Daniel P. W. Ellis - Curriculum Vitæ

office address:
Google, Inc.
111 8th Ave, 4th floor
New York, NY 10011

phone: (917) 617-0442
email: dan.ellis@gmail.com
web: https://ai.google/research/people/DanEllis/

Research interests:

- Signal processing and machine learning for analysis and classification of general audio, speech, and music.
- Detection and classification of nonspeech audio events.
- Large-scale audio datasets.
- Evaluation of audio classifiers.
- Audio source separation in underconstrained conditions.
- Computational models of human sound processing and organization.
- Automatic speech recognition in adverse environments.
- Visualization and browsing tools for audio and speech databases.

Positions held:

Staff Research Scientist, Google Inc, New York NY — 2015-present

Leading a team developing general purpose sound event classifiers, including releasing the AudioSet dataset: `http://g.co/audioset/` .

Professor of Electrical Engineering, Columbia University, New York NY — 2014-2015
Associate Professor of Electrical Engineering — 2005-2013
Assistant Professor of Electrical Engineering — 2000-2005

Leading LabROSA, the laboratory for the Recognition and Organization of Speech and Audio (`http://labrosa.ee.columbia.edu/`), investigating all aspects of intelligent sound processing from auditory scene analysis to robust speech recognition to music recommendation. Teaching includes introductory Digital Signal Processing and Speech and Audio Processing.

External Fellow, International Computer Science Institute, Berkeley CA — 2000-2014
Senior Research Scientist — 1998-2000
Postdoctoral Researcher — 1996-1998

Researching automatic speech recognition robustness with Prof. Nelson Morgan, including applications of perceptual models. Advised graduate students and managed grants.

Research Assistant, M.I.T. Media Lab, Cambridge MA — 1989-1996

Researching audio signal processing, sound synthesis and analysis under Prof. Barry L. Vercoe, while pursuing degrees at M.I.T.

Intern, Interval Research Corporation, Palo Alto CA — 1994

Developed sound analysis/synthesis techniques based on auditory models.

Technical staff, AWARE Inc., Cambridge MA — 1991-1993

Research and development of psychoacoustic-based compression schemes for high-quality audio including interactive tools. Contributed to MPEG audio 'committee code'.

Degrees:

Ph.D., Dept. of Elec. Eng. and Comp. Science, M.I.T. June 1996

Dissertation title: "A prediction-driven model of computational auditory scene analysis," advisor Prof. Barry L. Vercoe, readers Prof. Louis Braida, Dr. Ben Gold.

Master of Science, Dept. of Elec. Eng. and Comp. Science, M.I.T. February 1992

Thesis title: "A perceptual representation of audio," advisors Prof. Barry L. Vercoe and Dr. Thomas F. Quatieri.

Bachelor of Arts (honors), Dept. of Engineering, Cambridge University, U.K. June 1987

Electrical & Information Sciences Tripos. First class with distinction (ranked top). Final year project: A microcomputer-based linear predictive coding system for musical applications.

Projects at LabROSA:

Music Information Retrieval for Jazz Discography Jul 2012–Jun 2014
(Mellon Foundation, PI Szwed (Center for Jazz Studies), $250,000 to LabROSA)
Joint project between discographers at Columbia's Center for Jazz Studies and us to apply music information retrieval techniques to the analysis of jazz music recordings.

Spoken Wordsearch Rapid Development with Frugal Invariant Subword Hierarchies (SWORDFISH) Mar 2012–Feb 2017
(IARPA, PI Nelson Morgan, UC Berkeley; $750,000 to LabROSA)
5 site team to develop rapid-deployment robust speech recognition for low-resource languages.

Recognizing Abnormally Vocalized Speech Sep 2011–Aug 2015
(NSF, PI Ellis, $499,000)
Single-PI NSF project to apply speech recognition to unusual cases, such as singing.

Video Outlining and Learning through Unit Models and Events Mar 2011–Dec 2015
(IARPA, PI Smith (IBM), approx. $800,000 to LabROSA)
Large project between IBM and Columbia to develop automatic description and indexing of unconstrained web video.

Large-Scale Robust Cover Song Detection Sep 2010–Aug 2011
(PI Ellis, $68,000 gift from Google)
Project to develop and apply techniques for identifying "covers" of musical pieces applicable to large corpora such as YouTube.

Joint Audio-Visual Signatures for Web Video Analysis and Forensics Sep 2009–Aug 2012
(PI Ellis, $450,000 from National Geospatial Intelligence Agency, $225,000 to LabROSA)
PI on collaboration with Columbia video researchers to represent raw web-style videos in terms of audio-visual atoms learned from co-occurrence in large archives.

An Industrial-Scale Database for Music Audio Information Research Sep 2009–Aug 2011
(PI Ellis, $59,000 from NSF, GOALI collaboration with The Echo Nest)
Extension to NSF Music Data Mining project to develop a million-song database with metadata to provide a realistic, common database for Music Information research.

Indexing Consumer Video Sep 2007–Aug 2011
(PI Chang (Columbia), $120,000 from Eastman Kodak Co. to LabROSA)
Co-PI on project to combine audio and video information for content-based indexing and retrieval of short, home-made videos of the kind recorded by current generation consumer digital cameras.

Data-Driven Music Understanding (0713334) Sep 2007–Aug 2010
(PI Ellis, $450,000 from NSF IIS to LabROSA)
PI on collaboration with Columbia's Computer Music Center to analyze the high-level content of music audio with a view to discovering the latent structure that 'defines' music. Also includes education

outreach component to local middle and high schools.

Trustworthy Media: Media Forensics for Content Integrity Verification Sep 2007–Aug 2010
(PI Chang (Columbia); $350,000 from NSF; LabROSA portion $173,000)

Collaboration with Digital Video/Multimedia group on analyzing audio-video recordings to detect possible forgeries. Approaches focus on synchrony and correlation between audio and video elements, as well as 'device signatures' in the audio stream.

Separating Speech from Speech Noise Jan 2006–Dec 2009
(PI Ellis; $747,000 from NSF IIS; LabROSA portion $180,000)
PI on cross-disciplinary collaboration between engineers and psychologists seeking to apply signal separation in real-world conditions with demonstrable improvements for human listeners. Involves rotating students between engineering and psychology labs, including a collaborator in Paris.

Audio LifeLogs Jan 2006–Dec 2006
(PI Ellis; $50,000 from Microsoft Research)
Gift plus specialized equipment and software to develop techniques for gathering and indexing everyday personal recordings.

NIGHTINGALE Sep 2005 - Aug 2008
(PI Israel (SRI); $10,000,000 from DARPA GALE; LabROSA portion $150,000)
Large consortium addressing the Global Autonomous Language Exploitation requirements, spanning speech recognition, machine translation, information extraction, access and presentation. LabROSA will collaborate with Hirschberg in CS to identify prosodic factors to support translation and extraction e.g. phrase segments, stressed words.

Music Engineering Art Collaboration Oct 2004 - Aug 2007
(PI Ellis; $332,319 from Columbia Academic Quality Fund; LabROSA portion $173,970)
Interdisciplinary collaboration with Columbia Computer Music Center to develop music signal analysis and organization tools, teach a mixed engineering/art projects class, and establish a more permanent entity, using multimedia art installations as a motivating goal.

Music Similarity Modeling Sep 2003 - Aug 2005
(PI Ellis; $63,494 from Google, Inc.)
Sponsorship of one student looking at automatic analysis of music signals to predict listener preference, with the goal of developing intelligent music browsing and recommendation for little-known music.

The Listening Machine: Sound Organization for Multimedia Feb 2003 - Jan 2009
(PI Ellis; $558,207 from NSF Career program and Columbia SEAS.)
Project to develop intelligent analysis of real-world sound mixtures in terms of their component sources using techniques drawn from speech recognition, machine learning, and human auditory modeling. Current results include a novel analysis, the "spectral deformation model", with applications in speech recognition and source separation.

Pushing the Envelope: Rethinking Speech Signal Representation Apr 2002 - Aug 2005
(PI Morgan; Ellis co-PI; $6,000,000 from DARPA EARS; LabROSA portion $729,430)
Large consortium to develop "novel approaches" to dislodge speech recognition from its diminishing incremental returns. Our contribution has been a novel model for temporal energy modulations in the speech signal (based on the dual of linear prediction), and current work includes methods to transform informal, hasty speech into more careful pronunciation.

Audio Signal Analysis Oct 2001 - Aug 2003
(PI Ellis; $104,626 from NEC Research Lab)

Sponsorship of one student working on models for music similarity and developing and evaluating a browsing tool to navigate large music collections based solely on signal properties rather than manually-provided metadata.

Mapping Meetings: Language Technology for Natural Interactions Sep 2001 - May 2005
(PI Morgan; Ellis co-PI; $1,402,851 from NSF ITR program; LabROSA portion $125,674)
Project to analyze a large corpus of natural meetings resulting from a project I set up while at ICSI. Our work included a variety of techniques to analyze these recordings, including locating interesting events (stressed utterances, laughter), recovering speaker turns from multiple distant mic channels, and inferring 'talkativity" models to characterize each participant.

Teaching:

Spring 2014:

Music Signal Processing (ELEN E4896, 12 students, overall student rating 4.8/5)
Latest presentation of this course developed in 2010 that uses aspects and applications of music audio processing to introduce, motivate, and illustrate core signal processing concepts such as sampling, Fourier domain, and filtering, as well as presenting more advanced music-specific techniques such as time scaling and transcription. Emphasis on hands-on learning with weekly practical sessions based around developing real-time implementations of algorithms being discussed. This year, we experimented with "flipped classroom" techniques by replacing traditional in-class lectures with videos recorded in 2013. All the materials for this course, including videos of lectures, freely available at `http://www.ee.columbia.edu/~dpwe/e4896/`.

Reproducing Computational Research (EECS E6891, 4 students, overall student rating 4.7/5)
For the second iteration of this experimental course in which students are tasked to exactly reproduce the results in a computational paper of their choice, I enlisted my post-doc Brian McFee to co-teach, and to share his excellent and systematic workflows for reproducible computational research. We also had a several guest lecturers, some via teleconferencing. `http://www.ee.columbia.edu/~dpwe/e6891/`.

Fall 2013:

Digital Signal Processing (ELEN E4810, 99 students, overall student rating 4.3/5)
Fundamental senior/masters level class emphasizing a combined mathematical and intuitive understanding of the Fourier domain, and providing tools for filter design. Extensive use of live in-class demos; 30% of grade on individual final projects. This class was 50% larger than typical, but worked out very successfully, with more lively and interactive sessions. All notes, demos, lecture videos, and other materials are available online: `http://www.ee.columbia.edu/~dpwe/e4810/`.

Spring 2013:

Music Signal Processing (ELEN E4896, 13 students, overall student rating 4.2/5)
Continued refinement of this hands-on course included introducing a "sign-off" procedure to ensure that each student effectively completed each of the in-class practical sessions.

Reproducing Computational Research (EECS E6891, 7 students, overall student rating 3.2/5)
First instance I am co-teaching this class with Victoria Stodden of the Statistics department as part of the "Data to Solutions" IGERT that began this year. The class discusses issues in achieving full reproducibility in computational research. Each student must exactly reproduce a published computational result for their course project (or come as close as they can!).

