8.2 Johns Hopkins University

Dr. Hynek Hermansky is a Full Professor of Electrical and Computer Engineering at the Johns Hopkins University in Baltimore, Maryland. He is also a Professor (on leave of absence) at the Brno University of Technology, Czech Republic; an Adjunct Professor at the Oregon Health and Sciences University, Portland, Oregon; and an External fellow at the International Computer Science Institute at Berkeley, California. He is a Fellow of IEEE for "invention and development of perceptually-based speech processing methods." His main area of research is robust acoustic processing of speech. Among the techniques that he has pioneered are: Perceptual Linear Prediction (PLP), RASTA modulation frequency filtering, multi-stream recognition of speech, data-guided feature extraction, frequency-domain PLP, and posterior-based features for speech recognition. He has applied these methods to speech recognition, speaker recognition, and speech and audio coding. A number of the robust processing techniques that he pioneered are now in wide usage worldwide.

Dr. Hermansky has been working in speech processing for over 30 years and has held the following positions: Director of Research at the IDIAP Research Institute, Martigny, Switzerland; Research Fellow at the University of Tokyo; Research Engineer at Panasonic Technologies in Santa Barbara, California; Senior Member of Research Staff at US WEST Advanced Technologies; and Professor and Director of the Center for Information Processing at OHSU, Portland, Oregon.

In his professional activities, Dr. Hermansky served as Technical Chair at the 1998 ICASSP in Seattle and an Associate Editor for IEEE Transaction on Speech and Audio, and is a member of the Organizing Committee for ICASSP 2011 in Prague. He is member of the Editorial Board of Speech Communication, holds 7 US patents, and authored or co-authored over 200 papers in reviewed journals and conference proceedings.

8.3 University of Maryland

Dr. Shihab Shamma is a Full Professor, Department of Electrical Engineering, University of Maryland (UMD), College Park, with a joint appointment at the Institute for System Research, UMD. His research deals with questions in computational neuroscience, speech and audio processing, neuromorphic engineering, and the development of microsensor systems for experimental research and neural prostheses. Primary focus has been on studying the computational principles underlying the processing and recognition of complex sounds in the auditory system, and the relationship between auditory and visual processing. Signal processing algorithms inspired by data from neurophysiological and psychoacoustical experiments are being developed and applied in a variety of systems such as speech and voice recognition, diagnostics in industrial manufacturing, and underwater and battlefield acoustics.

Dr. Shamma has developed novel models of auditory cortical processing, inspired by physiological and psychoacoustic experiments over the last decade that have revealed a rich and flexible representation of the perceptually significant features of sound. The resultant higher-dimensional cortical representation (a frequency/rate/scale vector as a function of time) is able to separate speech sounds from other unwanted signals, such as various types of noise or music. He has been able to use this unique separation property to design multi-dimensional filters that enhance the quality of noisy speech and suppress the noise. He has also used the cortical representation to design front ends to a speech activity detector (SAD) for many types of noise. He has tested his SAD with animal vocalizations, music, and environmental sounds, including jet engine noise, factory noise, destroyer engine sounds, military vehicles, cars, and speech babble recorded in different environments like restaurants, airports, and exhibition halls, and, more recently, reverberant speech. His speech activity detector has been found to be superior to other SAD methods in independent tests by several government agencies.

Dr. Shamma is a Fellow of the Acoustical Society of America, Senior Member of IEEE, member of the Association for Research in Otolaryngology, and member of the Society for Neuroscience. He is also Action Editor for the Journal of Computational Neuroscience, Academic Editor for PLoS, and Academic Board Member for Trends in Cognitive Sciences. He has been Co-organizer and Director of numerous Workshops and Symposia including most recently the Annual Telluride Workshops on Neuromorphic Engineering (1997- present, partially funded by NSF, ONR, DARPA, and the Whitaker and Gatsby Foundations), and the NIPS and COSYNE workshops on Neural Mechanisms of Music Perception (1999), Thalamo-cortical Processing (2002), and Attention and Streaming (2006). He is the holder of a patent on "A Cochlear Filter Bank with Switched Capacitor Circuits", and has patents pending on "Intelligibility Assessment Using Spectro-temporal Modulations" and "Speech Discrimination Based on Multiscale Spectro-Temporal Modulations."

8.4 Cambridge University

Dr. Mark Gales will serve as the primary person in charge of the RATS work at Cambridge University. Professor Phil Woodland will serve as senior technical advisor.

Dr. Mark Gales is Reader in Information Engineering at Cambridge University where he has been a member of faculty staff since 1999. Prior to this he was a Research Staff Member at IBM. He has worked on speech processing for over 18 years, with a particular interest in acoustic modeling, acoustic environment robustness, and speaker adaptation. He has participated in a range of DARPA/NIST automatic speech recognition evaluations since 1994, at both Cambridge and IBM. These include recent evaluations under the DARPA EARS and GALE programs. Dr. Gales has been awarded multiple IEEE paper awards for his work on speech recognition. From 2001 to 2004 he was a member of the IEEE Speech Technical Committee. He is currently an Associate Editor of the IEEE Signal Processing Letters, IEEE Transactions on Audio Speech and Language, and on the editorial board of Computer Speech and Language journal. He has published over 100 papers, primarily in the area of acoustic modeling for speech recognition.

Dr. Gales has extensive experience in noise robustness, speaker adaptation, and speech recognition. He developed one of the first approaches to model-based compensation for acoustic models to handle noise conditions – parallel model combination. This model-based approach to handling high levels of background noise has been refined to handle a range of important techniques to make it applicable to state-of-the-art speech processing applications, including discriminative adaptive training , and improving efficiency to allow its application to large systems. He developed one of the standard approaches for speech recognition and adaptive training, Constrained Maximum Likelihood Linear Regression (CMLLR), as well as refining this to be more suited for speaker adaptation in the presence of high levels of background noise. Under the DARPA EARS and GALE programs he has been involved in developing large vocabulary transcription systems for English, Mandarin and Arabic. In addition, he has worked on sequence kernels that form an important part of speaker identification and verification systems and has recently been applied to handling high-levels of background noise.

Phil Woodland is Professor of Information Engineering at the University of Cambridge where he leads the Speech Group. He has been a member of faculty staff at Cambridge since 1989. He has published almost 200 academic papers in the area of speech and language technology. Many of these papers are highly cited, including the most highly cited paper to appear in the journal *Computer Speech and Language*. He has won best paper awards for work in the areas of speaker adaptation and discriminative training of large vocabulary speech recognition systems. He has also worked on statistical language modeling, auditory modeling, statistical speech synthesis, and spoken document retrieval. His team has developed large-vocabulary speech recognition systems that have frequently given the lowest error rates in international research evaluations organized by the US Government. He was one of the original co-authors of the widely-used HTK toolkit, which now has about 100,000 registered users worldwide, and has continued to play a major role in its development. He was a member of the IEEE SPS Speech Technical Committee from 1999 to 2003, a member of the Editorial Board of Computer Speech and Language from 1994 until 2009, and is a current member of the Editorial Board of Speech Communication.

Prof. Woodland has extensive experience in many areas of speech technology, particularly in speech recognition systems, and has worked in the field for the last 25 years. Of special relevance to RATS is his work on linear transform-based adaptation approaches, in which he developed the widely-used MLLR technique, as well as techniques for unsupervised adaptation in low recognition rate situations, via lattice-based adaptation, and discriminative adaptation based methods. He has developed widely used discriminative training techniques and shown that such techniques are useful for situations where test data is poorly matched to the training. He has worked on developing speech transcription technology for many years, including work on the DARPA EARS and GALE programs, for both of these he was the PI at Cambridge University. During this time his team has developed systems in a number of languages including English, Mandarin and Arabic.

8.5 Brno University of Technology

Dr. Lukas Burget will serve as the primary person in charge of the RATS work at BUT. Working closely with Dr. Burget will be Dr. Pavel Matejka.

Dr. Lukas Burget has been an Assistant Professor at the Faculty of Information Technology (FIT), University of Technology, Brno, Czech Republic, since 2004. He serves as Scientific Director of the Speech@FIT research group. From 2000 to 2002, he was a visiting researcher at OGI, Portland, OR, under supervision of Professor Hynek Hermansky. In 2008, he was invited to lead the "Robust Speaker Recognition over Varying Channels" team at the Johns Hopkins University CLSP summer workshop, and will lead the team of BOSARIS workshop in 2010.

Dr. Burget's scientific interests are in the field of speech processing, namely acoustic modeling for speech, speaker and language recognition, including their software implementations. He has authored or co-authored more than 40 papers in journals and conferences. He was leader of the top-performing teams in the NIST LID 2005 and 2007 evaluations and the NIST SID 2006 and 2008 evaluations, scoring first in a remarkable number of conditions. He contributed significantly to the team developing AMI LVCSR systems successful in the NIST RT 2005, 2006 and 2007 evaluations.

Dr. Burget has participated in EU-sponsored projects "Multimodal Meeting Manager" (M4), "Augmented MultiParty Interaction" (AMI), and "Augmented MultiParty interaction with

Distant Access" (AMIDA), as well as in several projects sponsored at the local Czech level. Currently, he is participating in the EU project "Mobile Biometry". He is principal investigator of US-Air Force EOARD sponsored project "Improving the capacity of language recognition systems to handle rare languages using radio broadcast data".

Dr. Pavel Matejka is senior researcher in the Speech@FIT research group, Department of Computer Graphics and Multimedia, Faculty of Information Technology, Brno University of Technology (BUT). He was a visiting scientist with the Anthropic speech processing group at the Oregon Graduate Institute of Science and Technology in 2002-2003. He has participated in the EU projects M4, AMI, and AMIDA mentioned above, and in language identification projects sponsored by US Air Force EOARD and the Czech Ministry of Defense. He took part in NIST 2005, 2007 and 2009 language recognition and NIST 2006 and 2008 speaker recognition evaluations, where BUT had excellent results. He is currently leading BUT's speaker identification activities in the Mobile Biometry EU project.

Dr. Matejka is author or co-author of more than 15 papers in journals and at reviewed international conferences. His research interests include speaker recognition, language identification, speech recognition (namely, phone recognition based on novel feature extraction using temporal patterns and neural networks). He is active in keyword spotting and on-line implementation of speech processing algorithms. He was finalist in the Student paper contest at ICASSP 2006 in Toulouse and participated in the JHU workshop group "Recovery from Model Inconsistency in Multilingual Speech Recognition" in 2007.

8.6 Time Commitments

Table 8-1 shows the planned time commitments for the Key Personnel for calendar years 2010, 2011, and 2012. The numbers for 2010 assume that the effort will start 15 August 2010, so 2010 is only a partial year. It is anticipated that a relatively small effort will be devoted for Technical Area 1 during the first six months of the project while data is being collected by the Data Collection team.

For the BBN Key Personnel, we have shown the time commitments both in terms of hours and percentage time. At certain universities, personnel time is accounted for by percentage of time rather than by hours, while at others, time is specified in months.

