2.6 INTELLECTUAL PROPERTY

1. FAR/DFARS Noncommercial Items IP Restrictions (Technical Data and Computer Software)

NONE.

2. FAR/DFARS Commercial Items IP Restrictions (Technical Data and Computer Software)

COMMERCIAL			
Technical Data / Computer Software To be Furnished With Restrictions	**Basis for Assertion**	**Asserted Rights Category**	**Name of Person Asserting Restrictions**
Decipher® Speech Recognition System	Commercial Computer Software	Commercial Rights (Vendor standard no cost agreement)	SRI International
SRILM – SRI Language Modeling Toolkit	Commercial Computer Software	Commercial Rights (www.speech.sri.com/projects/srilm/docs/ License)	SRI International
Algemy	Commercial Computer Software	Commercial Rights (Equivalent of Unlimited Rights, with waiver of warranty and liability)	Bosch & Volkswagen of America
Matlab	Commercial Computer Software	Commercial Rights (Vendor standard agreement)	Mathworks
SVMLight implementation of Support Vector Machines in C	Commercial Computer Software	Commercial Rights (Free for non-commercial use, no redistribution allowed)	Thorsten Joachims, Cornell University
HTK Hidden Markov Model Tool-kit	Commercial Computer Software	Commercial Rights (htk.eng.cam.ac.uk/docs/ license.shtml)	University of Cambridge
Speech data	Commercial Data	Commercial Rights (www.ldc.upenn.edu/ Catalog/)	Linguistic Data Consortium
R Statistics Packs	Commercial Computer Software	Commercial Rights (GNU General Public License)	Free Software Foundation (FSF)
Octave	Commercial Computer Software	Commercial Rights (GNU General Public License)	Free Software Foundation
Python	Commercial Computer Software	Commercial Rights (Python 2.4.2 license www.python.org/download/releases/2.4.2/license/)	Python Software Foundation (PSF)
Perl	Commercial Computer Software	Commercial Rights (GNU General Public License)	Free Software Foundation
DynaSpeak® Speech Recognition System	Commercial Computer Software	Commercial Rights	SRI International
QT User Interface Libraries	Commercial Computer Software	Commercial Rights (http://www.qtsoftware.com/products/ licensing)	Nokia, Inc.
XMLRPC-c	Commercial Computer Software	Commercial Rights (http://xmlrpc-c.sourceforge.net/)	Sourceforge.net
Apache XML RPC	Commercial Computer Software	Commercial Rights (http://ws.apache.org/xmlrpc/)	Apache open source project
Microsoft C++ 2008 Redistributable Package	Commercial Computer Software	Commercial Rights	Microsoft

COMMERCIAL			
Technical Data / Computer Software To be Furnished With Restrictions	**Basis for Assertion**	**Asserted Rights Category**	**Name of Person Asserting Restrictions**
Java Standard Edition Runtime Environment Version 6 (or later)	Commercial Computer Software	Commercial Rights (http://java.sun.com/javase/downloads/index.jsp)	Sun microsystems
Postgresql	Commercial Computer Software	Commercial Rights	Postgresql
Postgresql jdbc	Commercial Computer Software	Commercial Rights	Postgresql
SPHINX	Commercial Computer Software	Commercial Rights	Carnegie Mellon University

3. Non-FAR/DFARS IP Restrictions (Technical Data and Computer Software)

NONE.

4. Patent Dependencies

US Patent or Application #	Inventor(s)	Title	Issue Date	Owner
7,254,538	Hynek Hermansky, Sangita Sharma, and Daniel Ellis	Nonlinear mapping for feature extraction in automatic speech recognition	August 7, 2007	ICSI
11/758,650	Bratt et al.	Method and Apparatus for Speaker Recognition		SRI International and Stanford University
US 7,177,810	Bratt et al.	Method and Apparatus for Performing Prosody-based Endpointing of a Speech Signal	13-Feb-2007	SRI International

5. IP Representation

SRI International represents in good faith that to the best of its knowledge as of the date of this proposal, that SRI owns or possesses appropriate licensing rights to all other intellectual property proposed for use herein. However, SRI has not conducted any patent search or engaged counsel to provide any legal opinion regarding freedom to operate and provides no warranty of non-infringement of any third party rights.

2.7 SCHEDULE AND MILESTONES

The following table presents the schedule and milestones for the SCENIC program.

TASKS	Phase 1						Phase 2				Phase 3			
	Q1	Q2	Q3	Q4	Q5	Q6	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Task 1: Robust Features														
Auditory Model Inspired Temporal Representations	[redacted]													
Multistream Features Based on Gabor Filtering														
Pitch Sync. Feats & Robust Pitch Estimation														
Improved Feature Mapping														
Task Specific Feature Design														
Feature Combination														
Task 2: Advanced Modeling														
Structured JFA														
Selective Weighting and Missing Feature Modeling														
Large-Margin Training														
Task 3: Decision Making														
Module-Specific Adaptive Calibration														
Cross-Module Joint Calibration														
On The Fly Calibration for New Targets														
Task 4: Speech and Metadata Extraction														
System Development														
SAD Specific Feature Development														
SAD Combination by Scene Analysis														
Metadata Generation														
Task 5: Language and Speaker Recognition														
System Development and Experimental Setup														
Integration of Modeling Technologies														
Selection of Feature Types and Modeling														
Task 6: Keyword Spotting														
Language Dependent Recognition Sys Development														
FST-Based Hybrid KWS System														
Re-Scoring KWS														
Task 7: Software and Team Integration														
Evaluation System														
Graphical User Interface														
Data Processing														
Reporting														

♦ = internal milestone; √ = evaluation system delivery; ■ = quarterly technical reports

The following table presents the performance targets by phase and task for the SCENIC program.

	SAD	Language ID	Speaker ID	Keyword Spotting
Phase 1	5% Miss 3% FA Error – 3.75%	7% Miss 10% FA Error – 8.2%	10% Miss 4% FA Error – 5.5%	30% Miss 4% FA Error – 7%
Phase 2	4% Miss 1.5% FA Error – 2.2%	5% Miss 5% FA Error – 5%	10% Miss 2.5% FA Error – 4%	20% Miss 3% FA Error – 5.2%
Phase 3	3% Miss 1% FA Error – 1.5%	5% Miss 1% FA Error – 1.7%	10% Miss 1.5% FA Error – 2.6%	15% Miss 2% FA Error – 3.5%

2.8 PERSONNEL, QUALIFICATIONS, AND COMMITMENTS

The following biographies for all key members of the SCENIC team summarize qualifications and unique capabilities. Key personnel from SRI are Andreas Stolcke, Horacio Franco, Martin Graciarena, Dimitra Vergyri, Lynn Voss, and Nicolas Scheffer. Key personnel from ICSI are Nelson Morgan and Daniel Ellis. Other key personnel are Abeer Alwan from UCLA and John Hansen from UTD.

Key Individual Time Commitments

Key Individual	Project	Current / Pending	2010 (hr)	2011 (hr)	2012 (hr)	2013 (hr)	2014 (hr)
[redacted]							

(b) (4)

[redacted] (b) (4)

Andreas Stolcke, Ph.D.