Fall 2012:

Digital Signal Processing (ELEN E4810, 64 students, overall student rating 3.8/5)
Improved the quality of my self-produced online videos of lectures to give students the choice of coming to class in-person, or watching online.

Spring 2012:

Music Signal Processing (ELEN E4896, 12 students, overall student rating 4.3/5)
This was the third time I offered this course, and included several new modules. I also tried to address problems students had encountered learning the Pure Data audio processing environment by starting with it earlier in the semester, and covering it in more detail.

Fall 2010:

Digital Signal Processing (ELEN E4810, 65 students, overall student rating 4.1/5)
Continued practice of publishing complete course materials online, including videos of lectures.

Spring 2011:

Music Signal Processing (ELEN E4896, 25 students, overall student rating 4.2/5)
This was the second time I offered this course, and included several new modules.

Fall 2010:

Digital Signal Processing (ELEN E4810, 65 students, overall student rating 4.1/5)
Continued practice of publishing complete course materials online, including videos of lectures.

Spring 2010:

Music Signal Processing (ELEN E4896, 20 students, overall student rating 4.0/5)
The first presentation of this course using music signals as a way to investigate signal processing, with emphasis on practical, real-time experimentation.

Spring 2009:

Speech and Audio Processing and Recognition (ELEN E6820, 6 students, overall student rating 4.3/5)
Latest version of my novel course straddling the LabROSA research interests. Weekly meetings include some traditional lecturing, paper presentation by student participants, and free-discussion 'brainstorming' on an example practical problem relevant to that week's topic. 50% of grade from semester project; part of grade comes from peer assessment of project presentations. All the materials for this course are freely available at `http://www.ee.columbia.edu/~dpwe/e6820/`; I have had numerous messages of thanks from many sources who have downloaded them.

Fall 2008:

Digital Signal Processing (ELEN E4810, 59 students, overall student rating 3.8/5)
Various innovations including creating & posting videos of all lectures to aid students with revision.

Spring 2008:

Speech and Audio Processing and Recognition (ELEN E6820, 9 students, overall student rating 4.2/5)
Further experimentation with shifting the classroom sessions towards more student-led activities.

Fall 2007:

Digital Signal Processing (ELEN E4810, 67 students, overall student rating 4.4/5)
Continued refinement of this core signal processing class; significant jump in student ratings.

Music Engineering Art Project (8 students)
This project class is our ongoing activity within the EMAC project described above. With an even balance of engineering and music students, all of whom are passionate about both technology and music, we are developing interactive audio/music analysis/synthesis systems, including MEAPsoft. `http://labrosa.ee.columbia.edu/meapsoft/`

Spring 2007:

Speech and Audio Processing and Recognition (ELEN E6820, 15 students, overall student rating 4.3/5)
Relatively large enrollment with several extremely good projects; several went on to be published.

Music Engineering Art Project (10 students)
Spent a lot of time discussing the user interface to MEAPsoft, what core tasks it is or could be used for, and ideas for how to improve it; one student was simultaneously taking a user interface class in CS.

Music Signal Processing (ELEN E4998/E4896, 12 students)
By popular demand, we ran a version of this class originally created by Prof. Eleftheriadis who was on sabbatical. Students presented papers on digital audio music effects algorithms, analyzed from an engineering perspective, and demonstrated and discussed practical implementation. Final project involved applying specially-developed effects algorithm to a real musical multitrack mixing task.

Fall 2006:

Digital Signal Processing (ELEN E4810, 50 students, overall rating 3.8/5)
Improved use of online tools such as discussion boards and online videos to help connect and empower students.

Music Engineering Art Project (7 students)
This project class is our ongoing activity within the EMAC project described above. With an even balance of technologically-oriented engineering and music students, we are developing interactive audio/music analysis/synthesis systems. We are planning an interactive art installation on the Columbia campus using our MEAPsoft software (see below).

Spring 2006:

Speech and Audio Processing and Recognition (ELEN E6820, 10 students, overall student rating 4.8/5)
With a smaller than average enrollment, ww experimented with a much more interactive classroom for mat, devoting about half of each week to discussion and brainstorming.

Music Engineering Art Project (10 students)
After one year of meeting for this project, we finally got down to building something - the experimental audio/music reorganization package, MEAPsoft version 1.0, was released over the summer. `http://labrosa.ee.columbia.edu/meapsoft/`

Fall 2005:

Digital Signal Processing (ELEN E4810, 68 students, overall student rating 3.4/5)
Continuing revisions to content and ordering to improve intelligibility.

Music Engineering Art Project (9 students)
Follow-on project class incorporated more structure, with each student (and both faculty!) responsible for making a presentation one week during the semester.

Spring 2005:

Speech and Audio Processing and Recognition (ELEN E6820, 20 students; rating 4.4/5)
Added more material on signal separation and large audio database analysis based on our recent research work. Half the grade comes from a final project, which amounts to a mini independent semester project with each student.

Music Engineering Art Project (8 students)
Our first activity within the EMAC project. Initial investigation into practicality and issues of running a class made up of both engineering and music graduate students.

Machine Learning Signal Processing Seminar (5-15 students)
I manage this weekly informal reading group for the Ph.D. students in the signal processing area. This semester we looked at neurophysiological underpinnings of perception and some other novel machine learning techniques.

Fall 2004:

Digital Signal Processing (ELEN E4810, 50 students, overall student rating 3.5/5)
Fewer students than average; some sessions extensively revised to improve flow of ideas.

Machine Learning Signal Processing Seminar (5-15 students)
In-depth coverage of normalized cuts and spectral clustering theory and applications.

Spring 2004:

Speech and Audio Processing and Recognition (ELEN E6820, 9 students, rating 4.5/5)
Since it was offered the same semester as a new adjunct class on speech recognition, I added new units on signal separation and musical signal analysis to replace the earlier detail on speech recognition.

Machine Learning Signal Processing Seminar (ELEN E9701, 8 registered students)
Most of the semester was spent working through the Scholkpf/Smola book on kernel methods. We ran the reading group as a formal course number with registration as an experiment.

Fall 2003:

Digital Signal Processing (ELEN E4810, 72 students, rating 3.5/5)
Largest ever enrollment; now a required course for the Biomedical Engineering Imaging MS.

Spring 2003:

Speech and Audio Processing and Recognition (ELEN E6820, 23 students, rating 4.0/5)
One senior undergrad went on to publish his final project with me at an IEEE conference.

Musical Content Analysis by Machine Learning (9 hour short course, 20 students)
A one-week seminar taught by invitation at the Music Technology Group of Barcelona's Pompeu Fabra University (UPF), combining pattern recognition background with applications in music analysis. Also a one-day session at the Johns Hopkins University Center for Spoken Language Processing Summer School, for which I created a self-paced practical. Materials available at: `http://www.ee.columbia.edu/~dpwe/muscontent/`

Fall 2002:

Digital Signal Processing (ELEN E4810, 58 students, rating 4.0/5)
Moved entire course over to Powerpoint after complaints about my handwriting.

Spring 2002:

Speech and Audio Processing and Recognition (ELEN E6820, 12 students, rating 4.3/5)
Rearranged material and added a new section on spatial audio.

Fall 2001:

Digital Signal Processing (ELEN E4810, 42 students, rating 3.8/5)
Course open to distance students (watching videos of lectures) for the first time.

Spring 2001:

Speech and Audio Processing and Recognition (ELEN E6820, 11 students, rating 3.9/5)
First offering of this new course which I based on my research interests.

Fall 2000:

Digital Signal Processing (ELEN E4810, 41 students, rating 3.6/5)
My first time teaching a full course; it took me a while to get the pacing right.

Other teaching:

Supervision and advising of UC Berkeley graduate students at ICSI, 1997-2001.

Tutor, MIT OME Tutorial Services Room, working with individual students on probability and signal processing, 1993-95.

Occasional guest lecturer for courses Digital Audio Signal Processing, and Introduction to Perception (Auditory Organization), MIT, 1991-1995.

Students supervised:

Completed Ph.D. students (Columbia):

Zhuo Chen, Ph.D., graduated Jan 2017, researcher at Microsoft.

Colin Raffel, Ph.D., graduated June 2016, researcher at Google.

Dawen Liang, Ph.D., graduated June 2016, researcher at Netflix.

Thierry Bertin-Mahieux, Ph.D., graduated Jan 2013, researcher at Google.

Courtenay Cotton, Ph.D., graduated Jan 2013, developer at Museami, Inc.

Byung-Suk Lee Ph.D., graduated Sep 2012, analyst at Boston Consulting Group, Korea.

Graham Grindlay, Ph.D. graduated Jan 2011.

Christine Smit, Ph.D., graduated Jan 2011, engineer at Telophase Corporation.

Michael Mandel, Ph.D., graduated Sep 2009, research scientist at Ohio State University.

Keansub Lee, Ph.D., graduated May 2009, researcher at Samsung.

Ron Weiss, Ph.D., graduated May 2009, researcher at Google.

Xanadu Halkias, Ph.D., graduated Oct 2008, researcher at Univ. Sud Toulon, France.

Graham Poliner, Ph.D., graduated May 2008, financial analyst with Kurt Salmon.

Marios Athineos, Ph.D., graduated Sep 2007, researcher at Sennheiser, Inc.

Adam Berenzweig, Ph.D., graduated Sep 2006, researcher at Google.

Manuel Reyes, Ph.D., graduated Apr 2005, researcher at Microsoft Research.

Completed Ph.D. students (other schools):

Brian Whitman, Ph.D. from MIT Media Lab Mar 2005, who I effectively co-supervised with my former Ph.D. supervisor from MIT, Barry Vercoe, currently running a startup.

Patricia Scanlon, Ph.D. from University College Dublin Sep 2005, spent 18 months at LabROSA from 2002-2003 with me as her local advisor, currently working for Lucent in Ireland.

Visitors (stays of 1-12 months): Hiroyuki Satoh (Univ. Tokyo), Diego Silva (University of Sao Paolo, Brazil), Matt McVicar (Bristol University, UK), Dr. Hélène Papadopoulos (CNRS, France), Prof. Sofia Cavaco (New University of Lisbon, Portugal), Bjørn Sand Jensen (DTU, Denmark), Diego Castàn (U. Zaragoza, Spain), Prof. Jon Barker (U. Sheffield, UK), Prof. Edward Jones (National Univ. Ireland, Galway), Jon Gudnason (U. Rekjavik, Iceland), Jesper Boldt (Oticon, Denmark), Mads Christensen (Aalborg U., Denmark), Mikkel Schmidt (DTU, Denmark), Jesper Jensen (Aalborg U., Denmark), Jouni Paulus (TUT Finland), Kofi Boakye (UC Berkeley), Elvira Perez (U. Liverpool), Thomas Blumensath (Queen Mary, U. London), Sunil Sivadas (OGI).

MS students: Lingyu Zhang, Yuan Gao, Xiaohong Zeng, Michael Groble, Jim Ogle, Nathan Lesser, John Arroyo, Sam Keene, Jack Chan, Alex Sheh, Joseph Hazboun, Banky Omodunbi, Lyndon Kennedy, Rob Turetsky, Uday Arya, Chi Wong, Wayzen Lin, Johanna Devaney (Music).

Undergraduates: Hilary Mogul, James Thompson, Adrienne Humblet, Sean Healy, Phill Snyder, Vishal Kumar, Jerry Liu, Suman Ravuri, Clement Tong, Rob Spinella, David Wilmot, Ezra Schneck.

