


Nicolas OBIN

Associate Professor
Ircam - Sorbonne Université

Ircam
Analyse et Synthèse des Sons
1, place Igor Stravinsky 75004 Paris
☎ (+33) 01 44 78 48 90
✉ Nicolas.Obin@ircam.fr
 nicolasobin



Nicolas Obin (born 1980) is associate professor (2013) in speech, audio, and music technologies at Sorbonne Université and researcher in the Sciences and Technologies of Music and Sound (STMS) lab at Ircam - Center Pompidou, Paris, France. My main research interests cover speech processing and machine learning with application to modeling of speech prosody, speech synthesis, and voice conversion with applications in sound design, robotics, and computational musicology. I have a PhD. in computer sciences on the modeling of speech prosody and speaking style for text-to-speech synthesis (2011), and a master degree in music, art, and aesthetics (2007). I am passionate and deeply involved in the promotion of speech technologies for creation, art, and culture. I have collaborated with many artists such as Georges Aperghis, Roman Polansky, Philippe Parreno, Eric Rohmer, André Dussolier, Hector Parra, Marguerite Humeau, and Djamel Deboze.

Career and Formation

- [2013-] Associate Professor, Ircam and Sorbonne Université, Paris, France.
- [2006-2011] PhD. thesis, Ircam-Upmc, MeLos: Analysis and Modelling of Speech Prosody and Speaking Style.
- [2005-2006] Master in Engineering ATIAM (Acoustics, Signal Processing and Informatics Applied to Music), UPMC. Major of the promotion.
- [2006-2008] Master in Musicology (Musicology, Arts, Aesthetics), Saint-Denis University. *Pierre Boulez, Franz Kafka: L'écriture de l'irrationnel.*

Coordination and participation to research programs

Coordination

- [2017-2021] Project ANR TheVoice (voice design for the creative industry): PI. Voice identity conversion, voice casting.
- [2015-2016] Project EMERGENCE Sorbonne Universités ROUTE (Robot à l'Ecoute, Listening robot) : PI, binaural sound source localization.
- [2011-2015] Project FEDER Voice4Games (Speech technologies for video games) : PI. Speech levelling and voice casting of expressive speech.
- [2006-2008] Project PEPS Automatic learning of a generative model of speech prosody: PI.

Participation

- [2014-2017] Project ANR ChaNTeR (Real-time controlled digital singing) : participation, singing voice synthesis.
- [2007-2011] Project ANR Rhapsodie (A reference prosodic corpus of spoken French): participation, responsible for speech to text alignment and prosodic labelling.
- [2005-2008] Project RIAM VIVOS (Creation of expressive voices : multimedia applications): participation, prosody modelling for speech transformation and synthesis.

PhD supervision and monitoring

Supervision

- [2017-2021] Rafael Ferro, Expressive conversion of voice identity, Ircam, Université Pierre et Marie Curie, école doctorale informatique, télécommunications et électronique, 2017.
- [2017-2021] Mathias Quillot, Voice recommendation by deep learning, LIA/Ircam, Université d'Avignon et des Pays du Vaucluse, 2017.
- [2016-2019] Maxime Sirbu, Real-time and programmable machine listening, Ircam, Université Pierre et Marie Curie, école doctorale informatique, télécommunications et électronique, 2015.
- [2013-2016] Olivier Migliore, Analyzing the musical prosody of French punk, rap, and ragga (1977-1992) à l'aide de l'outil informatique, Université Montpellier 3 (PhD in musicology, ED 58 Langues, littératures, cultures, civilisations).

Member of the monitoring committee

- [2015-2018] Neil Zeghidour, Toward unsupervised speech recognition. Learning of informative and invariant representations. LSCP, Paris Sciences et Lettre / Ecole Normale Supérieure, Facebook A.I Research, ED Cerveau, cognition, comportement.
- [2016] Olivier Migliore, Analyzing the musical prosody of French punk, rap, and ragga (1977-1992) à l'aide de l'outil informatique, Université Montpellier 3 (PhD in musicology, ED 58 Langues, littératures, cultures, civilisations). Jury member.

