European Research Council

ERC Starting Grant 2013
Research proposal [Part B1]

CREAM: Cracking the Emotional Code of Music

PI: Jean-Julien AUCOUTURIER

Host institution: Comité National de la Recherche Scientifique (CNRS), UMR 9912 (IRCAM)

Title: Cracking the Emotional Code of Music

Short name: CREAM

Duration: 60 months

ABSTRACT

This project aims to "crack" the emotional code of music, i.e. to provide, for the first time, a precise characterization of **what type of music signal is able to activate** one mechanism of emotion induction or another, one neural pathway or another, **one emotion or another**. Research into this problem so far has been mainly correlating indistinct emotional reactions to relatively uncontrolled musical stimuli, with much technical sophistication but to little avail. Project CREAM builds on the PI's unique bi-disciplinary career spanning both computer science and cognitive neuroscience, to propose a radically **novel approach**: instead of using audio signal processing to simply *observe* musical stimuli a posteriori, we will harvest a series of recent developments in the field in order to build **powerful new tools of experimental control**, able *to engineer* musical stimuli that can activate specific emotional pathways, while inhibiting others.

For instance:
- because music is hypothesized to invade neural circuits involved in the emotional evaluation of speech, we will develop precisely-calibrated acoustic manipulations to create musical stimuli which e.g. tremble like anxious voice or exult like joyful speech, and use these stimuli to guide our exploration of such circuits
- as music is also hypothesized to trigger brainstem reflex reactions because of its sensory similarity to some survival-relevant environmental sounds, we will develop a signal processing model of the auditory cells of the human brainstem and use it to create musical stimuli which sound like e.g. impact sounds or animal roars, then use these stimuli to trigger and study such reactions.

By combining this **creative use of new technologies** with a **well-concerted mix of methods** from experimental psychology and cognitive neuroscience (incl. psychoacoustics, fNIRS brain imaging, EEG/ERP paradigms, intercultural studies, infant studies), project CREAM will produce the first functional description of the neural and cognitive processes involved in the induction of emotions by music, and establish new avenues for interdisciplinary research between the life sciences and the information sciences.

But most spectacularly, the fundamental breakthroughs brought by project CREAM will **unlatch the therapeutic potential of musical emotions**. Music will become a cognitive technology, with algorithms able to "engineer" it to mobilize one neuronal pathway or another, non-intrusively and non-pharmacologically. Within the proposed 5-year plan, support from the ERC will allow to implement a series of **high-impact clinical studies** which are direct applications of our findings, incl. linguistic rehabilitation of aphasic stroke victims with musical stimuli which sound like voice; emotion modulation for fibromyalgia patients with affective disorders, inhibition of the consolidation of traumatic memories in PTSD victims.

Section a: ***Extended Synopsis of the project proposal (max 5 pages)***

> *"Music can lift us out of depression or move us to tears - it is a remedy, a tonic, orange juice for the ear. But for many of my neurological patients, music is even more - it can provide access, even when no medication can, to movement, to speech, to life. For them, music is not a luxury, but a necessity."*
>
> - Oliver Sacks, *Musicophilia: Tales of Music and the Brain*

Music has incredible power on our emotions: it can make the lone listener cry with joy or sorrow, and a crowd of thousands jump and shout as one (Mannes, 2011). For the researcher, this power holds the key to understand the evolution and development of some of our most advanced cognitive capacities: language, empathy and social cohesion (Mithen, 2007). For the clinician, the emotional power of music holds promise for therapeutic applications ranging from linguistic rehabilitation, regulation of affective disorders and even stimulation of neural plasticity (Sacks, 2009).

If we could only understand it, that is.

However, despite years of research, the question of how music induces emotions, of what *emotional code* is spoken by music to our brains, remains one of the most vexing mysteries in modern cognitive neuroscience. Classical cognitive psychology has much to say about how one may cry, run and sweat upon encountering a bear in the woods (James, 1884). But why should a Beethoven or a Lady Gaga even have the same effect on us?