High school students: Juliette Kim, Ali Rehmatullah, Jeff Bauer, Ben Chang, Angel Umpierre.

I have also served on, or am currently serving on, the following Ph.D. committees:

Spencer Russell (Paradiso, MIT), Anurag Kumar (Raj, CMU), Sourish Chaudhuri (Raj, CMU), Katherine Kinnaid (Pauls, Dartmouth), Sarah Angelini (Maul, APAM), Erinc Tokluoglu (Sen), Pasi Saari (Eerola, U. Jyväskylä, Finland), Anna Choromaska (Jebara, CS), Taemin Cho (Bello, NYU), Junfeng He (Chang), Sourish Chaudhuri (Raj, CMU), Edwin Ahn (Longman, ME), Mandis Beigi (Chang), Erik Schmidt (Drexel Univ.), Wei Liu (Chang), Kevin Xu (Longman, ME), Lise Regnier (Peeters, U. Pierre et Marie Curie, Paris), Sriram Ganapathy (Hermansky, Johns Hopkins), Hari Parthasarahti (Bourlard, EPFL Switzerland), Wei Jiang (Chang), Yao Li (Longman, ME), Lingyun Gu (Stern, CMU), Laurent Oudre (Fevotte, Telecom Paris), Fadi Biadsy (Hirschberg, CS), Wei Jiang (Chang), Jean-Louis Durrieu (Richard, Telecom Paris), Andrew Rosenberg (Hirschberg, CS), Frank Enos (Hirschberg, CS), Jonathan Le Roux (de Cheveigne/Sagayama, ENS France), Michel Galley (McKeown, CS), Jackson Lipscombe (Hirschberg, CS), Kai Li (Wang), Raju Hormis (Wang), Boonsit Yimwadsana (Coffman), Winston Hsu (Chang), Lexing Xie (Chang), Dongqing Zhang (Chang), Ana Benitez (Chang), Di Zhong (Chang), Danny Hong (Eleftheriadis), Lai-Tee Cheok (Eleftheriadis), Ari Klein (Tsividis), Aya Aner (Kender, CS), Alex Klein (Pederson, App. Phys.), Lisa Margulis (Lerdahl, Music), Peter LeVoci (Longman, ME), Scott Otterson (Ostendorf, U. Washington), Mitch Parry (Essa, Georgia Tech), Jitendra Ajmera (Bourlard, IDIAP), Sofia Cavaco (Lewicki, CMU), Tuomas Virtanen (TUT Finland), Olivier Gillet (ENST France), Nayeemulla Khan (Yegnanarayanan, IIT Madras), Andrij Temko (Nadeau, UPC Barcelona).

Distinctions/Awards:

Elevated to IEEE Fellow, January 2018, for contributions to speech, audio, and music processing.

Best student paper, ISMIR 2013 (with Dawen Liang).

NAE "Frontiers of Engineering" participant/session organizer, 2009 and 2010.

NSF Career Award, 2003-2008.

Speech Communication Best Paper Award for 2005-2007 (one paper chosen from around 200 published in Speech Communication over two years).

Speech Communication Best Paper Award for 2001-2003

Co-developer of best-ranked systems for Audio Chord Recognition (2009), Audio Cover Song Detection (2006), and Audio Artist Identification (2004, 2005) at MIREX Evaluations of the International Conference on Music Information Retrieval (among many teams including Stanford University, Fraunhofer IDMT, University of Montreal, Austrian Research Institute for AI, HP Labs, and Sun Microsystems).

Co-developer of best-performing system in the ETSI Aurora-2 Noisy Speech Recognition evaluation at the Eurospeech Special Event, 2001 (among 17 teams including IBM, Microsoft, Bell Labs, and UCLA).

Journal papers:

(note: **citation counts** are from scholar.google.com, updated 2020-01-27, and shown when ≥ 60.
Overall Google-based h-index including conference papers: **69**.)

1. R. Oliver, D. Ellis, H. Chmura, J. Krause, J. P'erez, S. Sweet, L. Gough, J. Wingfield, N. Boelman, "Eavesdropping on the Arctic: Automated bioacoustics reveal dynamics in songbird breeding phenology," Science Advances 4(6), 2018.
2. S. Popham, D. Boebinger, D. Ellis, H. Kawahara, J. McDermott, "Inharmonic speech reveals the role of harmonicity in the cocktail party problem," Nature communications 9 (1), 2018.
3. D. Silva, V. Souza, D. Ellis, E. Keogh, G. Batista, "Exploring low cost laser sensors to identify flying insect species," J. Intelligent & Robotic Systems 80 (1), 313–330, 2015.
4. J. Salamon, E. Gómez, D. Ellis, G. Richard, "Melody Extraction from Polyphonic Music Signals," IEEE Sig. Proc. Magazine, 118–134, March 2014. DOI: 10.1109/MSP.2013.2271648 **(Cited by 178)**
5. J. Devaney, M. Mandel, D. Ellis, I. Fujinaga, "Automatically extracting performance data from recordings of trained singers," Psychomusicology: Music, Mind & Brain 21(1–2), 108–136, 2011.
6. G. Grindlay & D. Ellis, "Transcribing Multi-instrument Polyphonic Music with Hierarchical Eigeninstruments," IEEE J. Sel. Topics Sig. Process., 5(6), 1159–1169, Oct 2011. **(Cited by 77)**
7. M. Müller, D. Ellis, A. Klapuri, & G. Richard, "Signal Processing for Music Analysis," IEEE J. Sel. Topics Sig. Process., 5(6), 1088–1110, Oct 2011. **(Cited by 262)**
8. M. Müller, D. Ellis, A. Klapuri, G. Richard, & S. Sagayama, "Introduction to the Special Issue on Music Signal Processing," IEEE J. Sel. Topics Sig. Process., 5(6), 1085–1087, Oct 2011.
9. R. Weiss, M. Mandel, & D. Ellis, "Combining Localization Cues and Source Model Constraints for Binaural Source Separation," Speech Communication, 53(5), 606–621, May 2011.

10. M. Mandel, S. Bressler, B. Shinn-Cunningham, & D. Ellis, “Evaluating Source Separation Algorithms With Reverberant Speech,” IEEE Tr. Audio, Speech, & Lang. Proc., 18(7), 1872–1883, Sep 2010. DOI: 10.1109/TASL.2010.2052252
11. K. Lee & D. Ellis, “Audio-Based Semantic Concept Classification for Consumer Video,” IEEE Tr. Audio, Speech, & Lang. Proc., 18(6), 1406–1416, Aug 2010. DOI: 10.1109/TASL.2009.2034776
12. M. Mandel, R. Weiss, & D. Ellis, “Model-Based Expectation-Maximization Source Separation and Localization,” IEEE Tr. Audio, Speech, & Lang. Proc., 18(2), 382–394, Feb 2010. **(Cited by 282)**
13. R. Weiss & D. Ellis, “Speech separation using speaker-adapted Eigenvoice speech models,” Computer Speech & Language, 24(1), 16–29, Jan 2010. **(Cited by 76)**
14. J. H. Jensen, M. G. Christensen, D. Ellis, & S. H. Jensen, “Quantitative analysis of a common audio similarity measure,” IEEE Tr. Audio, Speech & Lang. Proc., 17(4), 693–703, May 2009.
15. M. Mandel & D. Ellis, “A Web-Based Game for Collecting Music Metadata,” J. New Music Research, 37(2), 151–165, 2008. **(Cited by 204)**
16. T. S. Huang, C. K. Dagli, S. Rajaram, E. Y. Chang, M. I. Mandel, G. E. Poliner, & D. Ellis, “Active learning for interactive multimedia retrieval,” Proc. IEEE, 96(4), 648-667, 2008. **(Cited by 123)**
17. J. Devaney & D. Ellis, “An Empirical Approach to Studying Intonation Tendencies in Polyphonic Vocal Performances,” J. Interdisc. Music Studies, 2(1–2), 141–156, Spring 2008.
18. M. Slaney, D. Ellis, M. Sandler, M. Goto, & M. M. Goodwin, “Introduction to the Special Issue on Music Information Retrieval,” IEEE Tr. Audio, Speech & Lang. Proc., 16(2), 253–254, Feb 2008.
19. M. Athineos & D. Ellis. “Autoregressive modeling of temporal envelopes,” IEEE Tr. Sig. Proc., 15(11), 5237-5245, Nov 2007. **(Cited by 105)**
20. G. Poliner, D. Ellis, A. Ehmann, E. Gomez, S. Streich, & B. Ong (2007) “Melody Transcription from Music Audio: Approaches and Evaluation,” IEEE Tr. Audio, Speech & Lang. Proc., 15(4), 1247-1256, May 2007. **(Cited by 226)**
21. D. Ellis. “Beat Tracking by Dynamic Programming,” J. New Music Research, Special Issue on Algorithms for Beat Tracking and Tempo Extraction, 36(1), 51–60, Mar 2007. **(Cited by 483)**
22. P. Scanlon, D. Ellis, & R. Reilly “Using Broad Phonetic Group Experts for Improved Speech Recognition,” IEEE Tr. Audio, Speech & Lang. Proc. 15(3), 803-812, Mar 2007.
23. G. Poliner & D. Ellis “A Discriminative Model for Polyphonic Piano Transcription,” EURASIP Signal Processing Journal, 2007 (2007), Article ID 48317, 9 pages. **(Cited by 284)**
24. D. Ellis & K.-S. Lee “Accessing minimal-impact personal audio archives,” IEEE Multimedia 13(4), 30-38, Oct-Dec 2006.
25. D. Ellis & G. Poliner “Classification-based Melody Transcription,” Machine Learning 65(2-3), 439-456, Dec 2006. **(Cited by 65)**
26. M. Mandel, G. Poliner, & D. Ellis “Support Vector Machine Active Learning for Music Retrieval,” Multimedia Systems Journal 12(1), 3-13, Aug 2006. **(Cited by 175)**
27. D. Ellis “Extracting Information from Music Audio”, Communications of the ACM 49(8), 32-37, Aug 2006.
28. X. Halkias & D. Ellis “Call detection and extraction using Bayesian inference,” Applied Acoustics, 67(11-12), 1164-1174, Nov-Dec 2006.
29. D. Ellis, B. Raj, J. Brown, M. Slaney, & P. Smaragdis “Editorial - Special Section on Statistical and Perceptual Audio Processing,” IEEE Tr. Audio, Speech & Language Proc., 14(1), 2-4, Jan. 2006.