Summary of Qualifications and Unique Capabilities:

Dr. Andreas Stolcke holds a Ph.D. in computer science from UC Berkeley, and has been with SRI International since 1994, currently as a Senior Research Engineer. Dr. Stolcke has published more than 200 journal articles and other peer-reviewed papers, with contributions to machine learning of languages, parsing, acoustic and language modeling for speech recognition, computational modeling of speech prosody, information extraction from speech, and ASR-based techniques for speaker recognition. He has served on the editorial boards or as co-editor for the journals *Computational Linguistics*, *IEEE Transactions on Audio Speech and Language Processing*, and *Computer Speech and Language*, and is a Senior Member of the IEEE Signal Processing Society. He wrote the widely used, open-source SRI Language Modeling (SRILM) toolkit.

Previous Accomplishments:

Dr. Stolcke was the PI for SRI's team on DARPA's EARS program, where he directed research in large-vocabulary speech recognition and recognition of structural metadata, heading a team of size similar to the proposed RATS effort. Subsequently, he was a Task Leader for SRI's team on the DARPA GALE program, responsible for coordinating speech recognition research and development among six partner sites, as well as with other team members working on machine translation and information distillation. Dr. Stolcke was part of the SRI team that fielded the winning system in the 2001 DARPA-sponsored evaluation of recognition of Speech in Noisy Environments (SPINE), and he was a key contributor to the SGI/OGI system for the NIST 2006 Spoken Term Detection evaluation. Dr. Stolcke has directed work at SRI and ICSI in meeting recognition, contributing to DARPA's PAL/CALO program, and consistently obtaining leading performance in NIST Rich Transcription evaluations. For the past six years, Dr. Stolcke has been co-leading SRI's speaker recognition program (funding from DoD, Sandia Labs, and IARPA), which resulted in numerous contributions to the state of the art in that field, and outstanding results in NIST evaluations.

Selected Publications:

A. Stolcke, B. Chen, H. Franco, V. R. R. Gadde, M. Graciarena, M.-Y. Hwang, K. Kirchhoff, A. Mandal, N. Morgan, X. Lin, T. Ng, M. Ostendorf, K. Sonmez, A. Venkataraman, D. Vergyri, W. Wang, J. Zheng, Q. Zhu (2006), Recent Innovations in Speech-to-Text Transcription at SRI-ICSI-UW. *IEEE Trans. Audio, Speech and Language Processing* 14(5), 1729–1744.

A. Stolcke, X. Anguera, K. Boakye, O. Cetin, A. Janin, M. Magimai-Doss, C. Wooters, J. Zheng (2008), The SRI-ICSI Spring 2007 Meeting and Lecture Recognition System, R. Stiefelhagen, R. Bowers, and J. Fiscus (eds.), Multimodal Technologies for Perception of Humans. International Evaluation Workshops CLEAR 2007 and RT 2007, Springer *Lecture Notes in Computer Science* 4625, pp. 450–463.

A. Stolcke, S. Kajarekar, L. Ferrer, E. Shriberg (2007), Speaker Recognition with Session Variability Normalization Based on MLLR Adaptation Transforms, *IEEE Transactions on Audio, Speech and Language Processing*, 15(7), 1987–1998. Special issue on speaker and language recognition.

S. S. Kajarekar, L. Ferrer, A. Stolcke, and E. Shriberg (2008), Voice-based speaker recognition combining acoustic and stylistic features, in *Advances in Biometrics: Sensors, Algorithms and Systems* (N. K. Ratha and V. Govindaraju, eds.), pp. 183--201, London: Springer.

A. Stolcke and S. Kajarekar (2008), Recognizing Arabic speakers with English phones, in *Proc. Odyssey Speaker and Language Recognition Workshop*, Stellenbosch, South Africa.

V. R. Rao Gadde, A. Stolcke, D. Vergyri, J. Zheng, K. Sonmez, A. Venkataraman (2002), Building an ASR System for Noisy Environments: SRI's 2001 SPINE Evaluation System, *Proc. Intl. Conf. on Spoken Language Processing*, vol. 3, pp. 1577–1580, Denver.

M. Akbacak, D. Vergyri, A. Stolcke (2008), Open-Vocabulary Spoken Term Detection Using Graphone-Based Hybrid Recognition Systems, *Proc. IEEE ICASSP*, pp. 5240–5243, Las Vegas.

E. Shriberg, A. Stolcke (2004), Direct Modeling of Prosody: An Overview of Applications to Automatic Speech Processing, *Proc. Inl. Conf. on Speech Prosody*, pp. 575--582, Nara, Japan.

Horacio Franco, Dr. Eng.

Summary of Qualifications and Unique Capabilities:

Horacio Franco is Chief Scientist in the Speech Technology and Research Laboratory (STAR) at SRI. He received his Engineer Degree in Electronics in 1978, and his Doctor in Engineering degree in 1996, both from the University of Buenos Aires. Since joining SRI in 1990, he has contributed to the areas of acoustic modeling, hybrid HMM-neural net speech recognition approaches, speech recognizer architectures, speech technology for language learning, noise-robust speech recognition, and speech-to-speech translation systems. He has co-authored more than 70 papers, and several communications and book chapters. Dr. Franco is co-author of several U.S. patents in aspects of speech recognition, speech technology for language learning, and noise-robust speech recognition. He has also been active in leading development and deployment of SRI's speech technology in different areas of government and commercial use.

Previous Accomplishments:

Dr. Franco has been a principal investigator for several DARPA-funded projects at SRI: Speech Modeling with Neural Networks, the TRP Affordable Language Education Consortium, and the speech recognition component of OneWay and SRI's Babylon speech translation projects, as well as leader of the Acoustic Modeling task in SRI's EARS project. Dr. Franco also led an effort, funded by the Collaborative Technology Alliance, to develop noise-robust command and control speech recognition for military vehicles. He led several projects to develop noise-robust speech recognition systems for military air vehicles such as the Joint Strike Fighter and the Apache helicopter. He led a Raytheon subcontract from the DARPA BOSS program to develop embedded wordspotting, speaker ID, language ID, and topic detection capabilities.

Selected Publications:

M. Akbacak, H. Franco, M. Frandsen, S. Hasan, H. Jameel, A. Kathol, S. Khadivi, X. Lei, A Mandal, S. Mansour, K. Precoda, C. Richey, D. Vergyri, W. Wang, M., Yang, J. Zheng, Recent Advances in SRI's IraqComm ™ Iraqi Arabic-English Speech-to-Speech Translation System, Proc. IEEE ICASSP (Taipei), Apr. 2009.

A. Stolcke, B. Chen, H. Franco, R. Gadde, M Graciarena, M.H. Hwang, K. Kirchhoff, A. Mandal, N. Morgan, T. Ng, M. Ostendorf, K. Sönmez, A. Venkataraman, D. Vergyri, W. Wang, J. Zheng, Q. Zhu, Recent Innovations in Speech-to-Text Transcription at SRI-ICSI-UW, IEEE Trans. Speech, Acoustics and Audio Processing 14(5), 1729–1744, 2006.

Jing Zheng, Horacio Franco, Andreas Stolcke (2004), Effective Acoustic Modeling for Rate-of-Speech Variation in Large Vocabulary Conversational Speech Recognition, Proc. ICSLP, pp. 401-404, Jeju, Korea.

H. Franco, J. Zheng, K. Precoda, F. Cesari, V. Abrash, D. Vergyri, A. Venkataraman, H. Bratt, C. Richey, A. Sarich (2003), Development of Phrase Translation Systems for Handheld Computers: From Concept to Field, Proc. Eurospeech, Geneve, Switzerland.