Research into this question so far has been mostly correlative. On the one hand, functional explorations of the reactions obtained to emotional musical stimuli have revealed wide networks of brain activity in both the sub-cortical regions of the limbic system (e.g. thalamus, hippocampus, amygdala - Blood & Zatorre, 2001) and cortical regions such as the pre-frontal and orbitofrontal cortices and the insula (Koelsch et al., 2006). On the other hand, computerized acoustical analysis of the musical stimuli causing emotional reactions has implicated a wealth of musical features including pitch, tempo, intensity or timbre (Liu & Zhang, 2006) but also physical properties of the signal that seem to have little musical justification, such as Alluri and Toiviainen's (2010) third statistical moment of the energy spectrum ("spectral skewness").

Yet, despite ever-increasing sophistication on both fronts, nothing seems to link these two domains of evidence - in fact, the more we know, the less we seem to understand. Looking for a fixed neural architecture, neuroscientists are instead finding vast and distributed networks of activity (Koelsch et al., 2006). Looking for hemispheric specialization, we're finding left preference (Schmidt & Taylor, 2001), or right (Blood & Zatorre, 2001), or both (Peretz & Zatorre, 2005). Looking for dedicated circuits, we're finding large overlap with regions involved in the emotional interpretation of expressive speech (Juslin & Laukka, 2003) or environmental sounds (Blumstein, Bryand and Kaye, 2012). Urged to find mathematical formulas to regress emotional reactions on some physical property of the signal, state-of-art signal processing software offers such incomprehensible factors as "*the entropy of the period of the magnitude of the maximum peak detected every 50 milliseconds in the signal's chromagram*" (Aucouturier & Bigand, 2012).

The problem is quite obviously more complicated that what the current approach presumes: first, it seems realistic to admit that music is able to activate not only one, but more likely several, functionally distinct mechanisms of emotion induction: it may prompt emotional images, remind emotional memories of past events, surprise by a particularly loud and sudden sound, etc. (see Juslin & Vastfjäll, 2008 for a tentative list of such mechanisms). Yet no methodological tool exists to selectively activate one mechanism or the other; most likely, we experimentalists trigger the whole range indiscriminately. Second, the sonic characteristics offered by modern signal processing (e.g. via the emerging discipline of *Music Information Retrieval* - MIR) are not designed for their biological or cognitive plausibility, but rather for their information-theoretical power in machine-learning applications (Pachet & Roy, 2009). Correlate, they do alright. Explain, they do very little.

Our proposal offers **a radically novel approach**: we propose to reposition state-of-art MIR technology as a means of experimental control, used to construct/select musical stimuli that will selectively activate some mechanisms of emotion induction while inhibiting others. This is a profound change of paradigm in how signal processing is used in cognition research: not to *observe* stimuli of interest a posteriori (i.e. signal processing as a sophisticated probe), but to *generate* stimuli for experimental control (i.e. signal processing as a musical scalpel).

Creative use of new technologies

The goal of the project is to establish a general methodology, and a set of computational and experimental tools, able to crack the emotional code of music, i.e. to characterize what exact type of musical signal is likely to activate this emotional mechanism or another, this brain area or another, this emotion or another. This new methodology will allow to explore possibly all the different types of emotion induction by music. However, as the first steps in this research program, we propose to focus on two specific mechanisms, for which state-of-art signal processing techniques offer **particularly creative solutions**:

- First, it has been suggested (Juslin & Laukka, 2003; Escoffier et al., 2012) that musical expression could be, partly, an amplification of the emotional expression involved in spoken language: music's trembling notes, hesitating phrases, bright or dark timbres may well be "heard" and processed "as if" they were emotional speech. However, so far, no methodology has allowed to test this mechanism empirically: we do not know what type of musical signal is able to activate such vocal circuits, or even if such stimuli exist at all. Interestingly, technology now exists that can manipulate the acoustic features of a vocal signal in real-time so as to "color" it with an emotional expression that wasn't intended when it was uttered (Bulut & Naranayan, 2008). For instance, adding vibrato (a modulation of the fundamental frequency of voice) on a neutral voice can make it sound more negative and anxious. Such acoustic manipulations are domain-agnostic: they can be used to process musical signals just as well as speech signals. We therefore propose to make a complete psycho-acoustical characterization, using voice stimuli, of such manipulations, measuring e.g. what exact modulation speed in vibrato imparts what emotional impression on the listener; then to apply the same manipulations to musical stimuli, i.e. **to create music that trembles like anxious voice or exults like joyful speech**, and to compare the resulting emotional reactions. Using such voice-transformed musical stimuli, we will be able, first the first time, to selectively activate, with music, the emotional circuitry involved in processing expressive speech.