30. N. Morgan, Q. Zhu, A. Stolcke, K. Sönmez, S. Sivadas, T. Shinozaki, M. Ostendorf, P. Jain, H. Hermansky, D. Ellis, G. Doddington, B. Chen, Ö. Çetin, H. Bourlard, & M. Athineos "Pushing the envelope – aside," IEEE Signal Processing 22(5), 81-88, Sep. 2005. **(Cited by 114)**
31. J. Barker, M. Cooke, & D. Ellis "Decoding speech in the presence of other sources," Speech Communication 45(1), 5-25, Jan 2005 (winner of Speech Communication's **Best Paper Award** 2005-2007). **(Cited by 178)**
32. A. Berenzweig, B. Logan, D. Ellis, & B. Whitman "A Large-Scale Evaluation of Acoustic and Subjective Music-Similarity Measures," Computer Music Journal (MIT Press) 28(2), 63-76, June 2004. **(Cited by 293)**
33. A. Robinson, G. Cook, D. Ellis, E. Fosler-Lussier, S. Renals, & G. Williams "Connectionist speech recognition of Broadcast News," Speech Communication 37(1-2), 27-45, 2002. **(Cited by 61)**
34. M. Cooke & D. Ellis "The auditory organization of speech in listeners and machines," Speech Communication 35(3-4), 141-177, 2001 (winner of Speech Communication's **Best Paper Award** 2001-2003). **(Cited by 203)**
35. D. Ellis "Using knowledge to organize sound: The prediction-driven approach to computational auditory scene analysis, and its application to speech/nonspeech mixtures," Speech Communication 27(3-4), 281-298, 1999. **(Cited by 75)**

Theses:

1. D. Ellis "Prediction-driven Computational Auditory Scene Analysis", Ph.D. dissertation, Dept. of Elec. Eng. and Comp. Sci. MIT, April 1996. **(Cited by 483)**
2. D. Ellis "A Perceptual Representation of Audio", M.S. Thesis, Dept. of Elec. Eng. and Comp. Sci. MIT, Feb. 1992.

Books / bookchapters:

1. T. Virtanen, M. D. Plumbley, D. Ellis (eds.) "Computational Analysis of Sound Scenes and Events," Springer, 2018. **(Cited by 60)**
2. B. Gold, N. Morgan, D. Ellis "Speech and Audio Signal Processing", second edition, J. Wiley, 2011. **(Cited by 1098)**
3. D. Ellis "An introduction to signal processing for speech", chap. 20 in: W. J. Hardcastle, J. Laver & F. E. Gibbon (eds.) *The Handbook of Phonetic Sciences, 2nd Ed.* (Blackwell), pp. 757-780, 2010.
4. D. Ellis "Model-Based Scene Analysis", chap. 4 in: D. Wang & G. Brown (eds.) *Computational Auditory Scene Analysis: Principles, Algorithms, and Applications* (Wiley/IEEE Press), pp. 115-146, 2006.
5. D. Ellis "Modeling the auditory organization of speech," chap. 24 in: S. Greenberg & W. Ainsworth (eds.) *Listening to speech: An auditory perspective* (Lawrence Erlbaum), pp. 393-407, 2006.
6. D. Ellis "Evaluating Speech Separation Systems", in: P. Divenyi (ed.) *Speech Separation by Humans and Machines* (Kluwer), pp. 295-304, 2004.
7. D. Ellis & D. Rosenthal "Mid-Level representations for Computational Auditory Scene Analysis," in: D. Rosenthal & H. Okuno (eds.) *Computational Auditory Scene Analysis* (Lawrence Erlbaum, Mahwah), pp. 257-272, 1998.

Software/Online resources:

(all available via `http://www.ee.columbia.edu/~dpwe/`).

- **AudioSet - A sound vocabulary and large-scale dataset:** An effort to provide a sufficiently large and general dataset for environmental sound events to provide a research stimulus comparable to the ImageNet object recognition task. Comprises a sound event ontology of over 600 items (the first attempt to provide complete coverage of sound events) along with pointers to over 2 million 10 second YouTube excerpts manually verified as representing the events.
`https://research.google.com/audioset/`
`https://github.com/audioset`
- **AUDFPRINT - Audio fingerprint database creation and query:** A ready-to-run open-source implementation of noise- and channel-robust audio fingerprint matching, capable of matching within databases of tens of thousands of tracks. **(Cited by 61)**
`http://www.ee.columbia.edu/~dpwe/resources/matlab/audfprint/`
`https://github.com/dpwe/audfprint`
- **Dynamic time warp (DTW) in Matlab:** Highly efficient Matlab implementatio of this difficult-to-vectorize operation. **(Cited by 79)**
`http://www.ee.columbia.edu/~dpwe/resources/matlab/dtw/`
- **PLP and RASTA (and MFCC, and inversion) in Matlab:** Easily-understood and flexible Matlab code to calculate this popular family of audio features. **(Cited by 193)**
`http://www.ee.columbia.edu/~dpwe/resources/matlab/rastamat/`
- **Audio processing examples in Matlab** including the Phase Vocoder, Dynamic Time Warping, and Sinusoidal Analysis/Synthesis. Many of these grew out of examples used in my classes. Posting these materials online has led to their widespread use, as measured by the hundreds of email messages I have received regarding them.
`http://www.ee.columbia.edu/~dpwe/resources/matlab/`
- **Online Course Materials** including complete slidepacks for the Speech and Audio Processing and Recognition course. These have been used at multiple other universities including MIT, U. Montreal, Washington University (St. Louis), U. Mississippi, U. Porto (Portugal).
`http://www.ee.columbia.edu/~dpwe/e6820/`
- **Music Similarity Ground Truth** data, distilled from projects and online surveys we have conducted in this area. This data was used as the basis for the first international evaluation of music similarity algorithms at the International Symposium on Music Information Retrieval in October 2004. We also created “uspop2002”, a standard database of basic features for 8700 pop music tracks, for researchers getting started in Music IR; we mailed out 42 copies of this 3 DVD set before switching to online distribution via Amazon S3 in 2009.
`http://labrosa.ee.columbia.edu/projects/musicsim/uspop2002.html`
- The **AUDITORY list** and homepage – archives and related materials for this research discussion list I have operated since 1992, currently with 2216 subscribers in 33 countries, and about 70 postings a month. `http://www.auditory.org/`

Peer-reviewed proceedings papers:

(Notes: **citation counts** are from scholar.google.com, updated 2014-04-08, and shown when ≥ 40.)

1. A. Jansen, D. Ellis, S. Hershey, R. C. Moore, M. Plakal, A. Popat, R. A. Saurous (2020) “Coincidence, categorization, and consolidation: Learning to recognize sounds with minimal supervision,” Proc. ICASSP, (accepted), Barcelona, Spain, May 2020.

2. Q. Huang, A. Jansen, L. Zhang, D. Ellis, R. A. Saurous, J. Anderson (2020) "Large-scale weakly-supervised content embeddings for music recommendation and tagging," Proc. ICASSP, (accepted), Barcelona, Spain, May 2020.
3. E. Tzinis, S. Wisdom, J. R. Hershey, A. Jansen, D. Ellis (2020) "Improving universal sound separation using sound classification," Proc. ICASSP, (accepted), Barcelona, Spain, May 2020.
4. E. Fonseca, M. Plakal, F. Font, D. Ellis, X. Serra (2020) "Audio tagging with noisy labels and minimal supervision," Detection and Classification of Acoustic Scenes and Events (DCASE) Workshop, New York, USA, October 2019.
5. E. Fonseca, M. Plakal, D. Ellis, F. Font, X. Favory, X. Serra (2019) "Learning sound event classifiers from web audio with noisy labels," Proc. ICASSP, 21–25, Brighton, UK, 2019.
6. E. Fonseca, M. Plakal, F. Font, D. P. W. Ellis, X. Favory, J. Pons, X. Serra (2018) "General-purpose Tagging of Freesound Audio with AudioSet Labels: Task Description, Dataset, and Baseline ," Detection and Classification of Acoustic Scenes and Events (DCASE) Workshop, Surrey, UK, 2018.
7. A. Jansen, M. Plakal, R. Pandya, D. Ellis, S. Hershey, J. Liu, R. Moore, R. Saurous (2018) "Unsupervised learning of semantic audio representations," Proc. ICASSP, 126–130, Calgary, Canada, 2018.
8. J. Gemmeke, D. Ellis, D. Freedman, A. Jansen, W. Lawrence, R. Moore, M. Plakal, M. Ritter (2017) "Audio set: An ontology and human-labeled dataset for audio events," Proc. ICASSP, 776–780, New Orleans USA, 2017. **(Cited by 486)**
9. A. Jansen, J. Gemmeke, D. Ellis, X. Liu, W. Lawrence, D. Freedman (2017) "Large-scale audio event discovery in one million youtube videos," Proc. ICASSP, 786–790, New Orleans USA, 2017.
10. S. Hershey, S. Chaudhuri, D. Ellis, J. Gemmeke, A. Jansen, R. Moore, M. Plakal, D. Platt, R. Saurous, B. Seybold, M. Slaney, R. Weiss, K. Wilson (2017) "CNN architectures for large-scale audio classification," Proc. ICASSP, 131–135, New Orleans USA, 2017. **(Cited by 532)**
11. C. Raffel, D. Ellis (2016) "Extracting Ground-Truth Information from MIDI Files: A MIDIfesto," Proc. ISMIR, 796-802, New York, USA, 2016.
12. C. Raffel, D. Ellis (2016) "Optimizing DTW-based audio-to-MIDI alignment and matching," Proc. ICASSP, Shanghai, China, 2016.
13. C. Raffel, D. Ellis (2016) "Pruning subsequence search with attention-based embedding," Proc. ICASSP, Shanghai, China, 2016.
14. C. Raffel, D. Ellis (2015) "Large-Scale Content-Based Matching of MIDI and Audio Files," Proc. ISMIR, 234-240, Malaga, Spain, 2015.
15. D. Liang, M. Zhan, D. Ellis (2015) "Content-Aware Collaborative Music Recommendation Using Pre-trained Neural Networks," Proc. ISMIR, 295-301, Malaga, Spain, 2015.
16. B. McFee, C. Raffel, D. Liang, D. Ellis, M. McVicar, E. Battenberg, O. Nieto (2015) "librosa: Audio and music signal analysis in python," Proc. 14th Python in Science conf, 2015. **(Cited by 480)**
17. J. Le Roux, E. Vincent, J. Hershey, D. Ellis (2015) "MICbots: Collecting large realistic datasets for speech and audio research using mobile robots," Proc. ICASSP, Brisbane, Australia, 2015.
18. C. Raffel, D. Ellis (2015) "Accelerating multimodal sequence retrieval with convolutional networks," Proc. NIPS Multimodal Machine Learning Workshop, Montr'eal, Canada, 2014.
19. C. Raffel, B. McFee, E. Humphrey, J. Salamon, O. Nieto, D. Liang, D. Ellis (2014) "mir_eval: A Transparent Implementation of Common MIR Metrics," Proc. ISMIR, Taipei, Taiwan, October 2014. **(Cited by 194)**