M. Graciarena, H. Franco, Greg Myers, Victor Abrash, Robust Feature Compensation in Nonstationary and Multiple Noise Environments, Proc. Eurospeech, Lisbon, Portugal, 2005.

M. Graciarena, H. Franco, Unsupervised Noise Model Estimation for Model-Based Robust Speech Recognition, Proc. IEEE Automatic Speech Recognition and Understanding Workshop, 2003

H. Franco, J. Zheng, K. Precoda, F. Cesari, V. Abrash, D. Vergyri, A. Venkataraman, H. Bratt, C. Richey, A. Sarich (2003), Development of Phrase Translation Systems for Handheld Computers: From Concept to Field, Proc. Eurospeech, Geneve, Switzerland.

J. Zheng, H. Franco, A. Stolcke (2003), Modeling Word-Level Rate-of-Speech Variation in Large Vocabulary Conversational Speech Recognition, Speech Communication, 41, pp. 273–285.

H. Franco, J. Zheng, J. Butzberger, F. Cesari, M. Frandsen, J. Arnold, V. R. Rao Gadde, A. Stolcke, V. Abrash (2002), DynaSpeak: SRI's Scalable Speech Recognizer for Embedded and Mobile Systems, Proc. Second Intl. Conf. on Human Language Technology Research, pp. 25–30, Mar. 24–27, San Diego, California.

Martin Graciarena, Ph.D

Summary of Qualifications and Unique Capabilities:

Martin Graciarena is a Research Engineer in the STAR Laboratory at SRI. He is a reviewer of noise-robust speech and speaker recognition papers on the most relevant publications and conferences in the area such as the *IEEE Transactions on Speech and Language* and the *IEEE International Conference on Acoustics, Speech and Signal Processing*. He has published more than 25 papers on noise-robust speech and speaker recognition, and he has done pioneering work on combination of standard and throat microphones for robust speech recognition. His paper about this research is one of the most cited papers in this area. He has improved feature mapping techniques and model-based noise-robust speech recognition, has improved state-of-the-art speech recognition systems by incorporating prosodic-based features, done research on noise-robust speaker identification in Arabic language as well as improving channel robustness of high-level prosodic-based speaker identification systems. He holds one patent on a noise-robust speech recognition technique and is an IEEE member.

His current research interests are improving the feature extraction processing based on human auditory modeling, improving feature mapping techniques by combining heterogeneous transformations, and improving speech activity detection in combination with speech and background improved feature extraction.

Previous Accomplishments:

Dr. Graciarena did pioneering work on combining speech- and throat-based microphones for robust speech recognition under the DARPA ROAR program. He was a researcher in the DARPA EARS program where he improved the performance of state-of-the-art speech recognition systems by the incorporation of voicing features in the acoustic front end. He improved the robustness to channel degradation of state-of-the-art prosodic-based speaker identification systems for Sandia National Laboratory.

Selected Publications:

M. Graciarena, S. Kajarekar, A. Stolcke, E. Shriberg (2007). Noise Robust Speaker Identification for Spontaneous Arabic Speech. *Proc. IEEE ICASSP*, vol. 4, pp. 245–248, Honolulu.

L. Ferrer, M. Graciarena, A. Zymnis, E. Shriberg (2008). System Combination Using Auxiliary Information for Speaker Verification. *Proc. IEEE Intl. Conf Acoustics, Speech and Signal Processing*, 2008.

M. Graciarena, H. Franco, G. Myers, V. Abrash (2005). Robust Feature Compensation in Nonstationary and Multiple Noise Environments. *Proc. Eurospeech*, pp. 985–988, Lisbon, Portugal.

M. Graciarena, H. Franco, J. Zheng, D. Vergyri, A. Stolcke (2004). Voicing Feature Integration in SRI's Decipher LVCSR System. *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing*, vol. 1, pp. 921–924, Montreal, Canada.

M. Graciarena, H. Franco, G. Myers, C. Cowan, F. Cesari, V. Abrash (2004).. Combination of Standard and Throat Microphones for Robust Speech Recognition in Highly Noisy Environments, *Proc. Intl. Conf. Spoken Language Processing*, Jeju Island, Korea.

M. Graciarena, H. Franco, K. Sonmez, H. Bratt (2003). Combining Standard and Throat Microphones for Robust Speech Recognition. *IEEE Signal Processing Letters*, vol. 10, no. 3, pp. 72–74, Mar. 2003

M. Graciarena, H. Franco (2003). Unsupervised Noise Model Estimation for Model-Based Robust Speech Recognition, *Proc. IEEE Automatic Speech Recognition and Understanding Workshop*, pp. 351–356, Virgin Islands.

G. Myers, H. Franco, M. Graciarena, C. Cowan, F. Cesari, V. Abrash (2003), Noise-Robust Spoken Language Interface, *CTA 2003 Symposium on Robotics*, pp. 135–139, Adelphi, New York.

M. Graciarena (2000). Adaptive Model-Based Technique for Robust Speech Recognition. *Proc. 34th Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA.

M. Graciarena (2000). Maximum Likelihood Noise HMM Estimation in Model-Based Robust Speech Recognition. *Proc. 6th International Conference on Spoken Language Processing*, vol. 3, pp. 598–601, Beijing.

Dimitra Vergyri, Ph.D.

Summary of Qualifications and Unique Capabilities:

Dimitra Vergyri is a Research Engineer at SRI. She obtained her Ph.D. degree from Johns Hopkins University in 2000, when she joined the Speech Technology and Research laboratory at SRI. She has published more than 40 papers in refereed conferences and journals. Her research interests include acoustic and language modeling for languages with sparse training data, spoken term detection, machine translation and information extraction from speech. She has served since 2009 as an Associate Editor for the *IEEE Transactions on Audio, Speech and Language Processing*.

Previous Accomplishments:

Dr. Vergyri has worked as one of the key personnel in four major DARPA-funded projects: EARS, CALO, GALE and TRANSTAC, where she participated in regular evaluations for speech recognition and speech translation. She led the ASR development for Arabic systems under GALE, and Pashto, Farsi and Arabic under TRANSTAC. She also participated as part of the SRI team in the Spoken Term Detection NIST evaluation in 2006, where SRI had the best submission for the Meetings condition, and in the SPeech In Noise Environments NIST evaluation in 2001, where SRI had the best performance.

Selected Publications:

D. Vergyri, A. Stolcke, and G. Tur, "Exploiting user feedback for language model adaptation in meeting recognition," in *Proc. IEEE ICASSP*, (Taipei), pp. 4737--4740, April 2009.

D. Vergyri, A. Mandal, W. Wang, A. Stolcke, J. Zheng, M. Graciarena, D. Rybach, C. Gollan, R. Schlüter, K. Kirchhoff, A. Faria, and N. Morgan, "Development of the SRI/Nightingale Arabic ASR system," in *Proc. Interspeech*, (Brisbane, Australia), pp. 1437--1440, September 2008.

D. Vergyri, I. Shafran, A. Stolcke, R. R. Gadde, M. Akbacak, B. Roark, and W. Wang, "The SRI/OGI 2006 Spoken Term Detection system," in *Proc. Interspeech/Eurospeech*, (Antwerp), pp. 2393--2396, August 2007.

M. Akbacak, D. Vergyri, and A. Stolcke, "Open-vocabulary spoken term detection using graphone-based hybrid recognition systems," in *Proc. IEEE ICASSP*, (Las Vegas), pp. 5240--5243, March 2008.