Responsibilities

- [2018] Ircam representative for the AI commission and strategy at Sorbonne Université
- [2017-] Member of the scientific council of the engineering UFR at UPMC
- [2017] Member of the selection committee for the associate professor position at Ircam-UPMC, section 60-61, "multi-physical systems and sound signals"

Awards and distinctions

- [2011] Best PhD thesis award, Fondation Des Treilles.

Organization of conferences, seminars, workshops...

- [2018] Living Lab Supramuros, Avignon theater festival, invited speaker, TheVoice project, 2018.
- [2018] "Workshop on Detection, Classification, Localization, and Density Estimation of Marine Mammals" (DCLDE 2018) : member of the scientific committee.
- [2016] Conference "Journées d'Étude de la Parole" (JEP), 300 people, Paris : member of the organization committee, member of the scientific committee.
- [2011] Workshop on "Advances in Speech Technologies", Paris : organizer.
- [2010] Journées Jeunes Chercheurs en Audition, Acoustique musicale et Signal audio (JJCAS), Paris : member of the organization committee.
- [2008] International conferences EMUS (Emotion in MUsic and Speech), Paris (France), Campinas (Brazil), Lyon (France) : member of the organization committee, member of the scientific committee.

Reviewing

- [1] Reviewer for the French Research National Agency (ANR, 2008, 2015, 2017)
- [2] Expertise pour l'Agence Nationale de la Recherche et de la Technologie (ANRT, 2018)

Journals

- [1] IEEE/ACM Transactions of Audio, Speech, and Language Processing (2013)
- [2] IEEE Transactions on Affective Computing (2015)
- [3] Elsevier Speech Communication (2015, 2018)
- [4] ACM Journal on Computing and Cultural Heritage (2017)
- [5] Acta Acustica (2017)
- [6] International Journal of Multimedia Information Retrieval (2017)
- [7] Elsevier Biomedical Signal Processing and Control (2015)
- [8] Journal of Speech Sciences (2014)
- [9] Springer Journal on Language Resources and Evaluation (2012)

Conferences

- [1] IEEE International Conference on Acoustics, Speech, and Signal Processing (2012, 2018)
- [2] ISCA Interspeech (2016, 2017, 2018)
- [3] International Symposium on Music Information Retrieval (2012, 2017)
- [4] Speech Prosody (2008, 2016, 2018)
- [5] International Conference on Digital Audio Effects (2017)
- [6] Swiss Workshop on Prosody (2014)
- [7] Journées d'Etude de la Parole (2016)
- [8] International Workshop on Detection, Classification, Localization, and Density Estimation of marine mammals using passive acoustics (2018)

Artistic

- [1] Reviewer for the musical residency program at Ircam (since 2011)

Artistic Diffusion

- [**2014-2015**] "Pourquoi j'ai pas mangé mon père", French movie directed by Jamel Debbouze. Recreation of the voice of the French actor Louis de Funès, by using spectral voice identity conversion.
- [**2014-2015**] "Juger Pétain", TV documentary directed by Richard Saada, Planète, France 5. Recreation of the voice of Maréchal Pétain and other French personalities of the second world war, by using spectral voice identity conversion.
- [**2014**] "La Voix" exhibition, Cité des Sciences et de l'Industrie, Paris : member of the organization committee, responsible for the realization of three interactive terminals on voice casting and voice transformation.
- [**2012-2013**] "Marilyn", short movie directed by Philippe Parreno, Fondation Beyeler (2012) et Palais de Tokyo (2013). Recreation of the voice of Marilyn Monroe, by using intonation conversion.

Broad diffusion of the scientific culture

Articles in newspaper

- [1] Des voix de synthèse presque humaines, Pour la Science, Décembre 2016, numéro 470, p. 54-62.
- [2] On va pouvoir ressusciter la voix de Marilyn Monroe, Sciences et Vie, Juin 2015, p. 104-108.
- [3] J'ai cherché Audrey, la synthèse vocale de mon ordinateur, Rue 89, 5 mai 2015.

[4] De Funès dans le film de Debboze ? Mais c'est la voix de qui ? Rue 89, 8 avril 2015.