20. B. McFee, D. Ellis (2014) “Analyzing Song Structure With Spectral Clustering,” Proc. ISMIR, Taipei, Taiwan, October 2014. **(Cited by 60)**
21. D. Liang, J. Paisley, D. Ellis (2014) “Codebook-based Scalable Music Tagging With Poisson Matrix Factorization,” Proc. ISMIR, Taipei, Taiwan, October 2014.
22. D. Ellis, H. Satoh, Z. Chen (2014) “Detecting proximity from personal audio recordings,” Proc. Interspeech, Singapore, September 2014.
23. Z. Chen, B. McFee, D. Ellis (2014) “Speech enhancement by low-rank and convolutive dictionary spectrogram decomposition,” Proc. Interspeech, Singapore, September 2014.
24. H. Papadopoulos, D. Ellis (2014) “Music-content-adaptive robust principal component analysis for a semantically consistent separation of foreground and background in music audio signals,” Proc. DAFx, Erlangen, Germany, September 2014.
25. Z. Chen, H. Papadopoulos, D. Ellis (2014) “Content-adaptive speech enhancement by a sparsely-activated dictionary plus low rank decomposition,” Proc. HSCMA, Nancy, France, May 2014.
26. D. Liang, D. Ellis, M. Hoffman, G. Mysore (2014) “Speech Decoloration Based On The Product-Of-Filters Model,” Proc. ICASSP, 2400-2404, Florence, May 2014.
27. B. McFee, D. Ellis (2014) “Better Beat Tracking Through Robust Onset Aggregation,” Proc. ICASSP, 2154–2158, Florence, May 2014.
28. B. McFee, D. Ellis (2014) “Learning To Segment Songs With Ordinal Linear Discriminant Analysis,” Proc. ICASSP, 5197–5201, Florence, May 2014.
29. M. McVicar, D. Ellis, M. Goto (2014) “Leveraging Repetition for Improved Automatic Lyric Transcription in Popular Music,” Proc. ICASSP, 3117–3121, Florence, May 2014.
30. C. Raffel, D. Ellis (2014) “Estimating Timing and Channel Distortion Across Related Signals,” Proc. ICASSP, 654–658, Florence, May 2014.
31. D. Silva, V. de Souza, G. Batista, E. Keogh, & D. Ellis (2013) “Applying Machine Learning and Audio Analysis Techniques to Insect Recognition in Intelligent Traps,” Proc. ICMLA, (to appear), Miami, December 2013.
32. D. Liang, M. Hoffman, & D. Ellis (2013) “Beta Process Sparse Nonnegative Matrix Factorization For Music,” Proc. ISMIR, 375–380, Curitiba, November 2013. **Best student paper.**
33. D. Silva, H. Papadopoulos, G. Batista, D. Ellis (2013) “A Video Compression-Based Approach To Measure Music Structure Similarity,” Proc. ISMIR, 95–100, Curitiba, November 2013.
34. Z. Chen & D. Ellis (2013) “Speech Enhancement By Sparse, Low-Rank, And Dictionary Spectrogram Decomposition,” Proc. IEEE WASPAA, 4pp, Mohonk, October 2013. DOI: 10.1109/WASPAA.2013.6701883
35. D. Gillespie & D. Ellis (2013) “Modeling nonlinear circuits with linearized dynamical models via kernel regression,” Proc. IEEE WASPAA, 4pp, Mohonk, October 2013. DOI: 10.1109/WASPAA.2013.6701830
36. M. Graciarena, A. Alwan, D. Ellis, H.Franco, L. Ferrer, J. Hansen, A. Janin, B.-S. Lee, Y. Lei, V. Mitra, N. Morgan, S. O. Sadjadi, T.J. Tsai, N. Scheffer, L. N. Tan, B. Williams (2013) “All for One: Feature Combination for Highly Channel-Degraded Speech Activity Detection,” Proc. Interspeech, paper #1338, Lyon, August 2013.
37. C. Cotton & D. Ellis (2013) “Subband Autocorrelation Features for Video Soundtrack Classification,” Proc. ICASSP-13, 8663–8666, Vancouver, May 2013.
38. T. Bertin-Mahieux & D. Ellis (2012) “Large-Scale Cover Song Recognition Using the 2D Fourier Transform Magnitude,” Proc. ISMIR, 241–246, Porto, October 2012. **(Cited by 85)**

39. J. McDermott, D. Ellis, & H. Kawahara (2012) "Inharmonic Speech: A Tool for the Study of Speech Perception and Separation," Proc. SAPA-SCALE 2012, 114–117, Portland, September 2012.

40. B.-S. Lee & D. Ellis (2012) "Noise Robust Pitch Tracking by Subband Autocorrelation Classification," Proc. Interspeech, paper P3b.05, Portland, September 2012.

41. K. Su, M. Naaman, A. Gurjar, M. Patel, & D. Ellis (2012) "Making a Scene: Alignment of Complete Sets of Clips based on Pairwise Audio Match," Proc. ACM Int. Conf. on Multimedia Retrieval (ICMR 2012), Hong Kong, June 2012.

42. B. McFee, T. Bertin-Mahieux, D. Ellis, & G. Lanckriet (2012) "The Million Song Dataset Challenge," Proc. WWW-2012 AdMIRe Workshop, 909–916, Lyon, April 2012. **(Cited by 165)**

43. T. Bertin-Mahieux, D. Ellis, B. Whitman & P. Lamere (2011) "The Million Song Dataset," Proc. ISMIR, 591–596, Miami, October 2011. **(Cited by 972)**

44. D. Ellis, B. Whitman, T. Jehan, and P. Lamere (2010) "Echoprint – An Open Music Identification System," ISMIR Late Breaking Abstracts, Miami, October 2011.

45. T. Bertin-Mahieux & D. Ellis (2011) "Large-scale Cover Song Recognition using Hashed Chroma Landmarks," Proc. IEEE WASPAA-11, 117–120, Mohonk, October 2011. **(Cited by 91)**

46. C. Cotton & D. Ellis (2011) "Spectral vs. Spectro-Temporal Features for Acoustic Event Detection," Proc. IEEE WASPAA-11, pp. 69–72, Mohonk, October 2011. **(Cited by 123)**

47. D. Ellis, X. Zeng, & J. McDermott (2011) "Classifying soundtracks with audio texture features," Proc. IEEE ICASSP-11, pp. 5880–5883, Prague, May 2011.

48. C. Cotton, D. Ellis, & A. Loui (2011) "Soundtrack classification by transient events," Proc. IEEE ICASSP-11, pp. 473–476, Prague, May 2011.

49. T. Bertin-Mahieux, G. Grindlay, R. Weiss, & D. Ellis (2011) "Evaluating music sequence models through missing data," Proc. IEEE ICASSP-11, pp. 177–180, Prague, May 2011.

50. C. Vezyrtzis, A. Klein, D. Ellis, & Y. Tsividis (2011) "Direct Processing of MPEG Audio Using Companding and BFP Techniques," Proc. IEEE ICASSP-11, pp. 361–364, Prague, May 2011.

51. Y.-G. Jiang, G. Ye, S.-F. Chang, D. Ellis, & A. C. Loui (2011) "Consumer Video Understanding: A Benchmark Database and An Evaluation of Human and Machine Performance," Proc. ACM ICMR, article #29, Trento, April 2011. **(Cited by 275)**

52. Y.-G. Jiang, X. Zeng, G. Ye, D. Ellis, S.-F. Chang, S Bhattacharya, M Shah (2010) "Columbia-UCF TRECVID2010 Multimedia Event Detection: Combining Multiple Modalities, Contextual Concepts, and Temporal Matching," TRECVID System Descriptions, December 2010. **(Cited by 124)**

53. G. Grindlay and D. Ellis (2010) "A Probabilistic Subspace Model for Multi-Instrument Polyphonic Transcription," Proc. ISMIR, pp. 21–26, Utrecht, August 2010.

54. T. Bertin-Mahieux, R. Weiss, and D. Ellis (2010) "Clustering beat-chroma patterns in a large music database," Proc. ISMIR, pp. 111–116, Utrecht, August 2010.

55. D. Ellis, B. Whitman, T. Jehan, and P. Lamere (2010) "The Echo Nest Musical Fingerprint," ISMIR Late Breaking Abstracts, Utrecht, August 2010.

56. D. Ellis and A. Weller (2010) "The 2010 LabROSA chord recognition system," MIREX 2010 system abstracts, August 2010.

57. S. Ravuri and D. Ellis (2010) "Cover Song Detection: From High Scores to General Classification," Proc. IEEE ICASSP-10, pp. 65–68, Dallas, March 2010. **(Cited by 63)**

58. C. Cotton and D. Ellis (2010) “Audio Fingerprinting to Identify Multiple Videos of an Event,” Proc. IEEE ICASSP-10, pp. 2386–2389, Dallas, March 2010.
59. K. Lee, D. Ellis, and A Loui (2010) “Detecting Local Semantic Concepts in Environmental Sounds using Markov Model based Clustering,” Proc. IEEE ICASSP-10, pp. 2278–2281, Dallas, March 2010.
60. A. Weller, D. Ellis, and T. Jebara (2009) “Structured Prediction Models for Chord Transcription of Music Audio,” Proc. Int. Conf. on Machine Learning and Applications, pp. 590–595, Miami Beach, December 2009.
61. C. Cotton and D. Ellis (2009) “Finding Similar Acoustic Events using Matching Pursuit and Locality-Sensitive Hashing,” Proc. WASPAA-09, pp. 125–128, Mohonk NY, October 2009.
62. C. Smit and D. Ellis (2009) “Guided Harmonic Sinusoid Estimation in a Multi-Pitch Environment,” Proc. WASPAA-09, pp. 41–44, Mohonk NY, October 2009.
63. G. Grindlay and D. Ellis (2009) “Multi-Voice Polyphonic Music Transcription Using Eigeninstruments,” Proc. WASPAA-09, pp. 53–56, Mohonk NY, October 2009.
64. J. Devaney, M. Mandel, and D. Ellis (2009) “Improving Midi-Audio Alignment with Acoustic Features,” Proc. WASPAA-09, pp. 45–48, Mohonk NY, October 2009.
65. M. Mandel and D. Ellis (2009) “The Ideal Interaural Parameter Mask: A Bound on Binaural Separation Systems,” Proc. WASPAA-09, pp. 85–88, Mohonk NY, October 2009.
66. W. Jiang, C. Cotton, S.-F. Chang, D. Ellis, and A. Loui (2009) “Short-Term Audio-Visual Atoms for Generic Video Concept Classification,” Proc. ACM MultiMedia-09, pp. 5–14, Beijing, October 2009. **(Cited by 73)**
67. J. Gudnason, M. Thomas, P. Naylor, and D. Ellis (2009) “Voice Source Waveform Analysis and Synthesis using Principal Component Analysis and Gaussian Mixture Modelling,” Proc. Interspeech-09, pp. 108-111, Brighton, September 2009.
68. J. Devaney & D. Ellis (2009) “Handling Asynchrony in Audio-Score Alignment,” Proc. Int. Comp. Music Conf. (ICMC), pp. 29–32, Montreal, August 2009.
69. J. B. Boldt & D. Ellis (2009) “A Simple Correlation-Based Model of Intelligibility for Nonlinear Speech Enhancement and Separation,” Proc. EUSIPCO-09, pp. 1849–1853, Glasgow, August 2009.
70. R. Weiss & D. Ellis (2009) “A Variational EM Algorithm for Learning Eigenvoice Parameters in Mixed Signals,” Proc. ICASSP-09, pp. 113–116, Taiwan, April 2009.
71. R. Weiss, M. Mandel, D. Ellis (2008) “Source Separation Based on Binaural Cues and Source Model Constraints,” Proc. Interspeech-08, pp. 419–422, Brisbane, Australia, September 2008.
72. K. Hu, P. Divenyi, D. Ellis, Z. Jin, B. Shinn-Cunningham, & D. Wang (2008) “Preliminary Intelligibility Tests of a Monaural Speech Segregation System,” Proc. SAPA-08, pp. 11–16, Brisbane, Australia, September 2008.
73. A. Lammert, D. Ellis, & P. Divenyi (2008) “Data-driven articulatory inversion incorporating articulator priors,” Proc. SAPA-08, pp. 29–34, Brisbane, Australia, September 2008.
74. M. Mandel & D. Ellis (2008) “Multiple-Instance Learning For Music Information Retrieval,” Proc. ISMIR 2008, pp. 577–582, Philadelphia, September 2008. **(Cited by 125)**
75. S. Ravuri & D. Ellis (2008) “Stylization of pitch with syllable-based linear segments,” Proc. ICASSP-08, pp. 3985–3988, Las Vegas, April 2008.
76. D. Ellis, C. Cotton, & M. Mandel (2008) “Cross-correlation of beat-synchronous representations for music similarity,” Proc. ICASSP-08, pp. 57–60, Las Vegas, April 2008.