K. Precoda, J. Zheng, D. Vergyri, H. Franco, C. Richey, A. Kathol, and S. Kajarekar, "IraqComm: A next generation translation system," in *Proc. Interspeech*, (Antwerp, Belgium), August 2007.

H. Lin, J. Bilmes, D. Vergyri, and K. Kirchhoff, "OOV detection by joint Word/Phone lattice alignment," in *Proc. Automatic Speech Recognition and Understanding Workshop*, (Kyoto, Japan), p. 478, December 2007.

K. Kirchhoff, D. Vergyri, J. Bilmes, K. Duh, and A. Stolcke, "Morphology-based language modeling for conversational Arabic speech recognition," *Computer Speech and Language*, vol. 20, pp. 589--608, October 2006.

D. Vergyri and K. Kirchhoff, "Automatic diacritization of Arabic for acoustic modeling in speech recognition," in *COLING Workshop on Arabic-script Based Languages*, (Geneva, Switzerland), August 2004.

D. Vergyri, K. Kirchhoff, R. Gadde, A. Stolcke, and J. Zheng, "Development of a conversational telephone speech recognizer for Levantine Arabic," in *Proc. Eurospeech*, (Lisbon), pp. 1613--1616, September 2005.

V. R. R. Gadde, A. Stolcke, D. Vergyri, J. Zheng, K. Sonmez, and A. Venkataraman, "Building an ASR system for noisy environments: SRI's 2001 SPINE evaluation system," in *Proc. International Conference on Spoken Language Processing*, vol. 3, (Denver, CO), pp. 1577--1580, September 2002.

L. Lynn Voss

Summary of Qualifications and Unique Capabilities:

Mr. Voss is a Program Manager in the Software Engineering Program at SRI. He received his B.S. in Computer Science (summa cum laude) from Montana State University-Bozeman and MBA in Technology Management from the University of Phoenix. He currently manages a team of 11 software engineers in the SRI Helena, Montana office. He has extensive experience in software systems engineering; software engineering management, design and development; graphical user interface design and implementation; automated build systems; configuration management; web design; and database programming.

Previous Accomplishments

Mr. Voss is currently the Architecture and Integration lead on FAUST – the DARPA Machine Reading project at SRI. This integrates the work of SRI with efforts from seven (mostly university) subcontractors in the areas of natural language processing, machine learning, and probabilistic reasoning to transform the knowledge contained in natural texts so that it can be utilized by artificial intelligence reasoning systems.

Previously, Mr. Voss was the Engineering Manager for the Meeting Assistance effort of the Cognitive Assistant that Learns and Organizes (CALO) project within the DARPA-sponsored Personalized Assistant that Learns (PAL) program. He was responsible for overall system design; managed design, implementation and project timeline of internal developers and university and commercial subcontractors; procured hardware; resolved licensing issues; and performed testing and integration with an independent test evaluator. The CALO Meeting Assistance project addressed multi-modal, multiparty dialogue understanding, including speech recognition, natural language and discourse and dialogue understanding; handwriting and sketch recognition; computer vision: coarse location and pointing, gaze estimation, and audio-visual speech recognition.

Selected Publications:

G. Tur, A. Stolcke, L. Voss, S. Peters, D. Hakkani-Tür, J. Dowding, B. Favre, R. Fernandez, M. Frampton, M. Frandsen, C. Frederickson, M. Graciarena, D. Kintzing, K. Leveque, S. Mason, J. Niekrasz, M. Purver, K. Riedhammer, E. Shriberg, J. Tien, D. Vergyri, F. Yang, "The CALO Meeting Assistant System," To appear in *IEEE Transactions on Audio, Speech, and Language Processing*, 2010.

G. Tur, A. Stolcke, L. Voss, J. Dowding, B. Favre, R. Fernandez, M. Frampton, M. Frandsen, C. Frederickson, M. Graciarena, D. Hakkani-Tür, D. Kintzing, K. Leveque, S. Mason, J. Niekrasz, S. Peters, M. Purver, K. Riedhammer, E. Shriberg, J. Tien, D. Vergyri, F. Yang, "The CALO Meeting Speech Recognition and Understanding System," in *Proc. IEEE Workshop on Spoken Language Technologies*, (Goa, India), pp. 69-72, 2008.

L. Voss, P. Ehlen, and the DARPA CALO Meeting Assistant Project Team, "The CALO Meeting Assistant (demonstration)," *Human Language Technologies: The Annual Conference of the North American Chapter of the Association for Computational Linguistics* (NAACL-HLT), (Rochester, NY), 2007.

Nicolas Scheffer, Ph.D.

Summary of Qualifications and Unique Capabilities:

Dr. Nicolas Scheffer obtained a Ph.D. in 2006 from University of Avignon, France, in Computer Science. Dr. Scheffer also holds a Master of Engineering in Control Systems Engineering as well as a Master of Science in EEE. He has been at SRI International since 2007, currently as a Research Engineer. Dr. Scheffer published numerous papers in international conferences and workshops. He received the IBM best paper award at the 2006 IEEE Odyssey Speaker and Language Recognition Workshop for his work on discriminative training in speaker recognition.

Previous Accomplishments:

Dr. Nicolas Scheffer has worked on speaker recognition for the past 6 years, including his Ph.D. thesis. His work has focused on acoustic modeling (SVMs and factor analysis) for speaker ID, and unsupervised feature extraction using universal background models. He has participated in all NIST speaker recognition evaluations since 2004, as well as in other evaluations for speaker diarization and tracking.

Dr. Scheffer interests also lie in technology transfer and software management best practices. As such, he is a previous member of the core team of the ALIZE project, an open-source tool to implement state-of-the-art speaker verification systems, widely used in the community. He acquired strong skills in software packaging, development, maintenance and community management.

Dr. Scheffer has also been leading the software development effort in SRI's speaker recognition program (funded by DoD, Sandia Labs, and IARPA), and has a proven track record in fast transfer of complex technologies to government clients.

Selected Publications:

N. Scheffer, R. Vogt "On the use of speaker superfactors for speaker recognition", in *Proc. IEEE ICASSP*, pp. 4410-4413, Dallas, March 2010.

S. S. Kajarekar, N. Scheffer, M. Graciarena, E. Shriberg, A. Stolcke, L. Ferrer, and T. Bocklet, "The SRI NIST 2008 speaker recognition evaluation system," in *Proc. IEEE ICASSP*, Taipei, pp. 4205-4209, April 2009.

Benoit G. B. Fauve, Driss Matrouf, N. Scheffer, Jean-François Bonastre, John S. D. Mason. State-of-the-Art Performance in Text-Independent Speaker Verification Through Open-Source Software, *IEEE Trans. Audio, Speech & Lang. Proc.* 15(7): 1960-1968, 2007.

Matrouf, D. and Scheffer, N. and Fauve, B. and Bonastre, J.F.: A Straightforward and Efficient Implementation of the Factor Analysis Model for Speaker Verification. In *Proc. Interspeech*, pp. 1242-1245, Antwerp, 2007.

Nicolas Scheffer, Jean-François Bonastre, "A multi-class framework within an Acoustic Event Sequence system for Speaker Verification", *Proc. Interspeech*, Pittsburgh, 2006.

Nicolas Scheffer, Jean-Francois Bonastre, "A UBM-GMM driven discriminative approach for Speaker Verification", *Proc. IEEE Odyssey Workshop*, San Juan, Puerto Rico, 2006 (IBM best paper student award).