- Second, it has been proposed (Goydke et al., 2004; Blumstein, Bryand and Kaye, 2012) that certain musical sounds, loud, fast or dissonant, trigger the same reflex emotional reactions that negative, survival-relevant environmental sounds would provoke. As for voice previously, this mechanism has not been explored empirically: as of yet, we do not know what type of musical signal is able to activate this type of emotional reaction. However, as before, new signal processing techniques have been recently introduced that can help here: it is now known that, under certain circumstances, sparse coding representations of audio (Plumbley et al., 2010) can simulate the sensory properties of cells in the human peripheral auditory system (the so-called *spectro-temporal receptive fields* - Chi, Ru & Shamma, 2005), the very same cells that are involved in the processing of environmental sounds (Ghazanfar & Nicolelis, 2001). We therefore propose to make a complete psychoacoustical characterization, using environmental sound stimuli, of such neuromimetic signal representations, measuring e.g. what exact spectro-temporal pattern in an impact sound or an animal roar imparts what emotional reaction on the listener; then to select musical sounds from a large dataset which present auditory characteristics that are maximally similar to these environmental sounds, i.e. **to create music that sound like impact sounds or animal roars** (from the point of view of the computational model), and to compare the resulting emotional reactions. As for voice before, such noise-similar musical stimuli will allow to selectively activate, with music, the emotional circuitry involved in processing environmental sounds, so we can explore its functionality.

Additionally, because these two mechanisms in particular are associated to fast perceptive processes (Juslin & Vastfjäll, 2008), we will use the general methodology to first work on very short musical stimuli (less than one second), and then to study their temporal integration into longer textures. Starting with short samples preserves the emotional effects of the mechanisms under study (Bigand et al., 2005), while inhibiting other, possibly confounding mechanisms that could be at play, such as syntactic expectancy (Meyer, 1953), and which require longer time-scales.

In summary, in order to study emotions in minute-long music sequences, we instead propose to study emotions in 500-millisecond bits of speech and environmental sounds. This may seem highly indirect, but this is in fact the first principled research program able to crack into the complexity of the general problem. This is only made possible by the creative use of modern signal-processing techniques, which have never been thought as a means of experimental control until now.

A well-concerted mix of experimental methods

The strength of our approach is to combine the state-of-art sophistication of audio signal processing with an experimental logics typical of cognitive psychology and neuroscience. By focusing on very short sounds, we inhibit all possible emotional induction mechanisms but the fastest ones. Further, by using computational techniques such as vocal transformations and neuro-mimetic modeling, we selectively activate the two target mechanisms of our study.

However, as creative as it is, this experimental device alone would not be sufficient. Even with our computationally manipulated stimuli, both mechanisms may still be activated simultaneously (with e.g. stimuli that both sound like voice and environmental sounds), and other mechanisms may also be involved (e.g. fear conditioning - Juslin & Vastfjäll, 2008). To control for such confounding mechanisms, our proposal utilizes a well-concerted mix of experimental methods:

- brain imaging: both our target mechanisms are highly modular, and correspond to well-identified brain areas. For each mechanism, we will therefore test predictions of increased activation in the corresponding ROIs, using **functional near-infrared spectroscopy** (fNIRS), as well as predictions of ERP patterns, using an **EEG mismatch negativity** (MMN**)** paradigm.
- intercultural study: there is a well-identified intra-group advantage to recognizing emotional prosody in one's own linguistic culture. This effect will likely be observed for voice-manipulated music, but not in other types of mechanisms. We will therefore test emotional reactions to our stimuli on **non-western subject populations** (Indian Kannad/Tamil speakers), and compare these reactions to those of Western subjects (French and English speakers). This will be a collaboration with a neuroscience institute in Bangalore, India.
- infant study: contrary to other, learned mechanisms, our two mechanisms of focus (voice and environmental sounds) have relatively early ontogenetic development - we will therefore ascertain the emotional reactions to our constructed material on **neonates, using a non-nutritive sucking rhythm paradigm** (in collaboration with a specialized laboratory in Tokyo, Japan).