77. J. H. Jensen, M. G. Christensen, D. Ellis, & S. H. Jensen (2008) “A tempo-insensitive distance measure for cover song identification based on chroma features,” Proc. ICASSP-08, pp. 2209–2212, Las Vegas, April 2008.
78. K. Lee & D. Ellis (2008) “Detecting music in ambient audio by long-window autocorrelation,” Proc. ICASSP-08, pp. 9–12, Las Vegas, April 2008.
79. M. Mandel & D. Ellis (2007) “EM localization and separation using interaural level and phase cues,” Proc. IEEE Workshop on Apps. of Sig. Proc. to Acous. and Audio WASPAA-07, pp. 275–278, Mohonk NY, October 2007.
80. R. Weiss & D. Ellis (2007) “Monaural speech separation using source-adapted models,” Proc. IEEE Workshop on Apps. of Sig. Proc. to Acous. and Audio WASPAA-07, pp. 114–117, Mohonk NY, October 2007.
81. C. Smit & D. Ellis (2007) “Solo voice detection via optimal cancelation,” Proc. IEEE Workshop on Apps. of Sig. Proc. to Acous. and Audio WASPAA-07, pp. 207–210, Mohonk NY, October 2007.
82. G. Poliner & D. Ellis (2007) “Improving generalization for polyphonic piano transcription,” Proc. IEEE Workshop on Apps. of Sig. Proc. to Acous. and Audio WASPAA-07, pp. 86–89, Mohonk NY, October 2007.
83. S.-F. Chang, D. Ellis, W. Jiang, K. Lee, A. Yanagawa, A. Loui, J. Luo (2007) “Large-scale multimodal semantic concept detection for consumer video,” Multimedia Information Retrieval workshop, ACM Multimedia, Augsburg, Germany, Sep 2007. (10pp) **(Cited by 143)**
84. A. Loui, J. Luo, S.-F. Chang, D. Ellis, W. Jiang, L. Kennedy, K. Lee, A. Yanagawa (2007) “Kodak’s consumer video benchmark data set: concept definition and annotation,” Multimedia Information Retrieval workshop, ACM Multimedia, pp. 245–254, Augsburg, Germany, Sep 2007. **(Cited by 143)**
85. D. Ellis (2007) “Classifying Music Audio with Timbral and Chroma Features,” Proc. Int. Conf. on Music Info. Retr. ISMIR-07, pp. 339–340, Vienna, Austria, Sep 2007. **(Cited by 123)**
86. M. Mandel & D. Ellis (2007) “A Web-Based Game for Collecting Music Metadata,” Proc. Int. Conf. on Music Info. Retr. ISMIR-07, pp, 365–366, Vienna, Austria, Sep 2007.
87. J. H. Jensen, D. Ellis, M. G. Chritensen, S. H. Jensen (2007) “Evaluation of Distance Measures Between Gaussian Mixture Models of MFCCs,” Proc. Int. Conf. on Music Info. Retr. ISMIR-07, pp. 107-108, Vienna, Austria, Sep 2007. **(Cited by 70)**
88. A. Doherty, A. Smeaton, K. Lee, & D. Ellis (2007) “Multimodal Segmentation of Lifelog Data,” Proc. 8th Int. Conf. on Computer-Assisted Information Retrieval RIAO 2007, Pittsburgh, May 2007. **(Cited by 68)**
89. J. Ogle & D. Ellis (2007) “Fingerprinting to Identify Repeated Sound Events in Long-Duration Personal Audio Recordings,” Proc. ICASSP-07, pp. I–233–236, Honolulu HI, Apr 2007. **(Cited by 63)**
90. D. Ellis & G. Poliner (2007) “Identifying Cover Songs With Chroma Features and Dynamic Programming Beat Tracking,” Proc. ICASSP-07, pp. IV–1429–1432, Honolulu HI, Apr 2007. **(Cited by 435)**
91. M. Mandel, D. Ellis, & T. Jebara (2006) “An EM algorithm for localizing multiple sound sources in reverberant environments,” Adv. NIPS, pp. 953–960, Vancouver BC, Dec 2006. **(Cited by 122)**
92. K. Lee & D. Ellis (2006) “Voice activity detection in personal audio recordings using autocorrelogram compensation,” Proc. Interspeech, pp. 1970-1973, Pittsburgh PA, Oct 2006.
93. M. Mandel & D. Ellis (2006) “A probability model for interaural phase difference,” Proc. Workshop on Statistical & Perceptual Audition SAPA-06, pp. 1-6, Pittsburgh PA, Oct 2006.

94. R. Weiss & D. Ellis (2006) "Estimating single-channel source separation masks: Relevance Vector Machine classifiers vs. pitch-based masking," Proc. Workshop on Statistical & Perceptual Audition SAPA-06, pp. 31-36, Pittsburgh PA, Oct 2006. **(Cited by 71)**
95. D. Ellis & R. Weiss (2006) "Model-Based Monaural Source Separation Using a Vector-Quantized Phase-Vocoder Representation," Proc. ICASSP-06, pp. V-957-960, Toulouse, May 2006. **(Cited by 93)**
96. X. Halkias & D. Ellis (2006) "Estimating the Number of Marine Mammals using Recordings of Clicks from One Microphone," Proc. ICASSP-06, pp. V-769-772, Toulouse, May 2006.
97. K. Dobson, B. Whitman, & D. Ellis (2005) "Learning Auditory Models of Machine Voices", IEEE Workshop on Apps. of Sig. Proc. to Acous. & Audio WASPAA-05, pp.339-342, Mohonk NY, October 2005.
98. G. Poliner & D. Ellis (2005) "A Classification Approach to Melody Transcription", Int. Conf. on Music Info. Retrieval ISMIR-05, pp. 161-166, London, September 2005. **(Cited by 74)**
99. M. Mandel & D. Ellis (2005) "Song-Level Features and Support Vector Machines for Music Classification", Int. Conf. on Music Info. Retrieval ISMIR-05, pp. 594-599, London, September 2005. **(Cited by 381)**
100. N. Lesser & D. Ellis (2005) "Clap Detection and Discrimination for Rhythm Therapy", Proc. ICASSP-05, pp. III-37-40, Philadelphia, March 2005.
101. C.-P. Chen, J. Bilmes, & D. Ellis (2005) "Speech Feature Smoothing for Robust ASR", Proc. ICASSP-05, pp. I-525-528, Philadelphia, March 2005.
102. M. Reyes-Gomez, N. Jojic, & D. Ellis (2005) "Deformable Spectrograms", AI & Statistics 2005, pp. 285-292, Barbados, Jan 2005.
103. D. Ellis & K.S. Lee (2004) "Minimal-Impact Audio-Based Personal Archives", Proc. 1st ACM workshop on Continuous Archiving and Recording of Personal Experiences CARPE-04, pp.39-47, New York, Oct 2004. **(Cited by 126)**
104. D. Ellis & J. Arroyo (2004) "Eigenrhythms: Drum pattern basis sets for classification and generation", International Symposium on Music Information Retrieval ISMIR-04, pp. 554-559, Barcelona, Oct 2004.
105. B. Whitman & D. Ellis (2004) "Automatic Record Reviews", International Symposium on Music Information Retrieval ISMIR-04, pp. 470-477, Barcelona, Oct 2004. **(Cited by 98)**
106. M. Athineos, H. Hermansky, & D. Ellis (2004) "LP-TRAP: Linear predictive temporal patterns", International Conference on Spoken Language Processing ICSLP-04, pp. 949-952, Jeju, Korea, Oct 2004.
107. M. Athineos, H. Hermansky, & D. Ellis (2004) "PLP^2: Autoregressive modeling of auditory-like 2-D spectro-temporal patterns", ISCA Tutorial and Research Workshop on Statistical and Perceptual Audio Processing SAPA-04, pp. 25-30, Jeju, Korea, Oct 2004.
108. M. Reyes-Gomez, N. Jojic, & D. Ellis (2004) "Towards single-channel unsupervised source separation of speech mixtures: The layered harmonics/formants separation-tracking model", ISCA Tutorial and Research Workshop on Statistical and Perceptual Audio Processing SAPA-04, pp. 37-42, Jeju, Korea, Oct 2004.
109. D. Ellis & K.S. Lee (2004) "Features for Segmenting and Classifying Long-Duration Recordings of Personal Audio", ISCA Tutorial and Research Workshop on Statistical and Perceptual Audio Processing SAPA-04, pp. 1-4, Jeju, Korea, Oct 2004.
110. D. Ellis & J. Liu (2004) "Speaker turn segmentation based on between-channel differences", NIST ICASSP Meeting Recognition Workshop, pp. 112-117, Montreal, May 2004.

111. L. Kennedy & D. Ellis (2004) “Laughter Detection in Meetings”, NIST ICASSP Meeting Recognition Workshop, pp. 118-121, Montreal, May 2004. **(Cited by 161)**
112. M.J. Reyes-Gomez, N. Jojic, & D. Ellis (2004) “Detailed graphical models for source separation and missing data interpolation in audio”, Learning Workshop, Snowbird, 2004.
113. M.J. Reyes-Gomez, D. Ellis, & N. Jojic (2004) “Multiband Audio Modeling for Single Channel Acoustic Source Separation”, Proc. ICASSP-04, pp. V-641-644, Montreal, May 2004. **(Cited by 64)**
114. L. Kennedy & D. Ellis (2003) “Pitch-based emphasis detection for characterization of meeting recordings”, Automatic Speech Recognition and Understanding Workshop IEEE ASRU 2003, pp. 243-284, St. Thomas, December 2003.
115. M. Athineos & D. Ellis (2003) “Frequency-domain linear prediction for temporal features”, Automatic Speech Recognition and Understanding Workshop IEEE ASRU 2003, pp. 261-266, St. Thomas, December 2003. **(Cited by 65)**
116. A. Sheh & D. Ellis (2003) “Chord Segmentation and Recognition using EM-Trained Hidden Markov Models”, 4th International Symposium on Music Information Retrieval ISMIR-03, pp. 185-191, Baltimore, October 2003. **(Cited by 409)**
117. R. Turetsky & D. Ellis (2003) “Ground-Truth Transcriptions of Real Music from Force-Aligned MIDI Syntheses”, 4th International Symposium on Music Information Retrieval ISMIR-03, pp. 135-141, Baltimore, October 2003. **(Cited by 154)**
118. A. Berenzweig, B. Logan, D. Ellis, & B. Whitman (2003) “A large-scale evaluation of acoustic and subjective music similarity measures”, 4th International Symposium on Music Information Retrieval ISMIR-03, pp. 103-109, Baltimore, October 2003. **(Cited by 430)**
119. M.J. Reyes-Gomez, B. Raj, & D. Ellis (2003) “Multi-channel Source Separation by Beamforming Trained with Factorial HMMs”, Proc. IEEE Workshop on Apps. of Sig. Proc. to Acous. and Audio, pp. 13-16, Mohonk NY, October 2003.
120. B. Logan, D. Ellis, & A. Berenzweig (2003) “Toward evaluation techniques for music similarity”, Keynote address, Workshop on the Evaluation of Music Information Retrieval (MIR) Systems at SIGIR 2003, Toronto, August 2003.
121. P. Scanlon, D. Ellis, & R. Reilly (2003) “Using Mutual Information to design class-specific phone recognizers”, Proc. Eurospeech-03, pp. 857-860, Geneva, September 2003.
122. M.J. Reyes-Gomez & D. Ellis (2003) “Selection, Parameter Estimation, and Discriminative Training of Hidden Markov Models for General Audio Modeling”, Proc. ICME-03, pp. I-73-76, Baltimore, July 2003.
123. A. Berenzweig, D. Ellis, & S. Lawrence (2003) “Anchor Space for Classification and Similarity Measurement of Music”, Proc. ICME-03, pp. I-29-32, Baltimore, July 2003. **(Cited by 90)**
124. S. Renals & D. Ellis (2003) “Audio Information Access from Meeting Rooms”, Proc. ICASSP-03, pp. IV-744-747, Hong Kong, April 2003 (invited).
125. M. Athineos & D. Ellis (2003) “Sound Texture Modelling with Linear Prediction in both Time and Frequency Domains”, Proc. ICASSP-03, pp. V-648-651, Hong Kong, April 2003. **(Cited by 65)**
126. M. Reyes, B. Raj, & D. Ellis (2003) “Multi-channel Source Separation by Factorial HMMs”, Proc. ICASSP-03, pp. I–664-667, Hong Kong, April 2003.
127. A. Janin, D. Baron, J. Edwards, D. Ellis, D. Gelbart, N. Morgan, B. Peskin, T. Pfau, E. Shriberg, A. Stolcke, & C. Wooters (2003) “The ICSI Meeting Corpus”, Proc. ICASSP-03, pp. I-364-367, Hong Kong, April 2003. **(Cited by 600)**