Key Individual	Project	Pending/Current	2010	2011	2012
John Makhoul	RATS	Proposed	90 hrs (15%)	528 hrs (30%)	616 hrs (35%)
	GALE	Current	350 hrs (60%)	403 hrs (23%)	n/a
Richard Schwartz	RATS	Proposed	100 hrs (17%)	792 hrs (45%)	880 hrs (50%)
	GALE	Current	410 hrs (70%)	440 hrs (25%)	n/a
Spyros Matsoukas	RATS	Proposed	153 hrs (25%)	1196 hrs (65%)	1472 hrs (80%)
	GALE	Current	429 hrs (70%)	460 hrs (25%)	n/a
Hynek Hermansky	RATS	Proposed	12%	25%	25%
	IARPA/ARL Speaker ID	Current	12.5%	12.5%	n/a
Shihab Shamma	RATS	Proposed	0.25 month	1 month	1 month
	IARPA/ARL Speaker ID	Current	1 month	2 months	n/a
	ONR	Pending	0.25 month	1 month	1 month
Mark Gales	RATS	Proposed	50 hrs	200 hrs	200 hrs
	GALE	Current	70	70	n/a
Phil Woodland	RATS	Proposed	20 hrs	50 hrs	50 hrs
	GALE	Current	70 hrs	70 hrs	n/a
Lukas Burget	RATS	Proposed	10%	50%	40%
	IARPA Speaker ID	Current	41%	40%	40%
Pavel Matejka	RATS	Proposed	10%	55%	40%
	IARPA Speaker ID	Current	61%	35%	40%

Table 8-1 Key personnel and their time commitments.

8.7 Selected Bibliographies

Below is a selected bibliography from each of the PATROL sites with references pertaining to the RATS program.

8.7.1 BBN

R. Prasad, S. Tsakalidis, I. Bulyko, C. Kao, and P. Natarajan, "Pashto Speech Recognition with Limited Pronunciation Lexicon," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Dallas, TX, March 2010.

S. Tsakalidis, R. Prasad, and P. Natarajan, "Context-dependent Pronunciation Modeling for Iraqi ASR," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Taiwan, April 2009.

M. Siu, H. Gish and X. Yang, "Discriminatively Trained GMMs for Language Classification Using Boosting Methods", *IEEE Trans. on Audio, Speech and Language Processing*, Vol. 17, No. 1, 2009.

L. Nguyen, T. Ng, K. Nguyen, Rabih Zbib, and John Makhoul, "Lexical and phonetic modeling for Arabic automatic speech recognition," *InterSpeech*, Brighton, UK, pp. 712-715, 2009.

S. Novotney, R. Schwartz, and J. Ma, "Unsupervised Acoustic and Language Model Training with Small Amounts of Labeled Data," *ICASSP*, Taipei, China, pp. 4297-4300, 2009.

J. Ma and R. Schwartz, "Unsupervised versus Supervised Training of Acoustic Models," *Interspeech*, Brisbane, Australia, pp. 2378-2381, 2008.

J. Au-Yeung and M. Siu, "Evaluation of the robustness of the polynomial segment models to noisy environments with unsupervised adaptation", *Speech Communications*, Vol. 50, 2008.

K. Subramanian, R. Prasad, and P. Natarajan, "Optimal Estimation of Rejection Thresholds for Topic Spotting," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Honolulu, HI, April 2007.

D. R. H. Miller, M. Kleber, C. Kao, O. Kimball, T. Colthurst, S. A. Lowe, R. Schwartz, and H. Gish, "Rapid and Accurate Spoken Term Detection", *Interspeech*, pp. 314-371, Antwerp, Belgium, 2007.

J. Ma, S. Matsoukas, O. Kimball, and R. Schwartz, "Unsupervised Training on Large Amounts of Broadcast News Data," *ICASSP*, Toulouse, France, vol.3, pp. 1056-1059, 2006.

S. Saleem, R. Prasad, and P. Natarajan, "Colloquial Iraqi ASR for Speech-to-Speech Translation," *Intl. Conf. on Spoken Language Proce*ssing, Pittsburg, PA, September, 2006.

B. Zhang, S. Matsoukas, R. Schwartz: "Recent progress on the discriminative region-dependent transform for speech feature extraction", *Interspeech*, Pittsburgh, PA, September, 2006.

H. Jin, R. Schwartz, F. G. Walls, and S. P. Sista, "Method and Apparatus for Score Normalization for Information Retrieval Applications", US Patent No. 6,651,057, issued on November 18, 2003, and US Patent No. 7,062,485, issued on June 13, 2006 (continuation).

L. Zhai, M. Siu, X. Yang and H. Gish, "Discriminatively trained Language Models Using Support Vector Machines for Language Identification," *Odyssey*, June 2006.

B. Zhang, S. Matsoukas, and R. Schwartz, "Discriminatively Trained Region Dependent Feature Transforms for Speech Recognition," *ICASSP*, Toulouse, France, vol.1, pp.313-316, 2006.

B. Zhang, S. Matsoukas, and R. Schwartz, "Recent Progress on the Discriminative Region-dependent Transform for Speech Feature Extraction," *Interspeech*, Pittsburgh, PA, pp. 1495-1498, 2006.

K.Y. Leung, M.W. Mak, M. Siu, and S.Y. Kung, "Speaker Verification Using Adapted Articulatory Feature-based Conditional Pronunciation Modeling," *IEEE Int. Conference on Acoustics, Speech and Signal Processing*, pp. 181-184, Philadelphia, PA, 2005.

B. Xiang, L. Nguyen, S. Matsoukas, and R. Schwartz, "Cluster-dependent acoustic modeling," *ICASSP 2005*, Philadelphia, Pennsylvania, 2005.

K.K. Wong, M. Siu, "Automatic language identification using discrete hidden Markov model," *International Conference on Spoken Language Processing*, pp. 1633-1636, Jeju Island, Korea, 2004.

W. Andrews and J. Hernandez-Cordero, "Using Speech Compression for Speaker Detection", *Speaker Odyssey Workshop*, Toledo, Spain, 2004.

L. Nguyen and B. Xiang, "Light supervision in acoustic model training," *ICASSP*, Montreal, Canada, 2004.

D. Reynolds, W. Andrews, J. Campbell, J. Navrtil, B. Peskin, A. Adami, Q. Jin, D. Kluscek, J. Abramson, R. Mihaescu, J. Godfrey, D. Jones, and B. Xiang, "SuperSID Project: Exploiting High-Level Information for High Accuracy Speaker Recognition", *ICASSP*, Hong Kong, China, 2003.

J. Navrtil, Q. Jin, W. Andrews and J. Campbell, "Phonetic Speaker Recognition Using Maximum-Likelihood Binary-Decision Tree Models", *ICASSP*, Hong Kong, China, 2003.

L. Nguyen, X. Guo, R. Schwartz, and J. Makhoul, "Japanese broadcast news transcription system," *ICSLP*, Denver, Colorado, 2002.

M. Siu, A. Chan, "Robust Speech Recognition Against Short-Time noise," *International Conference on Spoken Language Processing*, Vol. 2, pp. 1049-1052, 2002.

R. Prasad, L. Nguyen, R. Schwartz, and J. Makhoul, "Automatic Transcription of Courtroom Speech," *Int. Conf. Spoken Language Processing*, Denver, CO, pp.1745-1748, 2002.

R. Prasad, L. Nguyen, R. Schwartz, and J. Makhoul, "MeetingLogger: Rich Transcription of Courtroom Speech," *Int. Conf. Human Language Technology*, San Diego, CA, Morgan Kaufmann Publishers, pp. 303-306, March 24-27, 2002.

S. Sista, R. Schwartz, T. Leek, and J. Makhoul, "An Algorithm for Unsupervised Topic Discovery from Broadcast News Stories," *Int. Conf. Human Language Technology*, San Diego, CA, Morgan Kaufmann Publishers, pp. 110-114, 2002.

W. Andrews, M. Kohler, and J. Campbell, "Phonetic Speaker Recognition," *EuroSpeech*, Aalborg, Denmark, 2001.

H. Jin, R. Schwartz, S. Sista, F. Walls,"Topic Tracking for Radio, TV Broadcast, and Newswire," *Eurospeech*, Budapest, Hungary, pp. 2439-2442, 1999.

H. Jin, S. Matsoukas, R. Schwartz, and F Kubala, "Fast Robust Inverse Transform SAT and Multi-stage Adaptation", *DARPA Broadcast News Transcription and Understanding Workshop*, Leesburg VA, 1998.

H. Jin, F. Kubala, and R. Schwartz, "Automatic speaker clustering," *DARPA Speech Recognition Workshop*, Chantilly VA, 1997.

S. Matsoukas, R. Schwartz, H. Jin, and L. Nguyen, "Practical Implementations of Speaker Adaptive Training", *DARPA Speech Recognition Workshop*, Chantilly VA, 1997.

T. Anastasakos, J. McDonough, and J. Makhoul, "Speaker Adaptive Training: A Maximum Likelihood Approach to Speaker Normalization", *ICASSP*, Washington DC, 1997.

S. Matsoukas, R. Schwartz, H. Jin, and L. Nguyen, "Practical Implementations of Speaker-Adaptive Training", *IEEE Recognition Workshop*, Chantilly, Virginia, pp 133-138, 1997.

A. Anastasakos, J. Mcdonough, R. Schwartz, and J. Makhoul, "A Compact Model for Speaker-Adaptive Training," *ICSLP*, 1996

R. Schwartz, A. Derr, A. Wilgus, and J. Makhoul, "Speaker Verification IR&D Final Report," BBN Technical Report No. 6661, Nov. 1987.

H. Gish, M. Krasner, W. Russell, and J. Wolf, "Methods and experiments for text-independent speaker recognition over telephone channels," *IEEE Int. Conf. Acoust., Speech, Signal Processing*, pp. 865-868, 1986.

V. Viswanathan, C.M. Henry, R. Schwartz, and S. Roucos, "Evaluation of Multisensor Speech Input for Speech Recognition in High Ambient Noise," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Tokyo, Japan, pp. 85-88, 1986.

S. Roucos, V. Viswanathan, C. Henry, and R. Schwartz, "Word Recognition Using Multisensor Speech Input in High Ambient Noise," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Tokyo, Japan, pp. 737-740, 1986.

H. Gish, K. Karnofsky, M. Krasner, S. Roucos, R. Schwartz, and J. Wolf, "Investigation of Text-Independent Speaker Identification Over Telephone Channels," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Tampa, FL, pp. 379-382, 1985.

M. Krasner, J. Wolf, K. Karnofsky, R. Schwartz, S. Roucos, and H. Gish, "Investigation of text-independent speaker identification techniques under conditions of variable data," *IEEE Int. Conf. Acoust., Speech, Signal Processing*, pp. 18B.5.1-18B.5.4, 1984.

J. Wolf, M. Krasner, K. Karnofsky, R. Schwartz, and S. Roucos, "Further Investigations of Probabilistic Methods for Text-Independent Speaker Identification", *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Boston, MA, pp. 551-554, 1983.

R. Schwartz, S. Roucos, and M. Berouti, "The Application of Probability Density Estimation to Text-Independent Speaker Identification," *ICASSP*, Paris, France, pp. 1649-1652, May 1982.