By intersecting such methods (e.g. observing short, voice-similar musical stimuli with intercultural variation, left temporo-frontal activation and gradual 0-12-month development), we will be able to ascertain, with an unprecedented degree of experimental control, that we are truly able to isolate each mechanism experimentally.

A background, environment and network of collaborations that puts it within reach

The project builds on the PI's unique bi-disciplinary career, which combines expertise in both audio signal processing and cognitive neuroscience.

Our work on audio signal processing has lead us, over the years, to introduce several influential computational models for signals of music (e.g. the bag-of-frame algorithm to compute timbre similarity between long musical sequences, which is now standard to the field of Music Information Retrieval - JNRSAS 2004, 270 citations), environmental sounds (e.g. extension of the same paradigm to environmental textures, or "soundscapes" - JASA 2007, 95 citations) and human voice (e.g. a segmentation algorithm for expiratory and inspiratory sounds in baby vocalizations, JASA 2011). While the two technologies to be developed in the project (vocal transformations and spectro-temporal neuromimetic models) are not the continuation of any of our previous work, they fall in the center of our core expertise, where we already have a proven track record of impactful work.

All the empirical methodologies to be used in the project are those in which I have gathered experience during my postdoctoral and PI years in cognitive neuroscience research units, albeit in projects which either did not focus on musical signals or emotional processes: the current project is a novel and creative application of these methodologies to a new, emerging topic. I have for instance studied the psychoacoustics of short environmental sounds to study categorization (JASA, 2009); used EEG measures and ERP paradigms to study interactions between music and osteopathic treatment (Psychophysiology, under review, 2012); Near-Infrared Spectroscopy (NIRS) brain imaging to study the context effect of music and environmental sounds on word encoding in memory (in prep., 2012); and physiological measures (galvanic skin response) to study emotional influences on cognition (in prep., 2012).

In addition to this direct expertise, I will surround myself and my team with a scientific and technological environment which is closely tailored to the need of such interdisciplinary research. The project will be implemented on two research sites*, distant 250km apart in France, each a world-leader in one of the project's core discipline:

- the primary site, the Institute of Research/Coordination in Musical Acoustics (IRCAM), is based in central Paris (http://www.ircam.fr), and is the world-leading research center for musical informatics, with notable expertise in the analysis and synthesis of audio signals and Music Information Retrieval (IRCAM was the organizer of the 2002 International ISMIR conference). Without direct cost, IRCAM will provide facilities for psychoacoustics, audio engineering and signal processing.
- the secondary site, the Learning and Developmental Psychology Laboratory (LEAD), is based in University of Burgundy in Dijon (http://leadserv.u-bourgogne.fr). It is the only French psychology laboratory devoted to the cognitive neurosciences of music, and the European coordinator of several high-profile European projects on clinical applications of music (incl. FP7 EBRAMUS). Without direct cost, LEAD will provide facilities for experimental psychology, EEG and NIRS imaging, as well as institutional partnerships with university hospitals for clinical studies.

At each institution, my research team and I will be able to seek punctual, one-off expertise from several key collaborators, at various critical points of the project:

- for algorithmic aspects: Dr. Geoffroy Peeters and Dr Axel Röbel (senior researchers, IRCAM), whose team is a world-leader in voice synthesis technologies; Dr. Daniel Pressnitzer and Prof. Shihab Shamma (ERC Advanced grant recipients, 2011, Ecole Normale Superieure), who are pioneers in neuromimetic auditory models.
- for brain imaging studies: Dr Aurelia Bugaiska (Assistant professor of cognitive psychology, LEAD), an expert in the functional exploration with NIRS of frontal activity in healthy adults; Dr Bénedicte Poulain-Charronnat (CNRS Senior researcher, LEAD, Dijon), an expert in music ERP studies.
- for clinical studies: Prof. Emmanuel Bigand (Professor of Cognitive Psychology, LEAD, Dijon), expert of linguistic rehabilitation with music and our proxy for clinical studies with the Cerebro-vascular accident unit at Dijon Univ. Hospital; Dr Gérard Mick, M.D. Rheumatology (Associate Researcher, LEAD, Dijon), expert of pain and affective disorders in fibromyalgia, and our proxy for clinical studies with Rheumatology units at Dijon & Lyon Univ. Hospitals