Nelson Morgan, Ph.D.

Summary of Qualifications and Unique Capabilities:

Nelson Morgan is the Director of the International Computer Science Institute (ICSI) and the leader of its Speech Group. He is also a Professor-in-residence at UC Berkeley, where he has mentored many Ph.D. students and postdoctoral Fellows. He is a former Editor-in-chief of *Speech Communication*, and is an associate member of the IEEE Speech and Language Processing Technical Committee. He is a Fellow of the IEEE. In 1997 he received the *Signal Processing Magazine* best paper award. Professor Morgan has roughly 200 publications including three books; his most recent book is a text (written jointly with Ben Gold) on speech and audio signal processing, which is currently being revised with Professor Dan Ellis. He holds a number of patents in speech processing methods, including one that has been used in millions of CDMA cell phones. He was the first to use neural networks for speech classification in a commercial application (at National Semiconductor in the early 1980s). He is also the co-inventor (along with Herve Bourlard) of connectionist HMM/MLP statistical sequence recognition, particularly as applied to speech. He developed methods for the acoustical measurement of speaking rate that have been used by many researchers. His initiative to develop processing techniques for use with speech from meetings led to the development of the largest U.S. public corpus of transcribed and annotated speech from meetings, and to a series of NIST-run evaluations of the technology.

His current research interests include the redesign from first principles of the primary signal processing used in speech recognition systems, particularly for noisy and reverberant signals. This work is particularly oriented toward the incorporation of new information about the successful working of the human auditory system (especially from measurements in primary auditory cortex) for speech processing tasks. Professor Morgan has led efforts in this direction at ICSI.

Previous Accomplishments:

The cell phone algorithms referred to above were based on the RASTA innovations that Professor Morgan developed (with co-inventor Hynek Hermansky) in the early 1990s. He was the Principal Investigator for the multi-site coalition funded by the DARPA EARS Novel Approaches project, which was the 2002-2005 U.S. Government program focusing on long-term progress in English language speech recognition. This work led to significant reductions in word error rate based on preliminary versions of the ICSI algorithms that are proposed here. These methods were then applied to Arabic and Mandarin speech recognition in the DARPA GALE project, with similar success, first for the SRI team and more recently for the IBM team. Results reported in a recent speech conference (Interspeech 2009) further showed that a new version of the algorithms proposed by Professor Morgan were particularly effective for noisy speech; this was partly funded by the Intelligence Community postdoctoral Fellow program (for postdoctoral Fellow Sherry Zhao, who worked with Professor Morgan).

Selected Publications:

S. Y. Zhao, S. Ravuri, and N. Morgan, "Multi-Stream to Many-Stream: Using Spectro-Temporal Features for ASR," in *Proceedings of the 10th International Conference of the International Speech Communication Association*, pp. 2951-2954, 2009

D. Gelbart, N. Morgan, and A. Tsymbal, "Hill-Climbing Feature Selection for Multi-Stream ASR," in *Proceedings of the 10th International Conference of the International Speech Communication Association (Interspeech 2009)*, 2009, pp. 2967-2970.

J. M. Baker, L. Deng, J. Glass, S. Khudanpur, C. Lee, N. Morgan, and D. O'Shaughnessy, "Research Developments and Directions in Speech Recognition and Understanding, Part 2," *IEEE Signal Processing Magazine*, vol. 26, no. 4, pp. 78-85, July 2009.

J. M. Baker, L. Deng, J. Glass, S. Khudanpur, C. Lee, N. Morgan, and D. O'Shaughnessy, "Research Developments and Directions in Speech Recognition and Understanding, Part 1," *IEEE Signal Processing Magazine*, vol. 26, no. 3, pp. 75-80, May 2009.

Morgan, N., Zhu, Q., Stolcke, A., Sonmez, K., Sivadas, S., Shinozaki, T., Ostendorf, M., Jain, P., Ellis, D., Doddington, G., Chen, B., Cetin, O., Bourlard, H., and Athineos, M. "Pushing the Envelope - Aside: Beyond the Spectral Envelope as the Fundamental Representation for Speech Recognition." IEEE Signal Processing Magazine, pp. 81-88, September 2005.

N. Morgan, D. Baron, S. Bhagat, H. Carvey, R. Dhillon, J. Edwards, D. Gelbart, A. Janin, A. Krupski, B. Peskin, T. Pfau, E. Shriberg, A. Stolcke, and C. Wooters, "Meetings about meetings: Research at ICSI on speech in multiparty conversations," in *Proc. 2003 IEEE Intl. Conf. on Acoustics, Speech, and Signal Processing*, Vol. 4. Piscataway, NJ: IEEE Press, 2003. pp. 740-743.

Gold, B., and Morgan, N., "Speech and Audio Signal Processing," Wiley Press, New York, 1999.

Morgan, N., and Bourlard, H., "Continuous Speech Recognition: An Introduction to the Hybrid HMM/Connectionist Approach," *IEEE Signal Processing Magazine*, pp 25-42, May 1995 (won IEEE 1997 SP Magazine Award).

Hermansky, H., and Morgan, N., "RASTA Processing of Speech", *IEEE Transactions on Speech and Audio Processing*, Special issue on Robust Speech Recognition, vol 2 no. 4, pp. 578-589, Oct., 1994.

Daniel P. W. Ellis, PhD

Summary of Qualifications and Unique Capabilities:

Daniel Ellis is an Associate Professor of Electrical Engineering at Columbia University as well as being an External Fellow of ICSI. He leads the Laboratory for Recognition and Organization of Speech and Audio, which is concerned with separating and recognizing information from all kinds of complex, real-world sounds, including speech, music, and environmental audio. He has more than 100 publications in these areas, and is currently collaborating with Prof. Nelson Morgan on a new edition of a textbook covering these topics. Dr. Ellis holds several patents on topics related to speech and audio processing, including one on Tandem Connectionist-HMM acoustic modeling (joint with Morgan), a technology that has contributed to recent improvements in the speech recognition systems developed by SRI and ICSI under DARPA projects EARS and GALE. He is well known for his work in computational auditory scene analysis, which seeks to separate out individual sounds in noisy mixtures in a manner analogous to human listeners. These techniques are currently being developed for applications in unedited video soundtracks. Dr. Ellis is also known as the author and maintainer of a range of software tools widely used by the speech and audio research community, including dpwcutils (a suite of sound file manipulation routines) and feacalc (a generalized feature extraction program).

Previous Accomplishments:

The Tandem technique mentioned above was developed under EU project RESPITE and achieved top performance in the 2001 Eurospeech AURORA noisy speech recognition evaluation (among 21 participants). Under DARPA EARS, he developed Frequency-Domain Linear Prediction speech features for parametric encoding of transient temporal events in speech. Recent NSF-funded research on personal audio and video soundtracks has demonstrated voice activity detection at SNRs ranges below the level of human intelligibility – much lower than the majority of work in the field. Another NSF-funded project has developed the MESSL system for separation of overlapping speech under reverberation, which shows best-in-the-world signal-to-noise ratio improvement for real-world conditions, and which can be combined with Eigenvoice models to separate speech signals to achieve state-of-the-art speech recognition without prior knowledge of speaker characteristics. Ongoing NGA-sponsored research has pioneered the detection and mining of speech and other sound categories in large archives of unedited, uncontrolled consumer video soundtracks.