M. Berouti, R. Schwartz, and J. Makhoul, "Enhancement of Speech Corrupted by Acoustic Noise," *ICASSP*, Washington, DC, pp. 208-211, 1979. Also in J.Lim (Ed.), *Speech Enhancement*, New Jersey; Prentice-Hall, 1983.

8.7.2 Johns Hopkins University

M. Athineos, H. Hermansky, and D. P. W. Ellis, "LP-TRAP: Linear Predictive Temporal Patterns," *Proc. ICSLP*, Jeju, S. Korea, pp. 1154-1157, 2004.

S. Ganapathy,S. Thomas and H. Hermansky, "Temporal Envelope Subtraction for Robust Speech Recognition Using Modulation Spectrum", *Proc. ASRU*, Merano, Italy,2009.

S. Ganapathy, S. Thomas and H. Hermansky, "Robust Spectro-temporal Features based on Autoregressive Models of Hilbert Envelopes", *Proc. ICASSP 2010*, Dallas, TX, 2010.

S. Thomas, S. Ganapathy and H. Hermansky, "Tandem Representations of Spectral Envelope and Modulation Frequency Features for ASR", *Proc. Interspeech*, Brighton, UK, 2009.

S. Thomas, S. Ganapathy and H. Hermansky, "Recognition of Reverberant Speech Using Frequency Domain Linear Prediction", *IEEE Signal Processing Letters*, 2008.

H. Hermansky, "Speculations on Knowledge Versus Data in ASR," *Proc. ESCA-NATO Tutorial and Research Workshop on Robust Speech Recognition for Unknown Communication Channels*, Pont-a-Mousson, France, 1997.

F. Valente and H. Hermansky, "Discriminant Linear Processing of Time-Frequency Plane," *Proc. International Conference on Spoken Language Processing*, Pittsburgh, PA, 2006.

H. Hermansky and D.P.W. Ellis and S. Sharma, "Connectionist Feature Extraction for Conventional HMM Systems", *Proc. ICASSP '00*, 2000.

H. Hermansky and N. Morgan, "RASTA Processing of Speech", *IEEE Transactions on Speech and Audio Processing*, Vol. 2, No. 4, pp. 587-589, 1994.

H. Hermansky, "Robust Speech Recognition," in Cole, Hirshman et al., "The Challenge of Spoken Language Systems, Research Directions for the Nineties," *IEEE Transactions on Speech and Audio Processing*, Vol. 3, No. 1, pp. 1-21, 1995.

H. Hermansky and S. Sharma, "TRAPS - Classifiers of Temporal Patterns," *Proc. International Conference on Spoken Language Processing*, Sydney, Australia, 1998.

H. Hermansky and S. Sharma, "Temporal Patterns (TRAPS) in ASR of Noisy Speech," *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, Phoenix, AZ 1999.

H. Hermansky and P. Fousek, "Multi-resolution RASTA filtering for TANDEM-based ASR," *Proc. European Conference on Speech Communication and Technology*, Lisbon, Portugal, 2005.

N. Mesgarani, G.S.V.S. Sivaram, S. Nemala, H. Hermansky, "Discriminant Spectrotemporal Features for Phoneme Recognition", *Proc. Interspeech 2009*, Brighton, UK, 2009.

G.S.V.S. Sivaram, S. Nemala, M.Elhilali, T. Tran, and H. Hermansky, "Sparse Coding for Speech Recognition", *ICASSP 2010*, Dallas, TX, 2010.

J. Pinto, Y. Bayya, H. Hermansky, M. Magimai-Doss, "Exploiting Contextual Information For Improved Phoneme Recognition", *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2008.

F. Valente and H. Hermansky, "Hierarchical and Parallel Processing of Modulation Spectrum for ASR applications", *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2008.

H. Fletcher and R.H. Galt, "The perception of speech and its relation to telephony", *The Journal of the Acoustical Society of America*, 1950.

A. Boothroyd and S. Nittrouer, "Mathematical treatment of context effects in phoneme and word recognition", *Journal of the Acoustical Society of America*, Vol 84(1), pp. 101-114, July 2008.

H. McGurk and J. MacDonald, "Hearing lips and seeing voices", *Nature 264*, pp. 746 – 748, December 1976.

F. Valente and H. Hermansky, "Combination of Acoustic Classifiers based on Dempster-Shafer Theory of Evidence." *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, Honolulu, HI, 2007.

8.7.3 University of Maryland

S. Shamma, "Encoding sound timbre in the auditory system" *IETE J. Res.* 49(2), 193-205, 2003.

L. Atlas and S. Shamma, "Joint acoustic and modulation frequency", *EURASIP J. Appl. Sig. Proc. 7*, Jeju, S. Korea, pp. 668-675, 2003.

P. Julián, A. Andreou, L. Riddle, S. Shamma, D. Goldberg, and G. Cauwenberghs. "A Comparative Study of Sound Localization Algorithms for Energy Aware Sensor Network Nodes," *IEEE Trans. Cir. Sys.*, 51(4), 2004

A. Schai, S Shamma "A Neuromorphic Sound Localizer for a Smart MEMS System," *Analog Integrated Circuits and Signal Processing*, 39, 267–273, 2004

N. Massgarani, M. Slaney, and S. Shamma, "Content-Based Audio Classification Based On Multiscale Spectro-Temporal Features", *IEEE Trans. Audio and Speech*, 2005.

N. Massgarani, M. Slaney, and S. Shamma, "Content-Based Audio Classification Based On Multiscale Spectro-Temporal Features", *IEEE Trans. Audio and Speech*, 2005.

D. Zotkin, S. Shamma, P. Ru, and R. Duraiswami, L. Davis "Neuromimetic sound representation for percept detection and manipulation," *EURASIP J. Applied Signal Processing*, 2005.

D. Zotkin, S. Shamma, P. Ru, and R. Duraiswami, L. Davis "Neuromimetic sound representation for percept detection and manipulation," *EURASIP J. Applied Signal Processing*, 2005.

J. Fritz, S. Shamma, M. Elhiliali, "Active Listening: Task-dependent plasticity of receptive fields in primary auditory cortex", *Hearing Res*, 206, 159-176, 2005.

T. Chi, P. Ru, and S. Shamma, "Multiresolution spectrotemporal analysis of complex sounds", *J. Acoust. Soc. Am*, 118, pp.887, 2006.

T. Chi and S. Shamma, "Spectrum restoration from multiscale auditory phase singularities by generalized projections", *IEEE Trans. Audio and Speech*, 14(4), 118, 887 2006.

J. Fritz, M. Elhilali, S. David, S Shamma "Auditory Attention – focusing the search light on sound", *Curr. Opin. Neurobiol*, 17(4):437-55, 2007

M. Elhilali and S. Shamma "The cocktail party problem-with a cortical twist", *J. Acoust. Soc. Am.*, 124(6), 2008

S. Shamma, "On the emergence and awareness of auditory objects," *PLoS biology*, 6(6) 2008

S. Shamma, "Characterizing Auditory Receptive Fields", *Neuron*, 58(6), p.829-831, 2008

M. Elhilali, J Xiang, S Shamma , and J Simon "Interaction between attention and bottom-up saliency mediates the representation of foreground and background in an auditory scene," *PLoS* 7(6) 2009.

N. Mesgarani, J Fritz, and S Shamma, "A Computational Model of Rapid Task-Related Plasticity of Auditory Cortical Receptive Fields," *J. Comp. Neuroscience*, 2009.

8.7.4 Cambridge University

F. Flego and M.J.F. Gales, "Discriminative Adaptive Training with VTS and JUD," *Proc. IEEE Automatic Speech Recognition and Understanding Workshop*, December 2009.

H. Xu, M.J.F. Gales, and K.K. Chin, "Improving Joint Uncertainty Decoding Performance by Predictive Methods for Noise Robust Speech Recognition," *Proc. IEEE Automatic Speech Recognition and Understanding Workshop*, December 2009.

M.J.F. Gales, "Maximum Likelihood Linear Transformations for HMM-based Speech Recognition," *Computer Speech and Language*, Volume 12, 1998.

D.K. Kim and M.J.F. Gales, "Noisy Constrained Maximum Likelihood Linear Regression for Noise-Robust Speech Recognition," submitted to *IEEE Transactions on Audio Speech and Language Processing*.

M. Tomalin, J. Park, F. Diehl, M.J.F. Gales, and P.C. Woodland, "Recent Improvements to the Cambridge Arabic Speech-to-Text Systems," *ICASSP 2010*, Dallas, TX, 2010.

C. Longworth and M.J.F. Gales, "Combining Derivative and Parametric Kernels for Speaker Verification," *IEEE Transactions on Audio Speech and Language Processing*, May 2009.

M.J.F. Gales and F. Flego, "Discriminative Classifiers with Adaptive Kernels for Noise Robust Speech Recognition," to appear *Computer Speech and Language*, 2010.

C.J. Leggetter and P.C. Woodland, "Maximum likelihood linear regression for speaker adaptation of continuous density Hidden Markov models," *Computer speech and language*, 1995.

L.F. Uebel and PC Woodland, "Speaker adaptation using lattice-based MLLR," *ITRW on Adaptation Methods for Speech Recognition*, 2001.

L. Wang and P.C. Woodland, "MPE-based discriminative linear transforms for speaker adaptation," *Computer Speech & Language*, Vol. 22, No. 3, 2008.

K. Yu, M. Gales and P.C. Woodland, "Unsupervised adaptation with discriminative mapping transforms," *IEEE Trans. ASLP*, Vol. 17, No. 4, 2009.

D. Povey and P.C Woodland, "Minimum phone error and I-smoothing for improved discriminative training," *Proc. ICASSP*, Orlando, FL, 2002.

R. Cordoba, P.C. Woodland, and M. Gales, "Improved cross-task recognition using MMIE training," *Proc. ICASSP*, Orlando, FL, 2002.

M. Tomalin, J. Park, F. Diehl, M.J.F Gales and P.C. Woodland "Recent improvements to the Cambridge Arabic speech-to-text systems" *Proc. ICASSP'2010*, Dallas, TX, 2010.

8.7.5 Brno University of Technology

N. Brümmer, L. Burget, J. Černocký, O. Glembek, F. Grézl, M. Karafiát, D. van Leeuwen, P. Matějka, P. Schwarz, and A. Strasheim, "Fusion of heterogeneous speaker recognition systems in the STBU submission for the NIST speaker recognition evaluation 2006," *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 15, No. 7, pp. 2072-2084, 2007.

N. Brümmer, A. Strasheim, V. Hubeika, P. Matějka, L. Burget, and O. Glembek, "Discriminative Acoustic Language Recognition via Channel-Compensated GMM Statistics," *Proc. Interspeech,* Brighton, UK, pp. 2187-2190, 2009.

L. Burget, P. Matějka, and J. Černocký, "Discriminative Training Techniques for Acoustic Language Identification", *Proc. ICASSP*, Toulouse, FR, pp. 209-212, 2006.

L. Burget, P. Matějka, P. Schwarz, O. Glembek, and J. Černocký, "Analysis of feature extraction and channel compensation in GMM speaker recognition system", *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 15, No. 7, pp. 1979-1986 2007.

