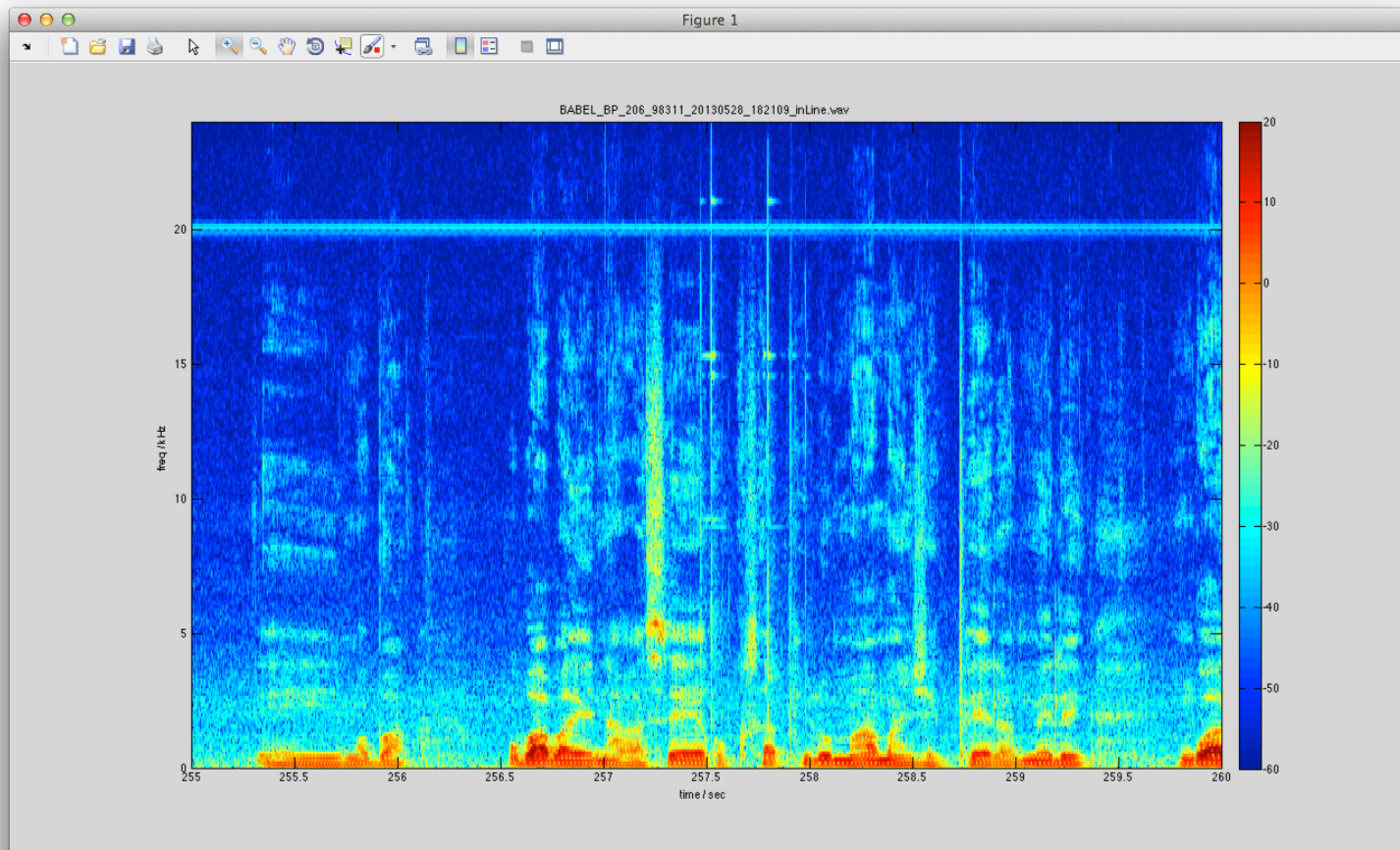
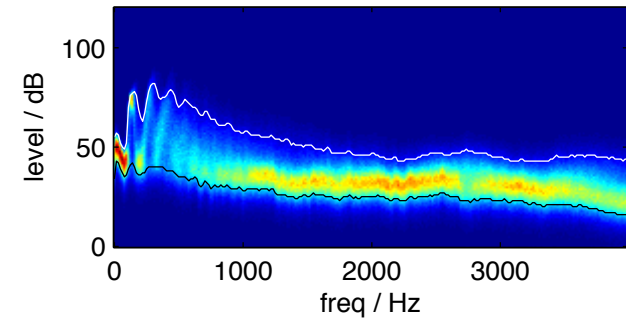