Selected Publications:

D. Ellis and M. J. Reyes Gomez (2001). Investigations into Tandem Acoustic Modeling for the Aurora Task, *Proc. Eurospeech*, Special Event on Noise Robust Recognition, pp. 189-192, Denmark, September 2001.

M. Athineos and D. Ellis (2007). Autoregressive Modeling of Temporal Envelopes, *IEEE Trans. Signal Processing*, vol. 15 no. 11, pp. 5237-5245, Nov. 2007.

K. Lee and D. Ellis (2006). Voice Activity Detection in Personal Audio Recordings Using Autocorrelogram Compensation. *Proc. Interspeech*, pp. 1970-1973, Pittsburgh, Oct 2006.

M. Mandel, R. Weiss, and D. Ellis (2009). Model-Based Expectation-Maximization Source Separation and Localization *IEEE Trans. Audio, Speech, and Lang. Proc.*, accepted for publication.

R. Weiss and D. Ellis (2010). Speech separation using speaker-adapted eigenvoice speech models, *Computer Speech and Language*, vol. 24 no. 1, pp. 16-29, Jan 2010.

C. Cotton and D. Ellis (2010). Audio Fingerprinting to Identify Multiple Videos of an Event, *Proc. IEEE ICASSP*, Dallas, pp. 2386-2389, March 2010.

Abeer Alwan, Ph.D.

Summary of Qualifications and Unique Capabilities:

Abeer Alwan received her Ph.D. in EECS from MIT in 1992. She has been with the Electrical Engineering Department at UCLA as an Assistant Professor (1992-1996), Associate Professor (1996-2000), Professor (2000-present), Vice Chair of the BME program (1999-2001), Vice Chair of EE Graduate Affairs (2003-2006), and Area Director of Signals and Systems (2006-present). She established and directs the Speech Processing and Auditory Perception Laboratory at UCLA. Her research interests include modeling human speech production and perception mechanisms and applying these models to improve speech-processing applications such as noise-robust ASR.

She is the recipient of the NSF Research Initiation Award (1993), the NIH FIRST Career Development Award (1994), the UCLA-TRW Excellence in Teaching Award (1994), the NSF Career Development Award (1995), and the Okawa Foundation Award in Telecommunications (1997). Dr. Alwan is an elected member of Eta Kappa Nu, Sigma Xi, Tau Beta Pi, and the New York Academy of Sciences. She served, as an elected member, on the Acoustical Society of America Technical Committee on Speech Communication (1993-1999, and 2005-present), on the IEEE Signal Processing Technical Committees on Audio and Electroacoustics (1996-2000) and on Speech Processing (1996-2001, 2005-present). She is a member of the Editorial Board of Speech Communication and was an editor-in-chief of that journal (2000-2003) and an Associate Editor of the IEEE Transactions on Audio, Speech, and Language Processing (2007-2010). Dr. Alwan is chair of the IEEE Flanagan Speech and Audio Processing Award committee. She is a Fellow of the Acoustical Society of America and of the IEEE. She was a 2006-2007 Fellow of the Radcliffe Institute for Advanced Study at Harvard University, and is a 2010-2012 Distinguished Lecturer for the International Speech Communication Association (ISCA).

Previous Accomplishments:

Dr. Alwan's work has been supported by DARPA (in the 1990 for work on speech coding and noise cancellation on mobile devices), NSF (since 1993 for work related to human and machine speech recognition in noise, landmark-based ASR, and voice quality), and NIH (human speech perception in noise). Dr. Alwan's group pioneered several noise-robust techniques such as peak isolation and adaptation (1997), variable frame rate analysis and amplitude demodulation of spectra (2000, 2004), weighted Viterbi recognition (2002), HMM-based signal reconstruction techniques (2008), and noise-robust voice activity detection (2010). This work led to significant reductions in word error rate in various tasks. Dr. Alwan's group is one of the few groups that have carefully studied the effects of channel packet loss and various wireless channel conditions on ASR (the other groups are in Denmark, Spain, Greece, and Germany). She continues to contribute to the above-mentioned areas, and the RATS proposed project will leverage that work.

Selected Publications:

L. N. Tan, B. J. Borgstrom and A. Alwan, "Voice Activity Detection using Harmonic Frequency Components in Likelihood Ratio Test," ICASSP 2010.

B. J. Borgstrom and A. Alwan, "HMM-Based Reconstruction of Unreliable Spectrographic Data for Noise Robust Speech Recognition", IEEE Trans. on Audio, Speech, and Language Processing (TASLP), Vol. 18, No. 5, July 2010.

A. Alwan, " Dealing with Limited and Noisy Data in ASR: A Hybrid Knowledge-Based and Statistical Approach," Keynote Speech at Interspeech 2008, pp. 11-15.

B. J. Borgstrom, A. Bernard, and A. Alwan, "Error Recovery - Channel Coding and Packetization," Chapter 8 in Automatic Speech Recognition on Mobile Devices and over Communication Networks, Springer-Verlag, Editors: Z.-H. Tan and B. Lindberg, pp. 163-185, 2008.

X. Cui and A. Alwan, "Robust Speaker Adaptation by Weighted Model Averaging Based on the Minimum Description Length Criterion," *Proc. IEEE Trans. Audio Speech Lang. Proc.*, vol. 15, no. 2, pp. 652-660, Feb. 2007.

Jintao Jiang, Marcia Chen, Abeer Alwan, "On the perception of voicing in syllable-initial plosives in noise", *Journal of the Acoustical Society of America*, vol. 119, no. 2, pp. 1092-1105, February 2006.

H. You, Q. Zhu, and A. Alwan, "Entropy-base Variable Frame Rate Analysis of Speech Signals and Its Application to ASR," in *Proc. IEEE ICASSP*, pp. 549-552. Montreal, Canada, May 2004.

Q. Zhu and A. Alwan, "Non-linear feature extraction for robust recognition in stationary and non-stationary noise," *Computer Speech, and Language*, 17(4): 381-402, Oct. 2003.

X. Cui, Al. Bernard, and A. Alwan, "A Noise-Robust ASR Back-end Technique Based on Weighted Viterbi Recognition," in *Proc. EUROSPEECH*, Switzerland, pp. 2169-2172, Sept. 2003.

A. Bernard and A. Alwan, "Low-bitrate Distributed Speech Recognition for Packet-based and Wireless Communication" *IEEE Trans. Audio Speech Lang. Proc.*, vol. 10, no. 8, pp. 570-580, Nov. 2002.

Q. Zhu and A. Alwan, "On the use of variable frame rate analysis in speech recognition," Proc. IEEE ICASSP, Istanbul, Turkey, Vol. III, pp. 1783-1786, June 2000.

B. Strope and A. Alwan, "Modeling the perception of pitch-rate amplitude modulation in noise," Proc. *of the NATO ASI on Computational Hearing*, pp. 117-122, July 1998.

B. Strope and A. Alwan, "A model of dynamic auditory perception and its application to robust word recognition," *IEEE Trans. Speech Audio Lang. Proc.*, vol. 5, no. 5, pp. 451-464, September 1997.

John Hansen, Ph.D.