L. Burget, N. Brümmer, D. Reynolds, P. Kenny, J. Pelecanos, R. Vogt, F. Castaldo, J. Dehak, R. Dehak, O. Glembek, Z Karam, J. Noecker Jr., Y. Na Hye, C. Costin Ciprian, V. Hubeika, S. Kajarekar, N. Scheffer, and J. Černocký, "Robust Speaker Recognition Over Varying Channels", *JHU* , Baltimore, MD, p. 81, 2008.

L. Burget, P. Matějka, V. Hubeika, and J. Černocký, "Investigation into variants of Joint Factor Analysis for speaker recognition", *Proc. Interspeech*, Brighton, UK, pp. 1263-1266, 2009.

L. Burget, M. Fapšo, V. Hubeika, O. Glembe, M. Karafiát, M. Kockmann, P. Matějka, P. Schwarz, and J. Černocký, "BUT system for NIST 2008 speaker recognition evaluation", *Proc. Interspeech*, Brighton, UK, pp. 2335-2338, 2009.

O. Glembek, L. Burget, N. Dehak, N. Brümmer and P. Kenny: "Comparison of scoring methods used in speaker recognition with joint factor analysis", accepted to *ICASSP*, 2009.

Z. Jancik, O. Plchot, N. Brümmer, L. Burget, O. Glembek, V. Hubeika, M. Karafiat, P. Matejka, T. Mikolov, A. Strasheim and J. Černocký, "Data selection and calibration issues in automatic language recognition", submitted to *Odyssey 2010*, Brno, 2010.

M. Kockmann, L. Burget, "Contour modeling of prosodic and acoustic features for speaker recognition", *Proc. IEEE Workshop on Spoken Language Technology*, p. 4, 2008.

P. Matějka, P. Schwarz, L. Burget, and J. Černocký, "Use of anti-models to further improve state-of-the-art PRLM language recognition system", *Proc. ICASSP*, Toulouse, FR, pp. 197-200, 2006.

P. Matějka, L. Burget,, P. Schwarz, O. Glembek, M. Karafiát, F. Grézl, J. Černocký, D. van Leeuwen, N. Brümmer, A. Strasheim, "STBU system for the NIST 2006 speaker recognition evaluation", *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, Honolulu, HI, pp. 221-224, 2007.

P. Matejka, L. Burget, O. Glembek, P. Schwarz, V. Hubeika, M. Fapso, T. Mikolov, and O. Plchot, "BUT system description for NIST LRE 2007", *NIST LRE 2007 Workshop*, Orlando, FL, 2007.

T. Mikolov, O. Plchot, O. Glembek, P. Matějka, L. Burget, and J. Černocký, "PCA-based Feature Extraction for Phonotactic Language Recognition", submitted to *Odyssey 2010, Brno*, 2010.

P. Oldrich, V. Hubeika, L. Burget, P. Schwarz, P. Matějka, and J. Černocký, "Acquisition of Telephone Data from Radio Broadcasts with Applications to Language Recognition: Technical Report", *technical report*, BUT, January 2009.

P. Schwarz, "Phoneme recognition based on long temporal context", *PhD Thesis, Brno*, BUT, 2009.

9 Project Management and Interaction Plan

9.1 Project Management

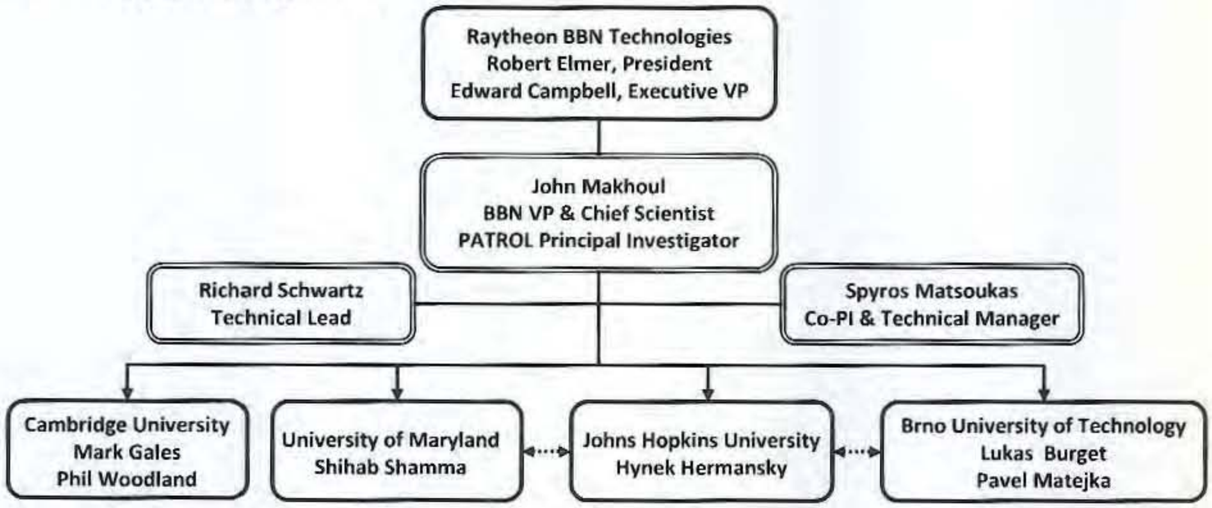

Figure 9-1 Organizational chart for the PATROL Team.

Figure 9-1 shows the management organization for the project. Dr. John Makhoul will serve as Principal Investigator and the overall Program Manager for this effort; he will coordinate and manage the activities of the BBN-led PATROL Team in the RATS Program. Dr. Makhoul is an experienced project leader and program manager. He has served as PI on a large number of DARPA and other government projects. Most recently, he managed and coordinated the BBN-led team under the DARPA EARS program and is currently the PI on the BBN-led team under the DARPA GALE program. In both programs, the BBN-led teams have been consistently top performers.

Under the BBN GALE effort, Dr. Makhoul oversees the activities of a dozen subcontractors, distributed geographically across multiple locations, from California to Europe. The work under the GALE program is quite diverse; it includes work on speech recognition, machine translation, distillation, the linguistic annotation of text, and the integration of technologies into operational systems that take speech or text from multiple languages and produce an English translation. The resulting technologies have been transitioned into BBN's Broadcast Monitoring and Web Monitoring products which have been deployed in at least 18 Government locations for 24/7 operation. The GALE program is expected to end in April 2011, but our experience in running such a large and successful project will serve us well in managing the BBN-led PATROL Team in the RATS program.

Dr. Makhoul will be assisted at BBN by two senior researchers, both of whom have had prominent and important roles in the EARS and GALES programs. Spyros Matsoukas, who has demonstrated in the GALES program first-rate skills in performing technical management across many sites, will serve as co-PI and Technical Manager. He will manage the technical activities at BBN and across the other four sites on the PATROL team. Rich Schwartz, who has been a fountain of innovation in various aspects of speech processing and pattern recognition for over three decades, will serve as Technical Lead for the effort.

Having been on the BBN-led team on GALE for the last four and a half years, Professors Mark Gales and Phil Woodland of Cambridge University have worked closely with Makhoul, Matsoukas, and Schwartz, and together have developed one of the most advanced multi-lingual speech recognition systems in the world. That strong working relationship will continue during the RATS project.

Professor Hynek Hermansky of Johns Hopkins University (JHU), who has an ongoing joint appointment with the Brno University of Technology, has a very strong relationship with Drs. Lukas Burget and Pavel Matejka there, both having performed research with him in the past and have attended summer workshops at JHU. Currently, Burget and Matejka serve as consultants on one of Hermansky's projects in the US.

Professor Shihab Shamma of the University of Maryland (UMD) currently has a joint project on speaker identification with Professor Hermansky, sponsored by IARPA under the BEST program. Furthermore, both of them work on auditory modeling of sound and will be able to join up to provide the PATROL team unparalleled expertise in that area, which we expect to be very important in isolating the speech part of the noisy speech signal. Shamma and Hermansky will be providing auditory-based feature extraction software to the other team members for use in building systems for the four RATS applications.

BBN and UMD have also had a very strong working relationship under the GALE program. Rich Schwartz from BBN, who is physically located in Virginia, has supervised the work of several students from UMD. Schwartz has been meeting with the students at the BBN office in Columbia, MD, on a weekly basis. Under RATS, Schwartz will continue his relationship with UMD and his interaction with the students there.

The above described working relationships that already exist among the PATROL team members will be a very good starting point for establishing a multilateral, open, and cooperative working relationship among all team members.

9.2 Team Size and Composition

From our experience in the EARS and GALE programs, we have found that, in programs with challenging goals, it is important to have the most qualified people at multiple sites work together, bounce off ideas against each other, and collaborate in implementing and integrating ideas and solutions that lead to the best performance and accuracy. By any measure, the goals of the RATS program are very challenging and, therefore, require diverse ideas from a team that has the requisite expertise and proven record of innovation and superior performance. We believe we have assembled such a team. Going with the old adage of "two heads are better than one", we have endeavored to ensure that, in each primary technical area, we have at least two sites that focus on each area of research.

Table 9-1 lists the five sites on the PATROL team and the areas of primary concentration in the proposed effort. In the area of robust feature extraction and speech enhancement, Dr. Hermansky at JHU and Dr. Shamma at UMD have each developed audio processing methods that tend to separate the speech from the noise. They will make their methods and software available to the PATROL team in order to see which features work best with large levels of noise.

Even with the use of robust features, there will still be the need for noise modeling. BBN and Cambridge University are both very experienced in this area and will continue to innovate solutions to this thorny problem. Brno University of Technology (BUT) will use their joint factor analysis (JFA) work to render the ID tasks more immune to noise.

	Robust Feature Extraction/ Enhancement	Noise Modeling	Speech Activity Detection	Language ID	Speaker ID	Keyword Spotting	Software Integration
BBN		X	X	X	X	X	X
Johns Hopkins U	X				X		
U Maryland	X		X				
Cambridge U		X				X	
Brno U Tech		X		X	X		

Table 9-1 Organizations of the PATROL team and their primary areas of concentration in RATS.

Dr. Shamma at UMD has obtained very good results in performing speech activity detection (SAD) in various types of noise by using his cortical representation of sound. He will continue to work on this problem while BBN will use advanced statistical methods to perform SAD.

BBN and BUT will work on language ID and speaker ID, both of which are the primary areas of strength at BUT. In addition, Dr. Hermansky will work on speaker ID, especially that he currently is working on that problem on an IARPA effort.

BBN is a recognized leader in the area of keyword spotting and has delivered systems to the Government for keyword spotting in various languages. Cambridge University – well known for their expertise in speech recognition – will work with BBN on keyword spotting.

Sites would be expected to deliver to BBN working software modules (a task that is not shown explicitly in the above table). BBN will need to integrate all this software into a working system that can be delivered to the Government. Naturally, BBN will be the only site performing integration and delivery.

Even though each site will concentrate on the tasks listed in Table 9-1, the sites, through joint meetings, will also be available to consult on the other tasks, as their expertise permits.