Finally, owing to my international career with 5 years spent in Japan, the project builds on the unique opportunity of international collaborations with groups that I have visited in the past: that of Prof. Kazuo Okanoya in Riken Brain Science Institute in Saitama, Japan (for infant developmental studies) and of Dr. Shantala Hegde in National Institute for Mental Health and Neuroscience, in Bangalore, India (for intercultural studies). The project will fund several research stays for team members at these 2 institutions.

Impact, potential to change research fields and open new interdisciplinary avenues

If successful, the project will impact the field of music cognition to the point of changing both *what* we know of musical emotions, but also *how* we acquire this knowledge. The project will:

- stop the current trend of using MIR techniques to correlate musical stimuli to their indistinct emotional reactions, a strategy that after 10 years of development has remained mostly fruitless, and
- reposition MIR as a powerful means of experimental control, able to generate stimuli that can selectively activate specific mechanisms of musical emotions.

This new methodological framework is a profound change of paradigm, which has potential to change the field of music cognition and establish **new avenues for interdisciplinary research** between the life sciences and the information sciences and technologies.

In the life sciences, this new paradigm will generate, for the first time in the history of the field, a computational characterization of what exact type of musical stimuli is able to activate one emotional pathway or another, one emotion or another. In other words, we will be able to create musical stimuli tailored to induce a specific emotion in a specific individual. This will provide **disruptive new tools for emotion research**, including:

* Both sites are managed by the same host institution, the French CNRS (Centre National de la Recherche Scientifique), with whom I am tenured researcher since October 2012.

- musical stimuli optimized to activate a given brain area, to subject it to functional imaging (the equivalent of variable frequency-modulated speech in e.g. Zatorre & Belin, 2001)
- normed musical stimuli optimized to generate a given emotion, to study e.g. emotional biases or emotional influences on cognition. (the equivalent of Lang, Bradley & Cuthbert's (2008) IAPS)

In the information sciences, the new formulation of MIR signal processing as a tool for experimental control in cognitive neurosciences will trigger programs of research aiming to establish novel algorithms with better biological and cognitive validity. This will initiate in audio research the same movement that changed the field of image processing in the past 5-10 years (Serre et al., 2007). MIR algorithms will evolve to become **more robust, more tailored to individual listeners, and more efficient**, all challenges currently faced by the community without much of a program to address them. (Aucouturier & Bigand, 2012; best presentation award at ISMIR 2012).

Most importantly perhaps, the fundamental breakthroughs brought by this project will unlatch the therapeutic potential of musical emotions. Music will become a cognitive technology, with algorithms able to "engineer" it to mobilize one neuronal pathway or another, non-intrusively and non-pharmacologically. Within the proposed 5-year plan, support from the ERC will allow to implement a series of **high-impact clinical studies** with are direct applications of our findings.:

- to stimulate, with musical stimuli which sound like voice, the language areas of the brain, for use in linguistic rehabilitation after strokes (proof of concept: Särkämö et al. 2008)
- to modulate emotions unconsciously (using non-detected vocal manipulations) in patients with affective disorders (proof of concept: Popova et al., 1998)
- to inhibit the consolidation of traumatic memories in patients with post-traumatic stress disorders (proof of concept: Rickard, Wong & Velik, 2011)

"Cracking the emotional code of music" therefore brings, for the first time, the very actual possibility of channeling and exploiting the formidable power that music has, on our neurons, on our minds, and our hearts.