[5] "Pourquoi j'ai pas mangé mon père" : l'Ircam refait parler Louis de Funès, Sciences et Avenir, 8 avril 2015.

Participation to broadcast programs

[1] A l'Ircam, la salle la plus silencieuse au monde dévoile ses secrets, journal télévisé, 20h, TF1, 9 septembre 2016.

[2] Donner de la Voix, FutureMag, Arte, 5 décembre 2015.

[3] Carte blanche d'Eva Bettan, A'live, France Inter, 18 Juin 2015.

[4] La Visite Musicale, Le Magazine de la Santé, France 5, 27 mars 2015.

[5] Pianotine, Visite de l'Institut de Recherche sur l'Acoustique et la Musique, Arte, 30 Novembre 2014.

[6] Petite Histoire de la Synthèse Vocale, France Culture, 12 Août 2014.

[7] L'Informatique cherche ses Voix, France Culture, 11 Août 2014.

[8] On a Volé la Voix d'André Dussolier, Le Monde, 7 Juillet 2014.

[9] Parler avec la Voix d'un Autre, On n'est pas que des cobayes, France 5, 2014.

[10] Voix et Identité, Universcience, Arte Creative, 2014.

[11] Vers une autre Voix, France Culture Plus, 2014.

[12] Les Intonations, France Culture Plus, David Christoffel, 2013.

Public lectures

[1] Les voix du numérique pour le design sonore et la création artistique, Festival VIno Voce, Saint-Emilion, France, 8-10 septembre 2017.

[2] Vox ex-machina. La voix et ses doubles numériques, Collège de France, Paris, France, 6 juin 2017.

[3] La voix fantôme, Exposition Polyphonies, Centre Pompidou, Paris, France, Décembre 2016.

[4] Voix d'outre-tombe : recréer les voix du passé à partir d'archives sonores, Philharmonie de Paris, Paris, France, Novembre 2015.

[5] Le futur du traitement de la voix, Future en Seine, Paris, France, Juin 2014.

[6] Machine théâtrale et nouvelles technologies, Je ne suis pas scientifique, mais... Forum Science, Recherche, et Société - Conservatoire National des Arts et Métiers, Paris, France, Mai 2014.

[7] Les technologies de traitement de la voix, La Semaine Du Son, Paris, France, Janvier 2014.

Publications

My publication indexes: **h-index: 14, i10-index : 23** (source google scholar, 07/2018)).

Articles in peer-reviewed journals (rank A)

[1] L. Benaroya, N. Obin, M. Liuni, A. Roebel, W. Rauml, S. Argentieri (2018). Binaural Localization of Multiple Sound Sources by Non-Negative Tensor Factorization. IEEE/ACM Transactions on Audio, Speech and Language Processing, en cours de publication. Facteur d'impact : 2.491

[2] N. Obin, A. Roebel (2016). Similarity Search of Acted Voices for Automatic Voice Casting. IEEE/ACM Transactions on Audio, Speech and Language Processing, vol. 24, no 9, p. 1638-1647, September 2016. Facteur d'impact : 1.877

[3] N. Obin, P. Lanchantin (2015). Symbolic Modelling of Speech Prosody: From Linguistics to Statistics. IEEE/ACM Transactions on Audio, Speech and Language Processing, vol.23, no. 3, p. 588-599, March 2015. Facteur d'impact : 1.877

Articles in peer-reviewed journals (others)

- [4] M. Avanzi, N. Obin, A. Lacheret and B. Victorri (2011). Vers une modélisation continue de la structure prosodique. Le cas des proéminences accentuelles. *Journal of French Languages Studies*, Vol. 1 (21), p. 53-71, 2011.

Articles in peer-reviewed journals (French)

- [5] G. Beller, C. Veaux, G. Degottex, N. Obin, P. Lanchantin and X. Rodet (2009). IrcamCorpusTools : plateforme pour les corpus de parole. *Traitement Automatique des Langues*, Vol. 3 (49), 2009.