As one can see from the above, BBN has built a team with some built-in redundancy, but with maximum effectiveness. We strongly believe that this team, both in size and composition, is both necessary and sufficient to meet the program objectives.

9.3 Working/Meeting Models

BBN has established teaming relationships with each of the PATROL team members. The Statement of Work for each team member contains specific technical requirements by task, project management reporting responsibilities, anticipated support for meetings and presentations, and the delivery of software. The teaming strategy draws upon unique capabilities of each participant to form a comprehensive execution plan for the RATS effort.

BBN plans to have a face-to-face kickoff meeting early in the project. The kick-off meeting will concentrate on working out an overall plan to be accomplished during Phase 1 of the project, and a timeline for executing the plan. After the kickoff meeting, most of the day-to-day coordination of activities will take place through regular weekly teleconferences for each of the major tasks. The technical reviews for each task will include coordination of activities across sites, planning for the research to be performed, the assignment of specific tasks, and the review of progress against the work plan. In addition to the frequent teleconferences, periodic face-to-face meetings will be scheduled at various locations, taking advantage of major conferences and DARPA meetings. Such team meetings will strengthen the ties among team members and allow for in-depth technical discussions and planning. In addition to the DARPA requested meetings and workshops specified upon award of the contract, meetings and teleconferences with the DARPA RATS Program Manager will be scheduled as needed.

9.4 Software/Code Management

9.4.1 Software Integration/Management

BBN is highly experienced in the integration of technology from other sites into functioning systems, with appropriate GUIs, which have been delivered to the Government, either for internal use or for deployment in the field for 24/7 operations. Our expertise includes developing COTS Human Language Technology (HLT) software systems, such as the Broadcast Monitoring System (BMS), which integrates software from two other sites; it operates with minimal human oversight at over 18 Government locations. In addition to commercial products, BBN delivers state-of-the-art HLT software suites directly to Government customers for their use in research and operations. One such system is Byblos, BBN's state-of-the-art, *trainable* speech-to-text (STT) system. Byblos is used by Government staff to run STT and to create new STT models using Byblos tools to train on noisy, real-world telephone data in a variety of languages. BBN also delivers keyword search (KWS) systems to the Government, including word and phone-based systems that operate at very high speeds. The KWS software suites are designed for use in larger Government systems, so they are callable via an API jointly designed by the Government and BBN. BBN has extensive experience supporting the Government customer to help them make STT, KWS, and other HLT work in a wide variety of languages and domains. We have also delivered a GUI-driven *trainable* OCR system to the Government (the OCR system is based on the BBN Byblos STT system).

In the RATS project, the integration of the PATROL software system will be performed at BBN. Well before integration, BBN, in consultation with the team members, will create APIs that clearly define interfaces of the modules each site is developing. Early integration of prototype modules in Phase 1 will take place to test API soundness and ensure successful final integration for delivery to the Evaluation Team. Our teammates will contribute software in the form of source code, enabling BBN to create a single tightly integrated software suite. Having the source code in a centralized location will also allow us to rapidly debug and fix any issues discovered during integration. Whenever sites submit new versions of software modules, they will also submit test programs that exercise API functionality and the results of test runs demonstrating the module's conformance to the defined API.

All source code will be checked into a central repository at BBN dedicated to the RATS project. This repository will be used for integration of all subcontractor components into the complete system. In addition, this repository will support the usual revision control functions necessary for BBN staff to efficiently conduct research exploring alternative software algorithms. Specifically, BBN researchers will check in to the repository all changes to the research code base, typically checking into a private branch of the repository until a change has been shown to be beneficial experimentally, at which time the change is merged into the repository's "main trunk." The RATS project will adopt a number of software practices that BBN currently uses to maintain a consistent software repository. This includes creation of a symbolically tagged, date-stamped "local release" every time the trunk of the repository is updated. Software will also be tagged, built, and released whenever new software contributions are received from subcontractors. The symbolic tags ensure our ability to analyze and identify the source of any unexpected changes in system behavior. Other software practices we will adopt include the use of automated nightly builds and regression tests, with emails automatically sent when any error or warning occurs in the process, along with information about the specific module and originating site. We will then work with that site to fix any problems.

BBN's software system will also adhere to the APIs and input and output formatting conventions defined by the Evaluation Team for periodic evaluation deliveries and by military end users for the trainable system deliveries. BBN will test all software components extensively prior to delivery to ensure robustness and conformance to external APIs.

BBN will support the Evaluation Team by providing access to all developed system and component software; addressing software instrumentation needs; identifying failure causes; meeting required delivery dates to support evaluation events; and responding to any other test-related issues. BBN's extensive experience supporting Byblos and other Human Language Technology systems at remote customer sites gives us the confidence to ensure that the Evaluation Team can successfully run our RATS system.

9.4.2 Secure Processing

The work under Technical Area 1 of the RATS program does not require the processing of any classified data. However, the Evaluation Team is slated to test the delivered systems on classified data, possibly at the SCI level. In the event that the Evaluation Team should meet with any issues when running the PATROL system on classified data, BBN has numerous staff in its speech processing group who are cleared at the SCI level and who will be able to interact with the Evaluation Team or any future military end users to ensure that the software suite is correctly configured and runs successfully on their data. In addition, BBN has a TS/SCI SCIF equipped with a secure connection to TS/SCI Government networks and, therefore, will be able to process the same classified data in its SCIF, should the need arise to do that.

9.5 University Participation

In this project, we are fortunate to have as partners four universities that have excelled in their areas of expertise which are relevant to the RATS program, as mentioned above. The Key Personnel professors in those universities will be supervising a mix of research staff, including post-docs, and graduate students who will be participating as research assistants. The staff and students, under the guidance of their professors, will be writing software and performing experiments to improve accuracy in the four RATS applications in noise.

In addition to developing new research ideas and writing code, these universities will be expected to deliver to BBN software that will be integrated into the BBN system for delivery to the Evaluation Team and to potential users. Under an existing IARPA contract, JHU, UMD, and BUT will be delivering software for an external evaluation team as well. So, some of their software will have already undergone hardening for delivery to third parties. Furthermore, BUT has a lot of experience in transitioning their software for government and commercial applications. We plan to take advantage of all this experience to integrate the software from these sites with ours for delivery. The speech group at Cambridge University has a long history of building solid software (especially their HTK speech recognition toolkit) for external use.

For over 20 years, BBN has had a cooperative relationship with Northeastern University, where Dr. Makhoul has an appointment as Adjunct Professor. In this relationship, graduate students from Northeastern work as research assistants and do their graduate thesis work at BBN, supervised by senior BBN staff, with Dr. Makhoul acting as their official thesis advisor. We expect to continue this arrangement, so we have budgeted one research assistant from Northeastern to work on this project.

9.6 Government's Role

We expect that the Government will provide us with the necessary data with which to build our models for the four application areas. We will work closely with the Program Manager to ensure that our work is well aligned with the Government's expectations.

9.7 Project Management at BBN

BBN has an extensive history of program management success that is predicated on tailoring the management approach to the needs of the project. Our project activities range from very large, formal programs to small, short-term projects with only a few engineers. Our project activities range from research and development to technical transition to commercial offerings. We have the flexibility and trained talent within BBN to work at all these levels. BBN is dedicated to the principle that program management must take a balanced and comprehensive approach to the achievement of technical, cost, and schedule objectives. To meet these objectives, BBN has developed a program organization and management approach proven effective by prior successful contract performances with multiple team members.

Every project has a designated program manager who is responsible for the overall management leading to the successful completion within time and budget. The program manager is responsible for day-to-day management, staffing plans, financial tracking and reporting, subcontractor relationships, and tracking of milestones and deliverables. Every project is required to have a project plan detailing that project's tasks, work organization, and budgeted costs. The plan, schedule, assignments and budgets are communicated to all program team members. Monthly tracking and control procedures measure the progress and cost against the plan. Each project is subjected to monthly reviews that report on project status, so that any difficulties or risks that may impact project completion can be identified and addressed in a timely manner.

The team holds regular project meetings to coordinate work, identify any potential problems, and ensure the project is proceeding on schedule. Program risks will be identified and discussed, and mitigation plans implemented. In a program of the magnitude of RATS the team meetings are required at both the program and task level.

A project web site is established to store, retrieve, and share technical and management documents. All work products undergo an appropriate level of peer review. We provide status reports prepared and submitted in accordance with procedures in the award document. At the end of each phase, we prepare and submit a Final Report that summarizes the project and tasks at the end of that phase.

BBN has the infrastructure to support the management of large government contracts. A number of systems (such as BBN's time reporting system, Microsoft Project for schedule creation and management, and PRISM for project finances) are in place to monitor project status and facilitate project control. Most importantly, we have customizable project management practices. We understand and appreciate the need to be flexible, efficient, and highly responsive to the evolving goals of the RATS Program.

10 Cost Summaries

The following four pages provide a complete cost summary for the proposed PATROL effort under the RATS program.

The first page contains a top-level view of the proposed costs, broken down by phase, major task, and performing site. There are seven major technical tasks: Feature Extraction, Noise Modeling/Compensation, Speech Activity Detection, Language Identification, Speaker Identification, Keyword Spotting, and System Integration. In addition to the seven major technical tasks, we have also priced Program Management and Program Infrastructure as additional major tasks. Program Management includes program administration, all BBN travel, and data services (such as low-level annotations, which will be made available to the other RATS performing sites). Program Infrastructure includes materials (computer equipment), computer system administration, and data acquisition. For Phases 2 and 3, we also show at the bottom of the page the costing for the proposed Speech Transcription option.

The second page contains the same costing information for Phase 1, but broken down by month, for the 18-month duration. The third and fourth pages contain the corresponding monthly costing information for Phases 2 and 3, each lasting 12 months.

RATS PATROL SUMMARY

Base Proposal	Phase 1	Phase 2	Phase 3
Feature Extraction			
Johns Hopkins University	$ 191,501	$ 128,763	$ 131,141
University of Maryland	$ 129,261	$ 95,315	$ 94,500
Noise Modeling/Compensation			
BBN	$ 664,876	$ 442,559	$ 441,829
Brno University	$ 51,520	$ 38,011	$ 37,687
Cambridge University	$ 134,649	$ 54,900	$ 55,978
Speech Activity Detection (SAD)			
BBN	$ 475,039	$ 276,875	$ 243,125
University of Maryland	$ 129,261	$ 95,315	$ 94,500
Language ID (LID)			
BBN	$ 644,170	$ 461,689	$ 495,500
Brno University	$ 103,040	$ 76,036	$ 75,372
Speaker ID (SID)			
BBN	$ 486,644	$ 398,574	$ 419,302
Brno University	$ 103,040	$ 76,036	$ 75,372
Johns Hopkins University	$ 191,501	$ 128,763	$ 131,140
Keyword Spotting (KWS)			
BBN	$ 616,423	$ 461,364	$ 483,307
Cambridge University	$ 134,650	$ 54,904	$ 55,978
Northeastern University	$ 43,063	$ 44,782	$ 46,573
System Integration			
BBN	$ 313,267	$ 342,757	$ 377,522
Program Management *			
BBN	$ 328,681	$ 247,608	$ 234,355
Program Infrastructure **			
BBN	$ 552,786	$ 470,811	$ 400,930
Total Base	**$5,293,372**	**$ 3,895,062**	**$3,894,111**
Speech Transcription Option			
BBN		$ 325,701	$ 337,100
Cambridge University		$ 157,084	$ 162,459
Total Option		**$ 482,785**	**$ 499,559**

* Program Management includes program administration, travel and data services.