OPI_201 HO_Mob vs. Mic



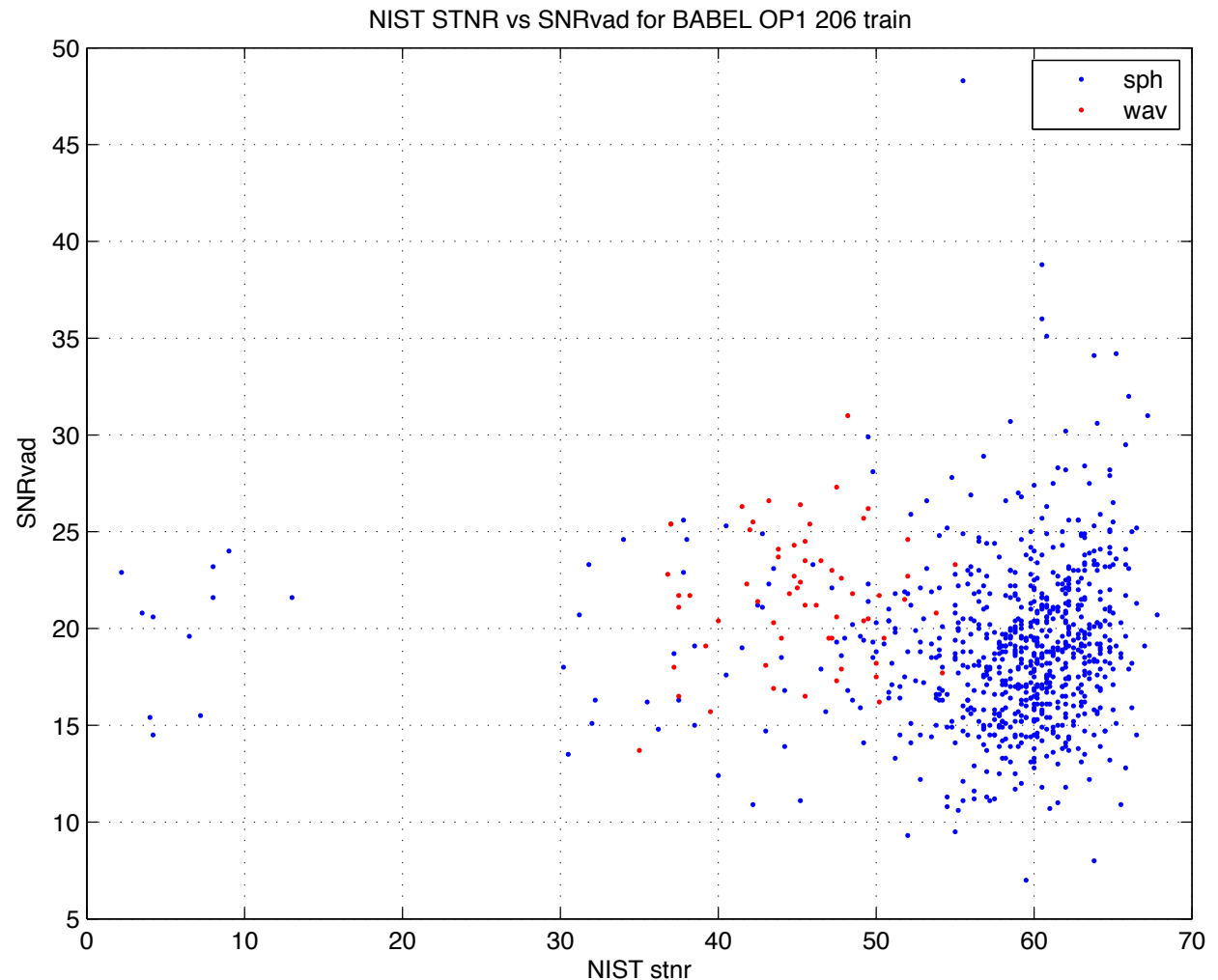
2xx Wideband Mic Channel

- 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



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