Article in peer-reviewed conferences (international, rank A)

Source : <http://www.conferenceranks.com>

- [1] D. Bouvier, N.Obin, M. Liuni, A. Roebel (2016). A Source/Filter Model with Adaptive Constraints for NMF-based Speech Separation International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Shanghai, China, 2016.
- [2] R. Gong, P. Cuvillier, N.Obin, A. Cont (2015). Real-Time Audio-to-Score Alignment of Singing Voice Based on Melody and Lyric Information. *Interspeech*, Dresden, Germany, 2015.
- [3] X. Favory, N.Obin, G. Degottex, A. Roebel (2015). The Role of Glottal Source Parameters for High-Quality Transformation of Perceptual Age. *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Brisbane, Australia, 2015.
- [4] G. Degottex, N.Obin (2014). Phase Distortion Statistics as a Representation of the Glottal Source: Application to the Classification of Voice Qualities. *Interspeech*, Singapore, Singapore, 2014.
- [5] N.Obin, A. Roebel, G. Bachman (2014). On Automatic Voice Casting for Expressive Speech: Speaker Recognition vs. Speech Classification. *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Florence, Italy, 2014.
- [6] N.Obin, F. Lamare, A. Roebel (2013). Syll-O-Matic: an Adaptive Time-Frequency Representation for the Automatic Segmentation of Speech into Syllables. *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Vancouver, Canada, 2013.
- [7] N.Obin (2012). Cries and Whispers - Classification of Vocal Effort in Expressive Speech. *Interspeech*, Portland, USA, 2012.
- [8] J. Lorenzo-Trueba, R. Barra-Chicote, T. Raitio, N. Obin, P. Alku, J. Yamagishi, J. M. Montero (2012). Towards Glottal Source Controllability in Expressive Speech Synthesis. *Interspeech*, Portland, USA, 2012.
- [9] M. Avanzi, P. Dubosson, S. Schwab, N.Obin (2012). Accentual Transfer from Swiss-German to French: A Study of "Français Fédéral". *Interspeech*, Portland, USA, 2012.
- [10] N. Obin, P. Lanchantin, A. Lacheret and X. Rodet (2011). Reformulating Prosodic Break Model into Segmental HMMs and Information Fusion. *Interspeech 2011*, Florence, Italy, 2011.
- [11] N. Obin, P. Lanchantin, A. Lacheret and X. Rodet (2011). Discrete/Continuous Modelling of Speaking Style in HMM-based Speech Synthesis: Design and Evaluation. *Interspeech 2011*, Florence, Italy, 2011.
- [12] N. Obin, A. Lacheret and X. Rodet (2011). Stylization and Trajectory Modelling of Short and Long Term Speech Prosody Variations. *Interspeech 2011*, Florence, Italy, 2011.
- [13] M. Avanzi, N. Obin, Anne Lacheret-Dujour and B. Victorri (2011). Toward a Continuous Modeling of French Prosodic Structure: Using Acoustic Features to Predict Prominence Location and Prominence Degree. *Interspeech 2011*, Florence, Italy, 2011.
- [14] N. Obin, A. Lacheret and X. Rodet. HMM-based Prosodic Structure Model Using Rich Linguistic Context. *Interspeech 2010*, p. 1133-1136, Makuhari, Japan, 2010.
- [15] N. Obin, V. Dellwo, A. Lacheret and X. Rodet. Expectations for Discourse Genre Identification: a Prosodic Study. *Interspeech 2010*, p. 3070-3073, Makuhari, Japan, 2010.

- [16] N. Obin, X. Rodet and A. Lacheret-Dujour. A Multi-Level Context-Dependent Prosodic Model Applied to Durational Modeling. *Interspeech 2009*, p. 512-515, Brighton, UK, 2009.
- [17] N. Obin, A. Lacheret-Dujour, C. Veaux, X. Rodet and A.-C. Simon. A Method for Automatic and Dynamic Estimation of Discourse Genre Typology with Prosodic Features. *Interspeech 2008*, p. 1204-1207, Brisbane, Australia, 2008.
- [18] N. Obin, X. Rodet and A. Lacheret-Dujour. French Prominence: A Probabilistic Framework. *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, p. 3993-3996, Las Vegas, USA, 2008.