** Program Infrastructure includes system administration, material, and data acquisition.

Phase 1 by Month

Phase 1 Base	Aug-10	Sep-10	Oct-10	Nov-10	Dec-10	Jan-11	Feb-11	Mar-11	Apr-11	May-11	Jun-11	Jul-11	Aug-11	Sep-11	Oct-11	Nov-11	Dec-11	Jan-12	Feb-12	Total
Feature Extraction																				
Johns Hopkins University	$ 5,341	$ 10,682	$ 10,682	$ 10,682	$ 10,682	$ 10,682	$ 10,682	$ 10,682	$ 10,682	$ 10,682	$ 10,682	$ 10,682	$ 10,682	$ 10,682	$ 10,682	$ 10,581	$ 10,413	$ 10,412	$ 5,206	**$ 191,501**
University of Maryland	$ 3,588	$ 7,182	$ 7,182	$ 7,182	$ 7,182	$ 7,180	$ 7,182	$ 7,182	$ 7,182	$ 7,181	$ 7,181	$ 7,182	$ 7,181	$ 7,181	$ 7,181	$ 7,181	$ 7,182	$ 7,182	$ 3,587	**$ 129,261**
Noise Modeling/Compensation																				
BBN	$ 14,523	$ 30,623	$ 30,707	$ 30,707	$ 30,159	$ 39,235	$ 38,122	$ 38,427	$ 38,391	$ 36,348	$ 38,202	$ 37,214	$ 38,549	$ 37,101	$ 36,202	$ 38,146	$ 35,798	$ 40,396	$ 35,026	**$ 664,876**
Brno University							$ 4,294	$ 4,294	$ 4,294	$ 4,294	$ 4,294	$ 4,294	$ 4,294	$ 4,294	$ 4,294	$ 4,294	$ 4,290	$ 4,290		**$ 51,520**
Cambridge University	$ 7,087	$ 7,087	$ 7,087	$ 7,087	$ 7,087	$ 7,087	$ 7,087	$ 7,087	$ 7,087	$ 7,087	$ 7,086	$ 7,086	$ 7,086	$ 7,086	$ 7,087	$ 7,087	$ 7,087	$ 7,087	$ 7,087	**$ 134,649**
Speech Activity Detection (SAD)																				
BBN	$ 11,264	$ 19,976	$ 20,855	$ 21,532	$ 21,532	$ 30,169	$ 29,800	$ 29,356	$ 29,026	$ 28,644	$ 28,666	$ 26,743	$ 24,889	$ 26,371	$ 25,860	$ 22,616	$ 24,335	$ 25,303	$ 28,101	**$ 475,038**
University of Maryland	$ 3,588	$ 7,182	$ 7,182	$ 7,182	$ 7,182	$ 7,180	$ 7,182	$ 7,182	$ 7,182	$ 7,181	$ 7,181	$ 7,182	$ 7,181	$ 7,181	$ 7,181	$ 7,181	$ 7,182	$ 7,182	$ 3,587	**$ 129,261**
Language ID (LID)																				
BBN	$ 15,246	$ 29,236	$ 29,057	$ 29,441	$ 29,479	$ 40,152	$ 38,798	$ 38,392	$ 36,002	$ 37,709	$ 35,452	$ 35,087	$ 35,002	$ 37,956	$ 37,338	$ 34,463	$ 33,061	$ 37,269	$ 34,030	**$ 644,170**
Brno University							$ 8,587	$ 8,587	$ 8,587	$ 8,587	$ 8,587	$ 8,587	$ 8,587	$ 8,587	$ 8,587	$ 8,587	$ 8,587	$ 8,583	$ -	**$ 103,040**
Speaker ID (SID)																				
BBN	$ 13,565	$ 23,807	$ 24,100	$ 24,393	$ 24,535	$ 27,232	$ 26,620	$ 26,229	$ 28,679	$ 27,344	$ 26,574	$ 26,883	$ 27,838	$ 28,073	$ 27,379	$ 25,266	$ 24,187	$ 27,650	$ 26,290	**$ 486,644**
Brno University							$ 8,587	$ 8,587	$ 8,587	$ 8,587	$ 8,587	$ 8,587	$ 8,587	$ 8,587	$ 8,587	$ 8,587	$ 8,587	$ 8,583	$ -	**$ 103,040**
Johns Hopkins University	$ 5,341	$ 10,682	$ 10,682	$ 10,682	$ 10,682	$ 10,682	$ 10,682	$ 10,682	$ 10,682	$ 10,682	$ 10,682	$ 10,682	$ 10,682	$ 10,682	$ 10,683	$ 10,581	$ 10,412	$ 10,412	$ 5,206	**$ 191,501**
Keyword Spotting (KWS)																				
BBN	$ 12,112	$ 29,376	$ 30,852	$ 30,252	$ 30,218	$ 34,581	$ 35,157	$ 33,786	$ 36,233	$ 35,536	$ 35,132	$ 35,497	$ 35,175	$ 34,924	$ 34,242	$ 34,154	$ 33,756	$ 34,460	$ 30,980	**$ 616,423**
Cambridge University	$ 7,087	$ 7,087	$ 7,087	$ 7,087	$ 7,088	$ 7,087	$ 7,087	$ 7,087	$ 7,087	$ 7,087	$ 7,087	$ 7,087	$ 7,087	$ 7,083	$ 7,087	$ 7,087	$ 7,087	$ 7,087	$ 7,087	**$ 134,650**
Northeastern University							$ 1,826	$ 1,826	$ 1,826	$ 1,826	$ 8,876	$ 8,876	$ 8,876	$ 1,826	$ 1,826	$ 1,826	$ 1,827	$ 1,826		**$ 43,063**
System Integration																				
BBN	$ 6,464	$ 13,777	$ 13,777	$ 13,777	$ 13,777	$ 20,329	$ 19,041	$ 18,367	$ 19,392	$ 17,871	$ 17,685	$ 17,359	$ 16,900	$ 18,291	$ 17,508	$ 17,157	$ 16,355	$ 18,484	$ 16,956	**$ 313,267**
Program Management																				
BBN	$ 19,031	$ 24,678	$ 12,546	$ 12,375	$ 12,071	$ 13,200	$ 12,930	$ 12,071	$ 17,290	$ 27,703	$ 25,211	$ 29,330	$ 15,766	$ 29,912	$ 12,020	$ 12,733	$ 16,993	$ 11,901	$ 10,920	**$ 328,681**
Program Infrastructure																				
BBN	$ 11,975	$ 12,161	$ 11,975	$228,757	$ 12,161	$ 11,975	$ 12,160	$ 11,976	$115,174	$ 12,562	$ 12,370	$ 12,371	$ 12,562	$ 12,370	$ 12,563	$ 12,370	$ 12,563	$ 12,370	$ 12,372	**$ 552,787**
Total Phase 1 Base	**$136,212**	**$233,536**	**$223,771**	**$441,136**	**$223,835**	**$266,771**	**$285,824**	**$281,800**	**$393,383**	**$296,911**	**$299,536**	**$300,729**	**$287,924**	**$298,187**	**$276,307**	**$269,897**	**$268,702**	**$280,477**	**$227,435**	**$5,293,372**

Phase 2 by Month

Phase 2 Base	Feb-12	Mar-12	Apr-12	May-12	Jun-12	Jul-12	Aug-12	Sep-12	Oct-12	Nov-12	Dec-12	Jan-13	Feb-13	Total
Feature Extraction														
Johns Hopkins University	$ 9,905	$ 9,905	$ 9,905	$ 9,905	$ 9,905	$ 9,905	$ 9,905	$ 9,905	$ 9,905	$ 9,905	$ 9,905	$ 9,905	$ 9,903	**$ 128,763**
University of Maryland	$ 4,039	$ 8,079	$ 8,079	$ 8,079	$ 7,977	$ 7,875	$ 7,875	$ 7,874	$ 7,875	$ 7,875	$ 7,875	$ 7,875	$ 3,938	**$ 95,315**
Noise Modeling/Compensation														
BBN	$ 34,140	$ 33,039	$ 34,196	$ 34,256	$ 35,020	$ 35,589	$ 35,832	$ 35,589	$ 34,196	$ 34,256	$ 34,196	$ 33,057	$ 29,192	**$ 442,558**
Brno University	$ 3,222	$ 3,222	$ 3,222	$ 3,222	$ 3,140	$ 3,141	$ 3,141	$ 3,141	$ 3,140	$ 3,140	$ 3,140	$ 3,140		**$ 38,011**
Cambridge University	$ 2,317	$ 4,635	$ 4,635	$ 4,635	$ 4,635	$ 4,635	$ 4,554	$ 4,519	$ 4,519	$ 4,519	$ 4,519	$ 4,519	$ 2,259	**$ 54,900**
Speech Activity														
BBN	$ 20,859	$ 21,188	$ 21,244	$ 22,243	$ 22,099	$ 21,716	$ 22,595	$ 22,128	$ 21,814	$ 21,959	$ 20,959	$ 20,124	$ 17,947	**$ 276,875**
University of Maryland	$ 4,039	$ 8,079	$ 8,079	$ 8,079	$ 7,976	$ 7,875	$ 7,875	$ 7,875	$ 7,875	$ 7,875	$ 7,875	$ 7,875	$ 3,938	**$ 95,315**
Language ID (LID)														
BBN	$ 33,190	$ 33,827	$ 34,352	$ 34,823	$ 36,090	$ 36,315	$ 37,156	$ 36,315	$ 35,836	$ 36,217	$ 35,491	$ 37,761	$ 34,316	**$ 461,689**
Brno University	$ 6,444	$ 6,444	$ 6,444	$ 6,444	$ 6,293	$ 6,281	$ 6,281	$ 6,281	$ 6,281	$ 6,281	$ 6,281	$ 6,281		**$ 76,036**
Speaker ID (SID)														
BBN	$ 29,956	$ 30,559	$ 30,720	$ 31,628	$ 31,289	$ 32,012	$ 31,603	$ 31,599	$ 31,289	$ 31,216	$ 30,720	$ 28,302	$ 27,679	**$ 398,572**
Brno University	$ 6,444	$ 6,444	$ 6,444	$ 6,444	$ 6,295	$ 6,281	$ 6,281	$ 6,281	$ 6,281	$ 6,281	$ 6,281	$ 6,281		**$ 76,038**
Johns Hopkins University	$ 9,905	$ 9,903	$ 9,905	$ 9,905	$ 9,905	$ 9,905	$ 9,905	$ 9,905	$ 9,905	$ 9,905	$ 9,905	$ 9,905	$ 9,905	**$ 128,763**
Keyword Spotting														
BBN	$ 34,624	$ 35,723	$ 36,315	$ 36,689	$ 36,660	$ 36,660	$ 35,478	$ 36,248	$ 35,836	$ 36,149	$ 35,492	$ 33,107	$ 32,383	**$ 461,364**
Cambridge University	$ 2,318	$ 4,635	$ 4,635	$ 4,636	$ 4,637	$ 4,637	$ 4,556	$ 4,519	$ 4,518	$ 4,518	$ 4,518	$ 4,518	$ 2,259	**$ 54,904**
Northeastern University	$ 1,899	$ 1,899	$ 1,899	$ 1,900	$ 9,229	$ 9,229	$ 9,230	$ 1,900	$ 1,899	$ 1,900	$ 1,899	$ 1,899		**$ 44,782**
System Integration														
BBN	$ 23,176	$ 22,803	$ 24,400	$ 23,703	$ 24,298	$ 28,507	$ 30,360	$ 28,507	$ 28,689	$ 29,251	$ 29,948	$ 24,364	$ 24,751	**$ 342,757**
Program Management														
BBN	$ 21,263	$ 28,165	$ 19,079	$ 14,460	$ 26,018	$ 30,577	$ 16,978	$ 20,653	$ 14,336	$ 14,048	$ 18,179	$ 11,663	$ 12,189	**$ 247,608**
Program Infrastructure														
BBN	$228,892	$ 12,110	$ 12,500	$ 12,499	$ 12,499	$ 12,499	$ 12,302	$105,017	$ 12,499	$ 12,497	$ 12,499	$ 12,500	$ 12,499	**$ 470,812**
Total Phase 2 Base	**$476,632**	**$280,659**	**$276,053**	**$273,550**	**$293,965**	**$303,639**	**$291,907**	**$378,256**	**$276,693**	**$277,792**	**$279,682**	**$263,076**	**$223,158**	**$3,895,062**
Speech Transcription Option														
BBN	$ 24,153	$ 24,698	$ 24,492	$ 25,847	$ 24,998	$ 25,562	$ 24,492	$ 25,562	$ 24,998	$ 25,847	$ 24,492	$ 25,562	$ 24,998	**$ 325,701**
Cambridge University	$ 12,084	$ 12,083	$ 12,083	$ 12,083	$ 12,084	$ 12,084	$ 12,083	$ 12,084	$ 12,084	$ 12,083	$ 12,083	$ 12,083	$ 12,083	**$ 157,084**
Total Option	$ 36,237	$ 36,781	$ 36,575	$ 37,930	$ 37,082	$ 37,646	$ 36,575	$ 37,646	$ 37,082	$ 37,930	$ 36,575	$ 37,645	$ 37,081	$ 482,785