Alluri, V & Toiviainen, P. (2010) Exploring perceptual and acoustic correlates of polyphonic timbre. Music Perception 27(3), 223-41

Bigand, E., Vieillard, S., Madurell, F., Marozeau, J. & Dacquet, A. (2005) Multidimensional scaling of emotions responses to music: The effect of musical expertise and of the duration of the excerpts. Cognition & Emotion, 19, 1113-39

Blood, A. J. & Zatorre, R. J. (2001). Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. Proceedings of National Academy of Sciences, 98, 11818-23

Blumstein, D., Bryant, G. & Kaye, P. (2012) The sound of arousal in music is context-dependent. Biology Letters, 8(5), pp.744-7.

Bulut, M., & Narayanan, S. S. (2008). F0 modifications in emotional speech. J. Acoust. Soc. Am., 123(6), 4547-4558.

Chi, T., Ru, P. & Shamma, S. (2005) Multiresolution spectrotemporal analysis of complex sounds J. Acoust. Soc. of Am.; vol. 118 (2), pp.887-906

Escoffier, N., Zhong, J., Schirmer, A. and Qiu, A. (2012), Emotional expressions in voice and music: Same code, same effect?. Human Brain Mapping.

Ghazanfar & Nicolelis (2001), The Structure & Function of Dynamic Cortical & Thalamic Receptive Fields, Cerebral Cortex

Goydke, KN, Altenmüller, E, Möller, J & Münte, TF (2004) Changes in emotional tone and instrumental timbre are reflected by the mismatch negativity. Brain Res Cogn Brain Res. 21(3):351-9

James, William (1884). What is an emotion? Mind, 9, 188-205

Juslin, P. & Laukka, P. (2003) Communication of emotions in vocal expression and music performance: Different channels, same code? Psychological Bulletin, 129, 770-814

Juslin, P. & Västfjäll, D. (2008) Emotional responses to music: the need to consider underlying mechanisms. Behav. & Brain Sciences, 31, 559-621.

Koelsch, S., Fritz, T., von Cramon, D; Y., Müller, K. & Friederici, A. D. (2006). Investigating emotion with music: An fMRI study. Human Brain Mapping, 27, 239-50

Lang, P.J., Bradley, M.M., & Cuthbert, B.N. (2008). International affective picture system (IAPS): Affective ratings of pictures and instruction manual. Technical Report A-8. University of Florida, Gainesville, FL.

Liu, D. & Zhang, H.-J. (2006) Automatic mood detection and tracking of music audio signal, IEEE Tr. Speech and Audio Proc. 14(1)

Mannes, Elena (2011). The Power of Music: Pioneering Discoveries in the new science of song. Walker & Company

Meyer, L. B. (1956) Emotion and Meaning in music. Chicago: CUP

Mithen, Steven (2007). The Singing Neanderthal: The Origins of Music, Language, Mind, and Body. Cambridge: Harvard University Press

Pachet, F. & Roy, P (2009) Analytical Features: a Knowledge-Based Approach to Audio Feature Generation. EURASIP Journal on Audio, Speech, and Music Processing, vol. 2009(1)

Peretz, I & Zatorre, R. J. (2005) Brain organization for music processing. Annual Review of Psychology, 57, 27-53

Plumbley, M., Blumensath, T., Daudet, L., Gribonval, R. & Davies, M. E. (2010) . Sparse representations in audio and music: from coding to source separation. Proceedings of the IEEE 98, 995-1005

Popova E, Mikheev V, Shuvaev V, Ivonin A, Chernyakov G. (1998) Functional rearrangements in the human brain during emotional self-regulation with biological feedback. Neurosci Behav Physiol. 28(1):8-16.

Rickard, N.,Wong, W. & Velik, L.(2011) Relaxing music counters consolidation of emotional memory Neurobiol Learn Mem.

Sacks, Oliver. (2008) Musicophilia: Tales of Music and the Brain. Vintage.