Summary of Qualifications and Unique Capabilities:

John Hansen is Professor and Dept. Head of Electrical Engineering at the Univ. of Texas at Dallas (the 13th largest EE Dept. in the USA), and holds the Distinguished University Chair in Telecommunications Engineering. He also holds a joint appointment as Professor in the School of Behavioral and Brain Sciences (Speech & Hearing). At UTD, he established the Center for Robust Speech Systems (CRSS) which is part of the Human Language Technology Research Institute. Previously, he served as Dept. Chair and Professor of Speech, Language and Hearing Sciences (SLHS), as well as Prof. in Electrical & Computer Engineering, at Univ. of Colorado Boulder (1998-2005), where he co-founded the Center for Spoken Language Research. In 1988, he established the Robust Speech Processing Laboratory (RSPL) – the first U.S.-based group specifically focused on robustness issues in speech systems, and continues to direct research activities in CRSS at UTD. He was named IEEE Fellow for contributions in *Robust Speech Recognition in Stress and Noise*, and is currently serving as a Member of the IEEE Signal Processing Society Speech Technical Committee (2005-08; 2010-13; elected Chair-elect in 2010), and Educational Technical Committee (2005-08; 2008-10). Previously, he served as Technical Advisor to the U.S. Delegate for NATO (IST/TG-01), as well as holding numerous positions in the IEEE Signal Processing Society, Acoustical Society of America, and Inter. Speech Communications Association. He supervised 50 (22 PhD, 28 MS/MA) thesis candidates, was recipient of The 2005 University of Colorado Teacher Recognition Award as voted on by the student body, author/co-author of 352 journal and conference papers and 8 textbooks in the field of speech processing and language technology, coauthor of the textbook *Discrete-Time Processing of Speech Signals* (IEEE Press, 2000), co-editor of *DSP for In-Vehicle and Mobile Systems* (Springer, 2004), *Advances for In-Vehicle and Mobile Systems: Challenges for International Standards* (Springer, 2006), *In-Vehicle Corpus and Signal Processing for Driver Behavior* (Springer, 2008), and lead author of the report *The Impact of Speech Under 'Stress'* on Military Speech Technology, (NATO RTO-TR-10, 2000). His research interests span the areas of digital speech processing, analysis and modeling of speech and speaker traits, speech enhancement, feature estimation in noise, robust speech recognition with emphasis on spoken document retrieval, and in-vehicle interactive systems for hands-free low-cognitive-load human-computer interaction.

Previous Accomplishments:

The noise and speaker compensation methods referred to in the proposed work were based on methods to improve word recognition in noise and stress for the U.S. Air Force and SPAWAR systems. Many of these advancements have been used by NATO RSG.10, and his enhancement algorithms based on Auto-LSP are in use by AT&T and other military groups. He served as Project Lead for the DARPA Communicator effort on In-Vehicle Speech Systems for Route Navigation (2002-2004), which included advancements in robust features and array processing for noisy speech recognition. He also served as co-PI and speech research lead for a project supported by the NSF Digital Library Initiative, entitled *National Gallery of the Spoken Word* which resulted in the first nationwide search engine for audio archives of historical content from the last 110 years (60,000 hours of historical documents). Many of the advancements to address unknown noise, channel, and speaker viabilities resulting from that 5-year study are now in use by researchers in speech recognition. More recently, he has made advancements in robustness for Speaker Recognition based on Lombard effect and small training/test data sizes, as well as accent and dialect classification of English, Arabic, and Spanish dialects, which have been supported by grants from various U.S. agencies. He has also served as a consultant, and held a U.S. clearance for more than ten years for work in speech/audio forensics.

Selected Publications:

J.H.L. Hansen, R. Huang, B. Zhou, M. Seadle, J.R. Deller, Jr., A.R. Gurijala, P. Angkititrakul, "SpeechFind: Advances in Spoken Document Retrieval for a National Gallery of the Spoken Word," *IEEE Trans. Speech & Audio Proc., Special Issue on Data Mining*, vol. 13, no. 5, pp. 712-730, Sept. 2005.

U. Yapanel, J.H.L. Hansen, "A New Perceptually Motivated MVDR-Based Acoustic Front-End (PMVDR) for Robust Automatic Speech Recognition, *Speech Communication*, vol. 50, pp. 142-152, Jan. 2008.

J.H.L. Hansen, V.S.Varadarajan, "Analysis and Normalization of Lombard Speech under different types and levels of noise with application to In-Set/Out-of-Set Speaker Recognition, *IEEE Trans. Audio, Speech & Language Processing*, vol. 17, no. 2, pp. 366-378, Feb. 2009

M. Akbacak, J.H.L. Hansen, "Environmental Sniffing: Noise Knowledge Estimation for Robust Speech Systems," *IEEE Trans. Audio, Speech and Language Processing*, vol. 15, no. 2, pp. 465-477, Feb. 2007.

R. Huang, J.H.L. Hansen, "Advances in Unsupervised Audio Classification and Segmentation for the Broadcast News and NGSW Corpora," *IEEE Trans. Audio, Speech and Language Processing*, vol. 14, no. 3, pp. 907-919, May 2006.

R. Huang, J.H.L. Hansen, P. Angkititrakul, "Dialect/Accent Classification using Unrestricted Audio," *IEEE Trans. on Audio, Speech and Language Processing*, vol. 15, no. 2, pp. 453-464, Feb. 2007.

G. Zhou, J.H.L. Hansen, and J.F. Kaiser, "Nonlinear Feature Based Classification of Speech under Stress," *IEEE Transactions on Speech & Audio Processing*, vol. 9, no. 2, pp. 201-216, March 2001

W. Kim, J.H.L. Hansen, "Time-Frequency Correlation Based Missing-Feature Reconstruction for Robust Speech Recognition in Band-Restricted Conditions," *IEEE Trans. Audio, Speech and Language Processing*, vol. 17, no. 7, pp. 1292-1304, Sept. 2009

J.H.L. Hansen, "Analysis and Compensation of Speech under Stress and Noise for Environmental Robustness in Speech Recognition," *Speech Communication*, Special Issue Speech Under Stress, 20(2):151-170, Nov. 1996.

J. Deller, J.H.L. Hansen, and J. Proakis (2000). Discrete-Time Processing of Speech Signals, (2nd Edition), IEEE Press / John Wiley Publishers, New York, NY.

N. Morales, D.T. Toledano, J.H.L. Hansen, J. Garrido, "Feature Compensation Techniques for ASR on Band-limited Speech," *IEEE Trans. Audio, Speech and Language Processing*, vol. 17, no. 4, pp. 758-774, May 2009.

N. Krishnamurthy, J.H.L. Hansen, "Babble Noise: Modeling, Analysis, and Applications," *IEEE Trans. Audio, Speech and Language Processing*, vol. 17, no. 7, pp. 1394-1407, Sept. 2009.

Y. Lei, J.H.L. Hansen, "Dialect Classification via Text-independent Training and Testing for Arabic, Spanish and Chinese," *IEEE Trans. Audio, Speech and Language Processing*, to appear, 2010. [Accepted Oct. 2009].

H. Boril, J.H.L. Hansen, "Unsupervised Equalization of Lombard Effect for Speech Recognition in Noisy Adverse Environments," IEEE Trans. Audio, Speech, and Language Processing, to appear, 2010. [Accepted Sept. 2009]

M. Akbacak, J.H.L. Hansen, "Spoken Proper Name Retrieval for Limited Resource Languages Using Multilingual Hybrid Representations," *IEEE Trans. Audio, Speech and Language Processing*, to appear, 2010. [Accepted Oct. 2009].

2.9 PROJECT MANAGEMENT AND INTERACTION PLAN

[redacted]

(b) (4)

Figure 12: SCENIC organizational chart

(b) (4)

[redacted]

performance. Integration, including the overall design, development, test, and evaluation of the system and the management of software deliverables from each technical task, will be the responsibility of SRI International and will be led from within the Software Engineering Program in SRI's Engineering Sciences Division.