Phase 3 by Month

Phase 3 Base

	Feb-13	Mar-13	Apr-13	May-13	Jun-13	Jul-13	Aug-13	Sep-13	Oct-13	Nov-13	Dec-13	Jan-14	Feb-14	Total
Feature Extraction														
Johns Hopkins University	$ 10,087	$ 10,088	$ 10,087	$ 10,088	$ 10,088	$ 10,088	$ 10,088	$ 10,088	$ 10,088	$ 10,088	$ 10,088	$ 10,088	$ 10,087	**$ 131,141**
University of Maryland	$ 7,269	$ 7,269	$ 7,269	$ 7,269	$ 7,269	$ 7,269	$ 7,270	$ 7,270	$ 7,270	$ 7,269	$ 7,269	$ 7,269	$ 7,269	**$ 94,500**
Noise Modeling/Compensation														
BBN	$ 33,219	$ 33,050	$ 36,019	$ 35,945	$ 37,967	$ 37,465	$ 38,067	$ 36,339	$ 35,235	$ 35,030	$ 34,382	$ 30,318	$ 18,793	**$ 441,829**
Brno University	$ 3,141	$ 3,141	$ 3,141	$ 3,141	$ 3,140	$ 3,140	$ 3,140	$ 3,140	$ 3,140	$ 3,141	$ 3,141	$ 3,141		**$ 37,687**
Cambridge University	$ 2,332	$ 4,665	$ 4,665	$ 4,665	$ 4,665	$ 4,665	$ 4,665	$ 4,665	$ 4,665	$ 4,665	$ 4,665	$ 4,664	$ 2,332	**$ 55,978**
Speech Activity Detection (SAD)														
BBN	$ 18,525	$ 17,938	$ 19,823	$ 19,347	$ 20,058	$ 21,630	$ 20,233	$ 20,980	$ 19,146	$ 18,697	$ 19,174	$ 16,962	$ 10,612	**$ 243,125**
University of Maryland	$ 7,269	$ 7,269	$ 7,269	$ 7,269	$ 7,269	$ 7,269	$ 7,269	$ 7,270	$ 7,270	$ 7,270	$ 7,269	$ 7,269	$ 7,269	**$ 94,500**
Language ID (LID)														
BBN	$ 37,082	$ 39,102	$ 40,033	$ 42,064	$ 40,750	$ 41,854	$ 38,053	$ 40,148	$ 40,750	$ 40,116	$ 38,085	$ 34,602	$ 22,861	**$ 495,500**
Brno University	$ 6,281	$ 6,281	$ 6,281	$ 6,281	$ 6,281	$ 6,281	$ 6,281	$ 6,281	$ 6,281	$ 6,281	$ 6,281	$ 6,281		**$ 75,372**
Speaker ID (SID)														
BBN	$ 31,519	$ 31,006	$ 32,623	$ 32,222	$ 32,196	$ 32,517	$ 32,947	$ 32,517	$ 32,196	$ 32,222	$ 32,623	$ 32,091	$ 32,623	**$ 419,302**
Brno University	$ 6,281	$ 6,281	$ 6,281	$ 6,281	$ 6,281	$ 6,281	$ 6,281	$ 6,281	$ 6,281	$ 6,281	$ 6,281	$ 6,281		**$ 75,372**
Johns Hopkins University	$ 10,087	$ 10,088	$ 10,088	$ 10,088	$ 10,087	$ 10,088	$ 10,088	$ 10,088	$ 10,088	$ 10,088	$ 10,088	$ 10,087	$ 10,087	**$ 131,140**
Keyword Spotting (KWS)														
BBN	$ 35,766	$ 36,323	$ 37,506	$ 37,594	$ 37,018	$ 37,269	$ 37,831	$ 37,269	$ 37,018	$ 37,594	$ 37,506	$ 37,594	$ 37,018	**$ 483,306**
Cambridge University	$ 2,332	$ 4,665	$ 4,665	$ 4,665	$ 4,665	$ 4,665	$ 4,665	$ 4,665	$ 4,665	$ 4,664	$ 4,665	$ 4,665	$ 2,332	**$ 55,978**
Northeastern University	$ 1,976	$ 1,976	$ 1,976	$ 1,976	$ 9,599	$ 9,599	$ 9,599	$ 1,976	$ 1,974	$ 1,974	$ 1,974	$ 1,974		**$ 46,573**
System Integration														
BBN	$ 28,420	$ 27,877	$ 29,415	$ 29,091	$ 29,177	$ 29,090	$ 29,415	$ 29,090	$ 29,175	$ 29,090	$ 29,414	$ 28,853	$ 29,415	**$ 377,522**
Program Management														
BBN	$ 21,095	$ 22,568	$ 18,482	$ 13,860	$ 13,305	$ 29,977	$ 16,382	$ 25,949	$ 13,305	$ 13,860	$ 18,406	$ 13,097	$ 14,069	**$ 234,355**
Program Infrastructure														
BBN	$208,835	$ 12,499	$ 12,903	$ 12,901	$ 12,901	$ 12,901	$ 12,697	$ 50,784	$ 12,901	$ 12,901	$ 12,903	$ 12,903	$ 12,902	**$ 400,931**
Total Phase 3 Base	$471,516	$282,086	$288,526	$284,747	$292,716	$312,048	$294,971	$334,800	$281,448	$281,231	$284,214	$268,139	$217,669	$3,894,111
Speech Transcription Option														
BBN	$ 24,998	$ 25,562	$ 25,349	$ 26,752	$ 25,873	$ 26,457	$ 25,349	$ 26,457	$ 25,873	$ 26,752	$ 25,348	$ 26,457	$ 25,873	$ 337,100
Cambridge University	$ 12,497	$ 12,497	$ 12,496	$ 12,497	$ 12,497	$ 12,497	$ 12,496	$ 12,497	$ 12,497	$ 12,497	$ 12,497	$ 12,497	$ 12,497	$ 162,459
Total Option	$ 37,495	$ 38,059	$ 37,845	$ 39,249	$ 38,370	$ 38,954	$ 37,845	$ 38,954	$ 38,370	$ 39,249	$ 37,845	$ 38,954	$ 38,370	$ 499,559

11 Organizational Conflict of Interest Affirmations and Disclosure

None.

12 Human Use

None.

13 Animal Use

None.

14 Statement of Unique Capability Provided by Government or Government-funded Team Member

Not Applicable.

15 Government or Government-funded Team Member Eligibility

Not Applicable.

Appendix A: References

[1] P. Mermelstein, "Distance measures for speech recognition, psychological and instrumental," in *Pattern Recognition and Artificial Intelligence*, C. H. Chen (Ed.), pp. 374–388, Academic, New York, 1976.

[2] H. Hermansky, "Perceptual Linear Prediction (PLP) Analysis for Speech," *J. Acoust. Soc. Amer.*, Vol. 87, pp. 1738-1752, April 1990.

[3] H. Hermansky and N. Morgan, "RASTA Processing of Speech," *IEEE Transactions on Speech and Audio Processing*, Vol. 2, No. 4, pp. 587-589, October, 1994.

[4] M. Berouti, R. Schwartz, and J. Makhoul, "Enhancement of speech corrupted by acoustic noise," *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 208-211, April, 1979.

[5] S. Ganapathy, S. Thomas, and H. Hermansky, "Robust Spectro-temporal Features based on Autoregressive Models of Hilbert Envelopes," *ICASSP*, Dallas, TX, March 2010.

[6] H. Hermansky, "The Modulation Spectrum in Automatic Recognition of Speech," *IEEE Workshop on Automatic Speech Recognition and Understanding*, S. Furui, B.-H. Juang and W. Chou (Eds.), 1997.

[7] T. Chi, P. Ru, and S. Shamma, "Multiresolution spectrotemporal analysis of complex sounds," *J. Acoust. Soc. Amer.*, Vol. 118, pp. 887-906, 2006.

[8] M. Mirbagheri, N. Mesgarani, and S. Shamma "Nonlinear filtering of spectrotemporal modulations in speech enhancement," *ICASSP*, Dallas, TX, March 2010.

[9] B. Zhang, S. Matsoukas, and R. Schwartz, "Discriminatively Trained Region Dependent Feature Transforms for Speech Recognition," *ICASSP*, Vol. 1, pp.I-313 - I-316, May 2006.

[10] H. Hermansky, D.P.W. Ellis, and S. Sharma, "Connectionist Feature Extraction for Conventional HMM Systems," *ICASSP*, 2000.

[11] H. Hermansky and S. Sharma, "Temporal Patterns (TRAPS) in ASR of Noisy Speech," *IEEE International Conference on Acoustics, Speech and Signal Processing*, Phoenix, AZ, 1999.