Article in peer-reviewed conferences (international, others)

- [1] N.Obin, J. Beliao (2018). Sparse Coding of Pitch Contours with Deep Auto-Encoders. *Speech Prosody, Poznań, Poland, 2018*.
- [2] O. Migliore, N.Obin (2018). At the Interface of Speech and Music: A Study of Prosody and Musical Prosody in Rap Music. *Speech Prosody, Poznań, Poland, 2018*.
- [3] R. Gong, N. Obin, G. Dzhambazov and X. Serra (2017). Score-Informed Syllable Segmentation for Jingju a Cappella Singing Voice with Mel-Frequency Intensity Profiles. *International Workshop on Folk Music Analysis (FMA), Malaga, Spain, 2017*.
- [4] N.Obin, J. Beliao, C. Veaux, A. Lacheret (2014). SLAM: Automatic Stylization and Labelling of Speech Melody. *Speech Prosody, Dublin, Ireland, 2014*.
- [5] A. Lacheret, J. Beliao, A. Dister, K. Gerdes, J.-P. Goldman, S. Kahane, N. Obin, P. Pietrandrea, A. Tchobanov (2014). Rhapsodie: a Prosodic-Syntactic Treebank for Spoken French. *Language Resources and Evaluation Conference (LREC), Reykjavik, Iceland, 2014*.
- [6] N.Obin, M. Liuni (2012). On the Generalization of Shannon Entropy for Speech Recognition. *IEEE workshop on Spoken Language Technology, Miami, USA, 2012*.
- [7] N.Obin, C. Veaux, P. Lanchantin (2012). Making Sense of Variations: Introducing Alternatives in Speech Synthesis. *Speech Prosody, Shanghai, China, 2012*.
- [8] N.Obin, M. Avanzi, G. Bordal, A. Bardiaux (2012). Regional Variations of Speech Rhythm in French: In Search of Lost Times. *Speech Prosody, Shanghai, China, 2012*.
- [9] G. Bordal, M. Avanzi, N.Obin, A. Bardiaux (2012). Variations in the Realization of the French Accentual Phrase in the Light of Language Contact. *Speech Prosody, Shanghai, China, 2012*.
- [10] M. Avanzi, A. Lacheret, N.Obin (2012). Transcription of Prosody in Continuous Speech: Prosodic Encoding the Spoken French Corpus Rhapsodie. *Advancing Prosodic Transcription for Spoken Language Science and Technology, LabPhon, Stuttgart, Germany, 2012*.
- [11] P. Lanchantin, S. Farner, C. Veaux, G. Degottex, N. Obin, G. Beller, Stefan Huber (2011). Vivos Voco: a Survey of Recent Research on Voice Transformations at IRCAM. *International Conference on Digital Audio Effects, Paris, France, 2011*.
- [12] N. Obin, M. Avanzi, and A. Lacheret (2011). Transcription of French Prosody: the Rhapsodie Protocole. *Interface Discours et Prosodie 2011, Manchester, UK, 2011*.
- [13] M. Avanzi, G. Bordal and N. Obin (2011). Typological Variations in the Realization of French Accentual Phrase. *International Congress of Phonetic Sciences (ICPHS), Hong-Kong, China, 2011*.
- [14] N. Obin, P. Lanchantin, M. Avanzi, A. Lacheret-Dujour and X. Rodet (2010). Toward Improved HMM-based Speech Synthesis Using High-Level Syntactical Features. *Speech Proosdy 2010, Chicago, USA, 2010*.
- [15] N. Obin, X. Rodet and A. Lacheret-Dujour (2009). A Syllable-Based Prominence Detection Model Based on Discriminant Analysis and Context-Dependency. *International Conference on Speech and Computer (SPECOM)*, p. 97-100, St-Petersburg, Russia, 2009.
- [16] G. Beller, N. Obin and X. Rodet (2008). Articulation Degree as a Prosodic Dimension of Expressive Speech. *Speech Prosody, Campinas, Brazil, 2008*.