Särkämö, T., Tervaniemi, M., Laitinen, S., Forsblom, A., Soinila, S., Mikkonen, M., Autti, T., Silvennoinen, H., Erkkilä, J., Laine, M., Peretz, I. & Hietanen, M. (2008) Music listening enhances cognitive recovery and mood after middle cerebral artery stroke. Brain vol. 131, pp. 866-876

Schmidt, L.A. & Trainor, L.J. (2001) Frontal brain electrical activity distinguishes valence and intensity of musical emotions. Cognition & Emotion, 15, 487-500

Serre, T., Wolf, L., Bileschi, S., Riesenhuber, M. & Poggio, T.(2007). Object recognition with cortex-like mechanisms. IEEE Transactions on Pattern Analysis and Machine Intelligence, 29 (3), pp. 411-426 , 2007

Zatorre, R.J. and Belin, P. (2001) Spectral and temporal processing in human auditory cortex. Cereb. Cortex 11, 946–953

Section b: Curriculum vitae

Jean-Julien Aucouturier, CNRS researcher (Cognitive Science). Age: 33; married, 3 children.

- From 2012: Researcher (**Cognitive Science**), CNRS (FR)
- 2009-2011: Director, Science Program & Assistant Professor (Computer Science/**Psychology**), Temple University, Japan Campus (JP)
- 2008-2011: Researcher, RIKEN **Brain Science** Institute, Tokyo (JP)

- 2006-2008: Postdoc, Institute of **Physics**, University of Tokyo (JP)
- 2006: PhD in **Computer Science**, University Paris 6
- 2001-2006: Assistant Researcher, SONY **Computer Science** Laboratory (FR)
- 2000-2001: MSc. **Audio processing**, King's College University of London (UK)
- 1998-2001: Ecole Supérieure d'**Electricité** (Supélec)

I have a radically bi-disciplinary career, with equal expertise in **computer science** and **cognitive neuroscience**.

I was trained as a signal processing engineer (in Supélec, France's 4th-ranked "Grande Ecole"). With a passion for sound and music, I then sought the mentoring of Prof. Mark Sandler at King's College London, who in 2000 was one of the early proponents of the nascent field of Music Information Retrieval (MIR). The field has grown, and so have I with it, for the best part of my graduate studies. In 2001, I joined the team of François Pachet (an ERC Advanced grant recipient, 2011) at Sony's Computer Science Labs in Paris, where I headed the development of Sony's experimental music distribution platform, the MusicBrowser. I subsequently obtained a PhD of computer science from University of Paris 6, in Sept. 2006.

The last months of 2006 were a turning point in my career. Within the same week, I received an offer for an lectureship in computer science in Prof. Sandler's group in Queen Mary University of London, and an invitation for a postdoctoral fellowship in Tokyo to work with a physicist (Prof. Takashi Ikegami) applying non-linear dynamics to model human cognition. My PhD work, while rooted in computer science, had convinced me that we needed better cognitive models. I rejected London, and embarked for Tokyo.

In the 6 years from 2006 to 2012, I gradually became an experimentalist and a cognitive neuroscientist. My two-year postdoctoral stay at University of Tokyo familiarized me with neuro-mimetic modeling, and culminated in a collaboration with a Japanese robot maker and an award in an international Artificial Life conference. However, I soon became frustrated with the position of remaining, first and foremost, a computer scientist, delegating experimental work to biologists and psychologists. In the field of music cognition, either you were a modeler, or you were an experimentalist. Truly, I wanted to be both.

In 2008, I therefore accepted a researcher position in a neuroscience institute, RIKEN BSI, where I got training in a variety of experimental techniques (behavioral measures, electrophysiology, brain imaging with fNIRS, EEG), combining them with audio signal processing to analyze animal vocalizations. In 2009, in recognition for this double expertise, I was appointed as an Assistant Professor, crossed-listed in both computer science and psychology, in Temple University (Philadelphia, PA), with the task of developing the Science Program of its international campus in Tokyo. In Temple, I lead a faculty of five (with three additional hires during my tenure), developed a series of courses in both disciplines, established an undergraduate research program which published more in one semester than the rest of the Department's faculty that same year, and obtained several American and Japanese research grants, the first ones in the campus' short history.

My family and I left Japan for France in the Spring of 2011in the wake of the Tokyo earthquake. In the Spring of 2012, I was ranked #1 nationally for the position of researcher in Cognitive Science at CNRS, and am now dividing my work between a computer science institute (IRCAM in Paris) and a cognitive psychology laboratory (LEAD, in Dijon). It is in this context (my establishing a new base in Europe for my work as a PI after 5 years spent in Japan) that I am applying for an ERC starting grant.

