

## ACOUSTICS2008/228

# Channel mitigation approach for automatic language identification

Allen Stauffer and Aaron Lawson  
RADC, Inc., 525 Brooks Road, Rome, NY 13441, USA

A major obstacle to overcome in language identification (LID) performance is the impact of varying channel conditions. In house experiments show that LID performance drops between 10-12 percent across channels. The focus of this project is to mitigate the impact of channel conditions and artifacts on LID, and to provide an understanding of how channel robust models may be created by combining data from across corpora. Our approach involved creating composite cross-channel language models from multiple corpora that were tested with data from three corpora whose results were compared to results obtained from same-channel and pure cross-channel experiments. Our hypotheses were that 1) same-channel models would be the most accurate, 2) purely cross-channel models would be considerably less accurate, and 3) composite model accuracy would fall in between that of the same-channel and cross-channel models. Results were surprising: while pure cross-channel tests performed the worst, with an average of 11% loss in accuracy, composite models outperformed matched condition models in tests by 0.5% to 15% for all four combinations tested. These findings provide evidence that cross-channel augmentation does not just make for more robust LID models across channel conditions, but improves model robustness within channel conditions as well.