Article in peer-reviewed conferences (French)

- [1] N. Obin , P. Pham, A. Roebel (2018). Conversion d'Identité de la Voix Chantée par Sélection et Concaténation d'Unités Spectrales. Journées d'Etude de la Parole, Aix-en-Provence, France, 4-8 juin 2018.
- [2] A. Roebel, L. Ardaillon, N. Obin (2018). Expressive Singing Synthesis using an Extended Source Filter Model. Congrès Français d'Acoustique, Le Havre, France, 23-27 Avril 2018.
- [3] Olivier Migliore, Nicolas Obin (2016). Analyser la prosodie musicale : l'exemple du rap français des débuts (1984-1991), La voix dans les chansons, Paris, France, 2016.
- [4] Olivier Migliore, Nicolas Obin, Jean Bresson (2014). Rap, Ragga et Rock Alternatif Français des Années 1980 : Vers une Musicologie Computationnelle. Journées d'Analyse Musicale (JAM), Paris, France, 2014.
- [5] A. Lacheret, J. Beliao, A. Dister, K. Gerdes, J.-P. Goldman, S. Kahane, N. Obin, P. Pietrandrea, A. Tchobanov (2014). Rhapsodie: un Treebank annoté pour l'Etude de l'Interface Syntaxe-Prosodie en Français Parlé. Congrès Mondial de Linguistique Française (CMLF), Berlin, Germany, 2014.
- [6] N.Obin, M. Avanzi, G. Bordal, A. Bardiaux (2012). A la Recherche des Temps Perdus : Variations sur le Rythme en Français. Journées d'Etude de la Parole, Grenoble, France, 2012.
- [7] M. Avanzi, N.Obin, G. Bordal, A. Bardiaux (2012). La Variation Prosodique Dialectale en Français. Données et Hypothèses. Journées d'Etude de la Parole, Grenoble, France, 2012.
- [8] N. Obin, Modélisation du Style en Synthèse de Parole par HMM, Journées Jeunes Chercheurs en Audition, Acoustique musicale et Signal audio, Paris, France, 2010.
- [9] M. Avanzi, N. Obin, A. Lacheret-Dujour. Vers une modélisation continue de la structure prosodique du français : le cas des prééminences accentuelles. Colloque Conscila, Paris, France, 2009.
- [10] N. Obin, X. Rodet and A. Lacheret-Dujour. Un modèle de durée des syllabes fondé sur les propriétés syllabiques intrinsèques et les variations locales de débit. Journées d'étude de la parole (JEP), p. 333-336, Avignon, France, 2008.
- [11] N. Obin, J.-P. Goldman, M. Avanzi and A. Lacheret-Dujour. Comparaison de trois outils de détection automatique de prééminences en français parlé. Journées d'étude de la parole (JEP), p. 85-88, Avignon, France, 2008.

Chapters in edited monographs

- [1] N. Obin, A. Lacheret, C. Veaux, B. Victorri (2018). Tonal Annotation: Stylization of Complex Melodic Contours over Arbitrary Linguistic Segments. Rhapsodie: a prosodic-syntactic Treebank for Spoken French, Benjamins, Amsterdam, en cours de publication, 2018.
- [2] M. Avanzi, N. Obin, G. Bordal, J. Sauvage (2018). The Annotation of Prominences and Disfluencies. Rhapsodie: a prosodic-syntactic Treebank for Spoken French, Benjamins, Amsterdam, en cours de publication, 2018.
- [3] N. Obin, C. Veaux, P. Lanchantin (2015). Exploiting Alternatives for Text-To-Speech Synthesis: From Machine to Human. Speech Prosody in Speech Synthesis: Modeling and generation of prosody for high quality and flexible speech synthesis Chapter 3: Control of Prosody in Speech Synthesis, p.189-202, Springer Verlag, February, 2015.
- [4] X. Rodet, G. Beller, N. Bogaards, G. Degottex, S. Farner, P. Lanchantin, N. Obin, A. Roebel, C. Veaux and F. Villavicencio. Transformation et Synthèse de la Voix Parlée et de la Voix Chantée. Parole et Musique: Aux Origines du Dialogue Humain, Odile Jacob, Paris, 2009.