128. D. Ellis, B. Whitman, A. Berenzweig, & S. Lawrence (2002) "The Quest for Ground Truth in Musical Artist Similarity", Proc. ISMIR-02, pp. 170-177, Paris, October 2002. **(Cited by 196)**
129. A. Berenzweig, D. Ellis, & S. Lawrence (2002) "Using Voice Segments to Improve Artist Classification of Music", Proc. AES-22 Intl. Conf. on Virt., Synth., and Ent. Audio. Espoo, Finland, June 2002. **(Cited by 147)**
130. M.J. Reyes-Gomez & D. Ellis (2002) "Error visualization for tandem acoustic modeling on the Aurora task," Proc. IEEE Int. Conf. on Acous., Speech & Sig. Proc., Student Session, Orlando, May 2002.
131. T. Pfau, D. Ellis & A. Stolcke (2001) "Multispeaker Speech Activity Detection for the ICSI Meeting Recorder," Proc. ASRU-01, 4 pp., Italy, December 2001. **(Cited by 123)**
132. A. Berenzweig & D. Ellis (2001) "Locating Singing Voice Segments within Music Signals," Proc. IEEE Workshop on Apps. of Sig. Proc. to Acous. and Audio, Mohonk, pp. 119-122, October 2001. **(Cited by 197)**
133. D. Ellis (2001) "Detecting Alarm Sounds," Proc. CRAC workshop, pp. 59-62, Aalborg, September 2001.
134. J. Barker, M. Cooke, & D. Ellis (2001) "Integrating bottom-up and top-down constraints to achieve robust ASR: The multisource decoder," Proc. CRAC workshop, pp. 63-66, Aalborg, September 2001.
135. D. Ellis & M. Reyes (2001) "Investigations into Tandem Acoustic Modeling for the Aurora Task," Proc. Eurospeech-01 (Special Event on Noise Robust Recog.), pp. 189-192, Denmark, September 2001. **(Cited by 60)**
136. D. Ellis, R. Singh, & S. Sivadas (2001) "Tandem acoustic modeling in large-vocabulary recognition," Proc. IEEE Int. Conf. on Acous., Speech & Sig. Proc., pp. I-517-520, Salt Lake City, May 2001. **(Cited by 132)**
137. N. Morgan, D. Baron, J. Edwards, D. Ellis, D. Gelbart, A. Janin, T. Pfau, E. Shriberg, & A. Stolcke (2001) "The Meeting Project at ICSI" Proc. Human Lang. Tech. Conf., pp. 1-7, San Diego, March 2001. **(Cited by 168)**
138. D. Ellis & J. Bilmes (2000) "Using mutual information to design feature combinations," Proc. Int. Conf. Spoken Lang. Proc., Beijing, October 2000.
139. J. Barker, M. Cooke, & D. Ellis (2000) "Decoding speech in the presence of other sound sources," Proc. Int. Conf. Spoken Lang. Proc., Beijing, October 2000.
140. J. Ferreiros-Lopez & D. Ellis (2000) "Using acoustic condition clustering to improve acoustic change detection on Broadcast News," Proc. Int. Conf. Spoken Lang. Proc., Beijing, October 2000.
141. H. Hermansky, D. Ellis, & S. Sharma (2000) "Tandem connectionist feature extraction for conventional HMM systems," Proc. IEEE Int. Conf. on Acous., Speech & Sig. Proc., pp. III-1365-1368, Istanbul, June 2000. **(Cited by 873)**
142. S. Sharma, D. Ellis, S. Kajarekar, P. Jain, & H. Hermansky (2000) "Feature extraction using non-linear transformation for robust speech recognition on the Aurora database," Proc. IEEE Int. Conf. on Acous., Speech & Sig. Proc., pp. II-1117-1120, Istanbul, June 2000. **(Cited by 104)**
143. D. Genoud, D. Ellis, & N. Morgan (1999) "Combined speech and speaker recognition with speaker-adapted connectionist models," Proc. IEEE Automatic Speech Recognition & Understanding Workshop, Keystone CO, December 1999.
144. D. Abberley, S. Renals, T. Robinson, & D. Ellis (1999) "The THISL SDR system at TREC-8," Proc. Text Retrieval Conference TREC-8, Washington, November 1999.

145. G. Williams & D. Ellis (1999) “Speech/music discrimination based on posterior probability features,” Proc. Eurospeech-99, Budapest, September 1999. **(Cited by 167)**
146. A. Janin, D. Ellis, & N. Morgan (1999) “Multistream: Ready for prime-time?” Proc. Eurospeech-99, Budapest, September 1999. **(Cited by 78)**
147. D. Ellis & N. Morgan (1999) “Size matters: An empirical study of neural network training for large vocabulary continuous speech recognition,” Proc. IEEE Int. Conf. on Acous., Speech & Sig. Proc., Phoenix, April 1999. **(Cited by 61)**
148. N. Morgan, D. Ellis, E. Fosler-Lussier, A. Janin, & B. Kingsbury (1999) “Reducing errors by increasing the error rate: MLP acoustic modeling for Broadcast News transcription,” DARPA Broadcast News Transcription and Understanding Workshop, Herndon VA, February 1999.
149. G. Cook, J. Christie, D. Ellis, E. Fosler-Lussier, Y. Gotoh, B. Kingsbury, N. Morgan, S. Renals, A. Robinson, & G. Williams (1999) “An overview of the SPRACH system for the transcription of Broadcast News,” DARPA Broadcast News Transcription and Understanding Workshop, Herndon VA, February 1999.
150. D. Ellis (1997) “Computational Auditory Scene Analysis exploiting speech recognizer knowledge,” Proc. IEEE Workshop on Apps. of Sig. Proc. to Acous. and Audio, Mohonk, Oct 1997.
151. D. Ellis (1997) “The Weft: a representation for periodic sounds,” Proc. IEEE Int. Conf. on Acous., Speech & Sig. Proc., Munich, April 1997.
152. S. Greenberg, J. Hollenback, & D. Ellis (1996) “Insights into spoken language gleaned from phonetic transcriptions of the Switchboard corpus,” Proc. Int. Conf. on Spoken Lang. Proc., Philadelphia, October 1996. **(Cited by 206)**
153. S. Greenberg, J. Hollenback, & D. Ellis (1996) “The Switchboard transcription project,” LVCSR Summer Workshop Technical Reports, August 1996. **(Cited by 116)**
154. D. Ellis (1996) “Prediction-driven computational auditory scene analysis for dense sound mixtures,” Proc. ESCA Workshop on the Auditory Basis of Speech Perception, Keele, July 1996.
155. D. Ellis (1995) “Underconstrained stochastic representations for top-down computational auditory scene analysis,” Proc. IEEE Workshop on Apps. of Sig. Proc. to Acous. and Audio, Mohonk, October 1995.
156. D. Ellis (1994) “A computer implementation of psychoacoustic grouping rules,” Proc. 12th Int. Conf. on Pattern Recog., Jerusalem, October 1994.
157. D. Ellis (1993) “Hierarchic models for sound analysis and reconstruction,” Proc. IEEE Workshop on Apps. of Sig. Proc. to Acous. and Audio, Mohonk, October 1993.
158. D. Ellis (1992) “Timescale modifications and wavelet representations,” Proc. Int. Computer Music Conf., San José , pp. 6-9, June 1992.
159. D. Ellis & B. Vercoe (1991) “A perceptual representation of audio for co-channel source separation,” Proc. IEEE Workshop on Apps. of Sig. Proc. to Acous. and Audio, Mohonk, 1991.
160. D. Ellis & B. Vercoe (1991) “A wavelet-based sinusoid model of sound for auditory signal separation,” Proc. Int. Computer Music Conf., Montréal, September 1991
161. B. Vercoe & D. Ellis (1990) “Real-time CSound: Software Synthesis with Sensing and Control,” Proc. Int. Computer Music Conf., Glasgow, pp. 209-211, September 1990. **(Cited by 69)**

Patents:

1. “Apparatus and methods for processing a signal using a fixed-point operation”, joint with C. Vezyrtzis, A. Klein, and Y. Tsividis, covers a technique for performing signal processing that maximizes the quality of limited-resolution processors, US patent 8,788,277 granted July 22, 2014.

2. "Methods, Systems, and Media for Identifying Similar Songs Using Two-Dimensional Fourier Transform Magnitudes", joint with Thierry Bertin-Mahieux, covers technique for summarizing musical fragments into a fixed-dimensional representation that is robust to small timing offsets, US 20130226957 A1 filed Feb 27, 2013.
3. "Musical fingerprinting", joint with Brian Whitman and Andrew Nesbit of the Echo Nest Corporation, describes technique for matching music audio soundtracks by matching timings of onset events in different frequency bands, US8,492,633 B2 granted Jul 23, 2013.
4. "System for measuring electric power consumption", sole inventor, filed Mar 2011.
5. "System for identifying cover versions of musical recordings", sole inventor, filed Sep 2007.
6. "Relevance feedback and active learning for music playlist generation", co-inventor with Graham Poliner and Michael Mandel, covers music similarity work based on audio alone, filed 2006.
7. "Systems and methods for speech recognition using frequency domain linear prediction polynomials to form temporal and spectral envelopes from frequency domain representations of signals", co-inventor with Marios Athineos and Hynek Hermansky, covers the FDLP and PLP^2 techniques for extracting and exploiting temporal envelopes in speech recognition, US 7,672,838 B1 granted Mar 2, 2010.
8. "Nonlinear mapping for feature extraction in automatic speech recognition", co-inventor with Hynek Hermansky and Sangita Sharma, covers the Tandem technique for using a neural-net classifier as a front-end to a conventional speech recognizer, US 7,254,538 B1 granted Aug 7, 2007.