Interdependencies

The seven defined project tasks are all highly interdependent. The three crosscutting tasks (Tasks 1-3), as well as the SAD/Metadata task (Task 4), benefit all others and are crucial for overall success. Tasks 5 and 6 have synergies but could be pursued independently. The software and team integration task (Task 7) is critical for all other work.

None of the subcontractors is entirely indispensable, but the loss of any one of them would seriously impact the overall outcome since they are expected to make key contributions to Task 1 (robust feature extraction) and Task 2 (advanced modeling), which are both critical to overall success, and, to a lesser extent, to Task 3 (SAD).

Managing Software Integration

In addition to the collaboration methods described above, we will support software collaboration and control by providing partner access to an integrated code base and repository on an SRI server. This will also support version control, code quality assurance, and a formal nightly release build. As team members will use each other's software during development, integration, test and evaluation, multiple users will have the opportunity to detect defects or integration incompatibilities as early as possible.

In addition to software components migrating from the SRI research team and partner sites to SRI's integration team, SRI will make common software components for SAD, LID, SID, and KWS available to the subcontractors to allow their research activities to be based on shared state-of-the-art software and ensure that research results can be readily integrated at SRI. We have successfully employed a similar methodology in past SRI-led multisite DARPA projects.

Risk Mitigation

We will continually work to identify, assess, and mitigate risks that may impact schedule, cost, and performance. Two key areas of risk are (1) failure to develop or integrate the necessary components, and (2) failure to deliver a system with real utility for users. Elements of our approach to mitigating these risks include:

- Wherever possible, using proven existing components as the starting point for SCENIC. SRI has built on, and will continue to build on, existing state-of-the-art software engines for the RATS program tasks.
- Performing diagnostic experiments with different types of degraded, as well as clean, data and/or replacing individual system modules with reference outputs, to test the utility and impact on an end-to-end system of the best achievable results. This will show us whether further development on a specific component or other improvement is worthwhile, before significant time is spent on research and implementation.
- Having more than one expert in a technical area (e.g., feature engineering) develop redundant approaches. If one approach proves systematically better, it will be adopted in the end-to-end system. If different approaches prove superior under different circumstances, we will research the best way to combine their outputs for maximum performance.
- Arranging for team members to use each other's software during development, integration, test and evaluation, so that multiple users will have the opportunity to detect defects or integration incompatibilities as early as possible.
- Engaging with DARPA and its contractor chosen for evaluation, providing feedback on software specifications prior to delivery and obtaining and acting on the feedback after system delivery.

Students and Postdoctoral Reseachers

The project will in part rely in the contributions of graduate students and postdoctoral researchers, via the three academic subcontractors: UT Dallas, UCLA, and ICSI.

UTD will employ two graduate students and two half-time postdocs. One postdoc will be responsible for development of restricted bandwidth-based feature and model compensation schemes for robust speech recognition and keyword spotting in noise and varying communications channels. This post-doc will concentrate on missing-feature-based compensation schemes. The second postdoc will focus on robust speech activity detection, and robustness with regard to speaker variability and competing background noise. One graduate student will work in collaboration with the first postdoc to develop novel automatic mask estimation schemes for missing-feature-based speech recognition in degraded channel conditions. The second graduate student will work with the second postdoc to help develop improved speaker and noise characterization schemes. These schemes will be evaluated in the context of speaker recognition, keyword/speech recognition, and language identification.

UCLA will employ one graduate student, who will be involved in the development and implementation of algorithms related to noise and channel robustness. Algorithms include variable frame rate, weighted Viterbi recognition, and error concealment techniques. The student will be conducting ASR in actual and simulated noise, and interacting with researchers at SRI, ICSI, and UTD to help deliver code and work on system integration. The student will also participate in PI meetings, if appropriate, and help write and present results at conferences and in technical reports.

ICSI will employ one graduate student, working on experiments incorporating ideas on channel and noise-robust spectral estimation and pitch detection, making use of ideas from Professors Morgan and Ellis (a senior researcher at ICSI).

Total Cost Summary Broken Down by Phase, by Task and Prime and Major Subcontractors [PHASE 1 (BASE)]

															GFY12					
	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 $
Task 1 Robust Feature Extraction																				
[redacted]																				
Task 3 - Decision Making																				
[redacted]																				
Task 4 Speech and Metadata Extraction																				
[redacted]																				
Task 5 - Language and Speaker Recognition																				
[redacted]																				
Task 6 - Keyword Spotting																				
[redacted]																				
Task 7 - Software and Team Integration																				
[redacted]																				

Note: In order to take full advantage of the available time to reach Phase 1 objectives, we plan to begin research and development with the start of Phase 1, while RATS data is still being collected. During this time, we will begin algorithm and software development, and carry out research using existing databases of degraded speech (such as the MIT-LL Tactical Speaker Identification Speech Corpus, the DARPA Spectrum Supremacy corpus, and various specialized databases at UT Dallas) and artificially degraded speech data. The level of effort during the first 6 months of Phase 1 will be limited compared to Months 7-18.

Total Cost Summary Broken Down by Phase, by Task and Prime and Major Subcontractors
[PHASE 2 (Option)]

	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 $
Task 1 Robust Feature Extraction														
[redacted]														
Task 2 - Advanced Modeling														
[redacted]														
Task 3 - Decision Making														
[redacted]														
Task 4 Speech and Metadata Extraction														
[redacted]														
Task 5 - Language and Speaker Recognition														
[redacted]														
Task 6 - Keyword Spotting														
[redacted]														
Task 7 - [illegible]														
[redacted]														

Total Cost Summary Broken Down by Phase, by Task and Prime and Major Subcontractors
[PHASE 3 (ROM)]

	[illegible]								[illegible]					
	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 $
Task 1 Robust Feature Extraction														
(b) (4) [redacted]														
Task 2 - Advanced Modeling														
(b) (4) [redacted]														
Task 3 - Decision Making														
(b) (4) [redacted]														
Task 4 Speech and Meta data Extraction														
(b) (4) [redacted]														
Task 5 - Language and Speaker Recognition														
(b) (4) [redacted]														
Task 6 - Keyword Spotting														
(b) (4) [redacted]														
Task 7 - Software and Team Integration														
(b) (4) [redacted]														

2.11 ORGANIZATIONAL CONFLICT OF INTEREST AFFIRMATIONS/DISCLOSURE

SRI International is not providing Scientific, Engineering, and Technical Assistance (SETA) to any DARPA technical offices.

SRI International participates in, and supports DARPA under the Intergovernmental Personnel Act (IPA) Mobility Program. SRI has IPA employees detailed or seconded as Program Managers in several offices within DARPA. We have listed those individuals below and SRI does not believe that any real or perceived conflict of interest exists, as these individuals are not directly involved with nor have access to any procurement-sensitive or budgetary information related to the effort proposed herein.

- Chris Earl, Office of the Director, DARPA
- Drew Dean, Transformational Convergence Technology Office (TCTO), DARPA

In addition, Mr. Dan Kaufman is an SRI IPA employee serving in the capacity of the Director of IPTO. It is our understanding that DARPA has policies and procedures that can be utilized to ensure that any conflict of interest, or appearance of conflict of interest, with respect to Mr. Kaufman and SRI can be addressed and avoided.

2.12 HUMAN USE

None.