[12] F. Valente and H. Hermansky, "Hierarchical and Parallel Processing of Modulation Spectrum for ASR applications," *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2008.

[13] J. Pinto, Y. Bayya, H. Hermansky, and M. Magimai-Doss, "Exploiting Contextual Information For Improved Phoneme Recognition," *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, 2008.

[14] B. Zhang, S. Matsoukas, and R. Schwartz, "Recent progress on the discriminative region-dependent transform for speech feature extraction," *Interspeech*, Pittsburgh, PA, September 2006.

[15] J. Ma and R. Schwartz, "Unsupervised versus supervised training of acoustic models," *Interspeech*, 2008.

[16] S. Novotney, R. Schwartz, and J. Ma, "Unsupervised acoustic and language model training with small amounts of labeled data," *ICASSP*, 2009.

[17] S. Novotney, and R. Schwartz, "Analysis of Low-Resource Acoustic Model Self-Training," *ICASSP*, 2009.

[18] H. Gish, M. Siu, A. Chan, and W. Belfield, "Unsupervised training of an HMM-based speech recognizer for topic classification," *ICASSP*, 2009.

[19] N. Mesgarani, M. Slaney, S. A. Shamma, "Discrimination of speech from nonspeech based on multiscale spectro-temporal modulations," *IEEE Transactions on Audio, Speech, and Language Processing*, vol.14, no.3, pp. 920- 930, May 2006.

[20] L. Deng, A. Acero, M. Plumpe, and X.D. Huang, "Large vocabulary speech recognition under adverse acoustic environments," *ICSLP*, Vol.3, pp. 806-809, October, 2000.

[21] M.J.F. Gales and S.J. Young, "An Improved Approach to the Hidden Markov Model Decomposition of Speech and Noise," *ICASSP*, Vol. 1, pp. 233-236, 1992.

[22] M.J.F. Gales, S.J. Young, "A Fast and Flexible Implementation of Parallel Model Combination," *ICASSP*, pp. 133-136, 1995.

[23] A. Acero, L. Deng, T. Kristjansson, and J. Zhang, "HMM adaptation using vector Taylor series for noisy speech recognition," *ICSLP*, September 2000.

[24] F. Flego and M.J.F., Gales, "Discriminative Adaptive Training with VTS and JUD," *IEEE Automatic Speech Recognition and Understanding Workshop*, December 2009.

[25] H. Xu, M.J.F., Gales, and K.K, Chin, "Improving Joint Uncertainty Decoding Performance by Predictive Methods for Noise Robust Speech Recognition," *IEEE Automatic Speech Recognition and Understanding Workshop*, December 2009.

[26] D.K, Kim and M.J.F., Gales, "Noisy Constrained Maximum Likelihood Linear Regression for Noise-Robust Speech Recognition," submitted to *IEEE Transactions on Audio Speech and Language Processing*, 2010.

[27] Z. Ghahramani and M. I. Jordan, "Factorial hidden Markov models," *Machine Learning*, 29:245–275, 1997.

[28] P. Kenny, G. Boulianne, P. Ouellet, and P. Dumouchel, "Joint factor analysis versus eigenchannels in speaker recognition," *IEEE Trans. Audio, Speech, Lang. Processing*, Vol. 15, no. 4, May 2007.

[29] G. Shafer, *A Mathematical Theory of Evidence*, MIT Press, 1976.

[30] F. Valente and H. Hermansky, "Combination of Acoustic Classifiers based on Dempster-Shafer Theory of Evidence," *IEEE International Conference on Acoustics, Speech and Signal Processing*, Honolulu, HI, 2007.

[31] R. Schwartz, A. Derr, A. Wilgus, and J. Makhoul, "Speaker Verification IR&D Final Report," *BBN Technical Report* No. 6661, Nov. 1987.

[32] E. Singer and D.A. Reynolds, "Analysis of multitarget detection for speaker and language recognition," *IEEE Odyssey: The Speaker and Language Recognition Workshop*, pp. 301-308, 2004.

[33] H. Jin, R. Schwartz, S. Sista, and F. Walls, "Topic Tracking for Radio, TV Broadcast, and Newswire," *Eurospeech*, Budapest, Hungary, pp. 2439-2442, September 1999.

[34] H. Jin, R. Schwartz, F. Walls, and S. Sista, "Method and apparatus for score normalization for information retrieval applications," *U.S. Patents* 7,062,485 and 6,651,057.

[35] D.R.H. Miller, M. Kleber, C. Kao, O. Kimball, T. Colthurst, S.A. Lowe, R.M. Schwartz, and H. Gish, "Rapid and Accurate Spoken Term Detection," *Interspeech*, Antwerp, Belgium, 2007.

[36] S. Matsoukas, L. Nguyen, J. Davenport, J. Billa, F. Richardson, M. Siu, D. Liu, R. Schwartz, and J. Makhoul, "The 1998 BBN Byblos Primary System Applied to English and Spanish Broadcast News Transcription," *DARPA Broadcast News Workshop*, Herndon, VA, Morgan Kaufmann Publishers, pp. 255-260, Feb. 28-Mar. 3, 1999.

[37] S. Galligan, "Automatic speech detection and segmentation of air traffic control audio using parametric trajectory models," *IEEE Aerospace Conference*, Big Sky, Montana, March 2007.

[38] A. Stolcke, L. Ferrer, S. Kajarekar, E. Shriberg, and A. Venkataraman, "MLLR Transforms as Features in Speaker Recognition," *Eurospeech*, pp. 2425-2428, September 2005.

[39] E. E. Shriberg "Higher level features in speaker recognition," in *Speaker Classification I*, C. Muller (Ed.), Vol. 4343 of *Lecture Notes in Computer Science*, Artificial Intelligence, Springer, pp. 241-259, 2007.

[40] C.J. Leggetter and P.C. Woodland, "Maximum likelihood linear regression for speaker adaptation of continuous density HMMs," *Computer Speech and Language*, Vol. 9, pp. 171-186, 1995.

[41] M.A. Zissman, "Comparison of four approaches to automatic language identification of telephone speech," *IEEE Trans. on Speech and Audio Processing*, 4(1):31-44, 1996.

[42] L. Burget, P. Matějka, and J. Černocý, "Discriminative Training Techniques for Acoustic Language Identification," *ICASSP*, Toulouse, France, pp. 209-212, 2006.

[43] T. Anastasakos, J. McDonough, R. Schwartz, and J. Makhoul, "A compact model for speaker-adaptive training," *Int. Conf. Speech Language Processing*, vol. 2, Philadelphia, PA, pp. 1137–1140, 1996.

[44] H. Li and B. Ma, "A Phonotactic Language Model for Spoken Language Identification," *43rd Annual Meeting of the ACL*, Ann Arbor, MI, pp. 515–522, June 2005.

[45] S. Sista, R. Schwartz, T. Leek, and J. Makhoul, "An Algorithm for Unsupervised Topic Discovery from Broadcast News Stories," *Human Language Technology Workshop*, San Diego, CA, 2002.

[46] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE ASSP-27*, pp. 113-120, April 1979.

[47] H. Gish, Y.-L. Chow, and J. R. Rohlicek, "Probabilistic Vector Mapping of Noisy Speech parameters for HMM word spotting," *ICASSP*, 1990.

[48] K. Ng, H. Gish, and J. R. Rohlicek, "Robust Mapping of noisy speech parameters for HMM word spotting," *ICASSP*, 1992.

[49] X. Zhu, "Multi-Stage speaker diarization for conference and lecture meetings," http://www.itl.nist.gov/iad/mig/tests/rt/2007/workshop/RT07-LIMSI-SPKR-Presentation_v2.pdf

[50] C. Wooters and M. Huijbregts, "The ICSI 2007 speaker diarization system," http://www.itl.nist.gov/iad/mig/tests/rt/2007/workshop/RT07-SRIICSI-SPKR-Presentation.pdf

[51] X. Anguera, M. Aguilo, C. Wooters, C. Nadeu, and J. Hernando, "Hybrid Speech/non-speech detector applied to Speaker Diarization of Meetings," *IEEE Odyssey Speaker and Language Recognition Workshop*, pp. 1-6, 2006.

[52] E. Scheirer and M. Slaney, "Construction and evaluation of a robust multifeature speech/music discriminator," *ICASSP*, 1997.

[53] B. Kingsbury, G. Saon, L. Mangu, M Padmanabhan and R. Sarikaya, "Robust speech recognition in noisy environments: The 2001 IBM SPINE evaluation system," *ICASSP*, 2002.

[54] D. A. Reynolds, R. C. Rose, and M. J. T. Smith, "A Mixture Modeling Approach to Text-Independent Speaker Identification," *Journal of the Acoustical Society of America*, Suppl. 1, Vol. 87, 1990.

[55] L. Burget, P. Matějka, P. Schwarz, O. Glembek, and J. Černocký, "Analysis of feature extraction and channel compensation in GMM speaker recognition system," *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 15, No. 7, 2007.

[56] W. Campbell, D. E. Sturim, D. A. Reynolds, "Support Vector Machines Using GMM Supervectors for Speaker Verification," *IEEE Signal Processing Letters*, Vol. 13, No. 5, pp. 308–311, May 2006.

[57] L. Burget, P. Matějka, V. Hubeika, and J. Černocký, "Investigation into variants of Joint Factor Analysis for speaker recognition," *Interspeech*, Brighton, UK, pp. 1263-1266, 2009.

[58] A. Solomonoff, W. M. Campbell, and C. Quillen, "Nuisance Attribute Projection," *Speech Communication*, Amsterdam, The Netherlands, May 2007.

[59] W. M. Campbell, J. P. Campbell, D. A. Reynolds, D. A. Jones, and T. R. Leek, "High-Level Speaker Verification with Support Vector Machines," *ICASSP*, Montreal, Quebec, Canada, pp. I: 73-76, May 2004.

[60] J. P. Campbell, D. A. Reynolds, and R. B. Dunn, "Fusing High- and Low-Level Features for Speaker Recognition," *Eurospeech*, 2003.

[61] N. Brümmer, L. Burget, J. Černocký, O. Glembek, F. Grézl, M. Karafiát, D. van Leeuwen, P. Matějka, P. Schwarz, and A. Strasheim, "Fusion of heterogeneous speaker recognition systems in the STBU submission for the NIST speaker recognition evaluation 2006," *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 15, No. 7, pp. 2072-2084, 2007.

[62] P. Matějka, P. Schwarz, L. Burget, and J. Černocký, "Use of anti-models to further improve state-of-the-art PRLM language recognition system," *ICASSP*, Toulouse, France, pp. 197-200, 2006.

[63] J.R. Rohlicek, P. Jeanrenaud, K. Ng, H. Gish, B. Musicus, M. Siu, "Phonetic training and language modeling for word spotting," *ICASSP*, 1993.

[64] J.G. Fiscus, J. Ajot, J.S. Garofolo, and G. Doddington, "Results of the 2006 Spoken Term Detection Evaluation," *Special Interest Group on Information Retrieval* (SIGIR-07) *Workshop in Searching Spontaneous Conversational Speech*, 2007.