Invited talks:

(slides viewable at `http://www.ee.columbia.edu/~dpwe/talks/`)

"On Communicating Computational Research," invited talk, Scholarly Communication Program, Columbia University, 2013-04-04.

"Augmenting and Exploiting Auditory Perception for Complex Scene Analysis," invited talk, Defense Science Research Council Workshop on Electronic Enhancement of Sensory Dead Space, Arlington VA, 2013-03-28.

"The State of Music at LabROSA," invited talk, North-East Music Information SIG (NEMISIG), at The Echo Nest, Somerville, MA, 2013-01-26.

"Music Information Retrieval for Jazz," invited talk, Center for Jazz Studies, Columbia University, 2012-11-15.

"Recognizing and Classifying Environmental Sounds," invited talk, Speech and Audio in the North East (SANE) workshop, MERL, Cambridge, MA, 2012-10-24.

"Handling Speech in the Wild," invited talk, Hearing Research Seminar, Boston University, 2012-10-05.

"Mining Audio," invited talk, Data to Solutions seminar, Columbia University, 2012-09-14.

"Mining Large-Scale Music Data Sets," invited talk, ITA-2012, San Diego, 2012-02-09.

"Engineering & the World," invited presentations, 4th grade science class, The School at Columbia University, 2012-01-19.

"Perceptually-Inspired Music Audio Analysis," invited talk, WISSAP, Indian Institute of Science, Bangalore, 2012-01-07.

"Speech Separation for Recognition and Enhancement," invited talk, DARPA Language Technologies Day, Tampa FL, 2011-10-27.

"Joint Audio-Visual Signatures for Web Video Analysis," project report, National Geospatial Intelligence Agency Annual Research Symposium, 2011-08-31.

“Environmental Sound Recognition and Classification,” Keynote talk, Hands Free Speech Communication and Microphone Arrays 2011, Edinburgh, 2011-06-01.

“Using the Soundtrack to Classify Videos,” invited talk, Visual Analytics Consortium 2011 meeting, University of Maryland, 2011-05-04.

“Extracting Information from Sound,” invited talk, Digital Media Analysis Search and Management International Workshop, CalIT2 center, UC San Diego, 2011-03-01.

“Music Audio Research at LabROSA,” invited talk, North East Music Information SIG, Drexel University, Philadelphia, 2011-01-28.

“Joint Audio-Visual Signatures for Web Video Analysis,” project report, at the National Geospatial Intelligence Agency Annual Research Symposium, 2010-09-14.

“A History and Overview of Machine Listening,” invited talk, Computational Audition workshop, UCL Gatsby unit, 2010-05-12.

“Using Speech Models for Separation,” invited talk, Acoustical Society Meeting, Baltimore, 2010-04-20.

“Music Audio Research at LabROSA,” invited talk, North East Music Information Special Interest Group (NEMISIG-2010), NYU, 2010-01-29.

“Some projects in real-world sound analysis ,” invited talk, NYU Music Technology, New York, 2009-12-10.

“Using Speech Models for Separation,” invited talk, Center for Speech and Language Understanding, Johns Hopkins University, Baltimore, 2009-10-13.

“Learning, Using, and Adapting Models in Scene Analysis,” invited talk, Scene Analysis Workshop, Berlin Institute of Advanced Studies, 2009-04-23.

“Mining for the Meaning of Music,” invited talk, Music Technology Seminar, New York University, 2008-10-17.

“Mining for the Meaning of Music,” invited talk, Distinguished Lecturer Series, Centre for Interdisciplinary Reearch on Music, Mind, and Technology, McGill University/University de Montréal, Montreal, 2008-03-27.

“Searching for Similar Phrases in Music Audio,” invited talk, Digital Music Research Network workshop, Queen Mary, University of London, 2007-12-18.

“Using source modes in speech separation,” invited presentation at the Next-Generation Statistical Models for Speech and Audio Signal Processing workshop, Radcliffe Institute for Advanced Study, 2007-11-09.

“Extracting and Using Music Audio Information,” invited ECE department seminar, UC San Diego, 2007-11-02.

“Analysis of everyday sounds,” invited talk, Kodak Research Labs, Rochester NY, 2007-07-24.

“Using sound source models to organize mixtures,” invited talk, ASIP-NET.DK hearing aid industry seminar, Denmark Technical University, 2007-05-24.

“Beat-synchronous chroma representations for music analysis,” invited talk, Intelligent Sound Workshop ’08, Karlslunde, Denmark, 2007-05-23.

“Sound Organization by Source Models in Humans and Machines,” invited talk, NIPS Workshop on Advances in Models of Acoustic Processing, Whistler BC, 2006-12-09.

“Extracting Information from Music Audio,” invited talk at joint Danish Technical University-Aalborg University Intelligent Sound Workshop, Sæby, Denmark, 2006-05-22.

“Auditory Scene Analysis in Humans and Machines,” tutorial at the AES Convention, Paris, 2006-05-20.

“VQ Source Models: Perceptual & Phase Issues,” invited talk at special session on source separation organized by Prof. Shoji Makino, IEEE ICASSP-06, Toulouse, 2006-05-16.

"Using Learned Source Models to Organize Sound Mixtures," invited talk at New Ideas In Hearing workshop organized by Dr. Alain de Cheveigné, Ecole Normale Supériere, Paris, 2006-05-12.

"Model-Based Separation in Humans and Machines," invited talk at the special session on approaches to audio separation organized by Emmanuel Vincent, ICA-2006, Charleston SC, 2006-03-08.

"Music Information Extraction," invited talk as guest of Prof. Ozgur Izmirli at Connecticut College, 2006-02-13.

"Speech Separation in Humans and Machines," **opening keynote**, IEEE Automatic Speech Recognition and Understanding ASRU-05, San Juan Puerto Rico, 2005-11-28.

"Computational Auditory Scene Analysis" and "Model-Based Scene Analysis", invited talks at the Hearing Aid Developers" Forum organized by Prof. Birger Kollmeier, Univ. Oldenburg, Germany, 2005-06-30.

"Audio signal recognition for speech, music, and environmental sounds," invited talk, Special session on classification, 146th meeting of the Acoustical Society of America, Austin, 2003-11-13.

Professional affiliations:

Member of the Audio Engineering Society since 2004.

Member of the International Speech Communications Association since 1999.

Member of the Institute of Electrical and Electronic Engineers since 1994.

Member of the Acoustical Society of America since 1993.

Legal work:

Analysis and explanation of patents; searches for prior art; expert witness reporting and testifying.

Deposition experience.

Other information:

Since 1993, administrator of the AUDITORY list, an email discussion list for researchers in auditory organization (currently over 2216 participants in 33 countries). `http://www.auditory.org/`

General Chair of the 2011 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, Mohonk NY (159 participants, 86 papers presented).

Co-guest editor of Special Issue of *IEEE Transactions on Selected Topics in Signal Processing* on Music Signal Processing, October 2011.

Organizer, special session on Engineering of Music, National Academy of Engineering 2010 Frontiers of Engineering Meeting, Armonk NY, September 2010.

Co-chair of the ISCA Tutorial and Research Workshop on Statistical and Perceptual Audition SAPA-2010, held as a satellite to Interspeech 2010 in Makuhari, Japan, September 2010 (9 presentations, around 30 participants). `http://www.sapa2010.org/`

General co-chair of the 9th International Conference on Music Information Retrieval (ISMIR-08), Philadelphia, September 2008 (105 papers and over 250 participants). `http://www.ismir2008.org/`

Co-chair of the ISCA Tutorial and Research Workshop on Statistical and Perceptual Audition SAPA-2008, held as a satellite to Interspeech 2008 in Brisbane, Australia, September 2008 (9 presentations, around 20 participants). `http://www.sapa2008.org/`

Co-guest editor of Special Issue of IEEE Transactions on Speech and Audio Processing on Music Information Retrieval, January 2008.

Co-organizer of the first North-East Music Information Special Interest Group (NEMISIG) meeting, a regional symposium for graduate students working in music informatics, Columbia University, 2008-01-

25.

Co-chair of the ISCA Tutorial and Research Workshop on Statistical and Perceptual Audition SAPA-2006, held as a satellite to ICSLP-2006 in Pittsburgh PA, September 2006 (12 papers presented, 37 participants). http://www.sapa2006.org/

Co-guest editor of Special Issue of IEEE Transactions on Speech and Audio Processing on Statistical and Perceptual Audio Processing (12 papers, Jan 2006).

Co-organizer of Workshop on Music Information Processing Systems, held as a satellite to the Neural Information Processing Systems conference, Whistler, BC, December 2004 (8 papers presented, 25 participants). http://www.iro.umontreal.ca/~eckdoug/mips/

Co-organizer of AFOSR/NSF-sponsored 2nd Montreal Workshop on Speech Separation, November 2004. 40 invited participants. http://labrosa.ee.columbia.edu/Montreal2004/

Co-chair of the ISCA Tutorial and Research Workshop on Statistical and Perceptual Audio Processing SAPA-2004, held as a satellite to ICSLP-2004 in Jeju-da, Korea, November 2004 (22 papers, 40 participants). http://www.sapa2004.org/

Co-guest editor of Special Issue of Speech Communication on Recognition and Organization of Real-World Sounds (vol 43 issue 4, published September 2004).

Co-organizer of NSF-sponsored workshop on Speech Separation, held in Montreal in November 2003. Twenty invited participants attended. http://www.ebire.org/speechseparation/

Invited Participant of NSF workshop on Next-Generation Automatic Speech Recognition, Atlanta, Oct 2003. http://users.ece.gatech.edu/~chl/ngasr03/

Co-chair of workshop on Consistent and Reliable Acoustic Cues for sound analysis (CRAC2001), held as a satellite event to Eurospeech in Denmark in September 2001 (24 papers presented, over 40 participants). http://www.ee.columbia.edu/crac/

Co-organizer of the first Workshop on Computational Auditory Scene Analysis in Montreal, 1995, and the following meeting in Nagoya, 1997.

Associate Editor for IEEE Multimedia Magazine (since 2005) and IEEE Signal Processing Magazine (Lecture Notes column, since 2006).

Co-chair of Audio Engineering Society Technical Committee on Semantic Audio Analysis (since 2004).

Treasurer of the 1997 IEEE Workshop on Apps. of Sig. Proc. to Acous. and Audio (Mohonk).

Reviewer for IEEE ICASSP, IEEE Tr. Speech & Audio Proc., Speech Communication, Computer Speech & Language, IEEE Mohonk WASPAA, AES Convention, IEEE Trans. Neural Networks, Free Speech Journal (online), J. VLSI Sig. Proc. Sys.

Regular contributor of software and book reviews published in IEEE Spectrum magazine.

Awarded a Harkness Fellowship of the Commonwealth Fund of New York to study at MIT.

Author of numerous public-domain sound analysis and processing tools for Unix including “dspB” (a soundfile viewer) and ‘pva” (a phase vocoder/short-time Fourier analyzer); maintainer of the ‘SPRACH-core” set of connectionist speech recognition tools.