Funding record

- Innovative teaching grant (10,000 GBP, 1 year), co-PI, University College, London, 2011
- Seed Research Initiatives (50,000 USD, 1 year), co-PI, Office of the Provost, Temple Univ., 2010
- CARAS Research grant (7,000 USD), PI, Office of the Provost, Temple Univ., 2009
- Grant-in-aid (1,6M JPY, c. 15,000 USD), PI, Japanese Society for Promotion of Science, Japan, 2008

Fellowships

- Long-term Post-doctoral Fellowship for Foreign Researcher, Japanese Society for Promotion of Science, Japan, 2006-2008
- AHRC Research Fellowship, Centre for Cross-Cultural Music & Dance Performance, School of Oriental and African Studies, University of London, UK, 2006.
- CIFRE Doctoral Fellowship, ANRT, France, 2003-2006
- Excellence in Engineering Fellowship, George Besse Foundation, France, 2001
- Entente Cordiale fellowship, British Council, 2001

Teaching and Research Leadership

- Between 2008-2011, I taught over 800 hours of college courses in the two departments of Computer Science and Psychology, Temple University Japan (12 undergraduate courses, 2 graduate courses)
- From 2008-2011, I directed the Science Program in Temple University, supervised a faculty of 5, and was responsible for curriculum and research development. I was the campus representative for Temple University's educational technology reform initiative TLTR2.
- In 2010, I designed an award-winning undergraduate research program (the 3m10p program - http://jjtok.io/3m10p) and supervised research projects by 8 students, leading to the submission of 3 articles in international psychology journals
- In 2009, I supervised 3 honors undergraduate students, leading to 2 publications in international journals and 1 international conference presentation.
- From 2003-2006, I was technical manager for one consortium partner in the FP6 IST project SemanticHIFI (team of 4, 120 m.m. over 3 years)

Academic Service

- Conference organization: Chair, Demos and Late Breaking Results, International Conference on Music Information Retrieval (ISMIR, 2012)
- Associate Editor: Pattern Recognition Letters (Special Issue: "Future Trends in the Pattern Recognition of Non-Speech Audio", vol. 31(12), 2010); Journal of New Music Research (Special Issue: "Music Information Retrieval in the era of folksonomies", vol. 32(7), 2008); IEEE Intelligent Systems (Special Issue: "Cheek to Chip: Dancing robots and the future of AI", vol. 23(2), 2008)
- Review committee (journals): IEEE Multimedia Magazine; IEEE Intelligent Systems Magazine; IEEE Transactions on Audio, Speech and Language Processing; IEEE Transactions on Signal Processing; Pattern Recognition Letters; Pattern Recognition; Machine Learning; Journal of Acoustical Society of America; Journal of Applied Signal Processing; Journal on Audio, Speech and Music Processing; Journal of New Music Research.
- Review committee (conferences): Int. Conference on Music Information Retrieval (ISMIR, 2001-2011); Int. Conference on Neural Information Processing (ICONIP, 2008-2010); Neural Information Processing Systems (NIPS, 2006-2008); Sound and Music Computing Conference (SMC, 2009-2010); Workshop on Learning the Semantics of Audio Signals (LSAS, 2006-2008)

Section c: Early achievements track-record (max 2 pages)

- Over **50 publications** in computer science (2001-2007) and cognitive neuroscience (2008~)
- Over 1600 citations, **H-index= 18** (source: Google scholar)
- **Second most-cited** MIR article of all times (218 cites, source: Lee, Jones & Downie, 2009)

My scientific contributions have impacted the fields of computer science (Music Information Retrieval) and cognitive neuroscience (music cognition), mirroring my transition from the former to the latter.

My early work is deeply connected to the emergence, in the years 2000-2001, of the field of Music Information Retrieval (MIR). In a series of articles during my PhD, I established a novel paradigm, the bag-of-frames, to compute acoustic similarities between whole music songs. This paradigm was to become foundational to a large part of the field: the task of music similarity as we defined it is part of the annual MIR algorithm evaluation campaign (MIREX) since 2006 (our algorithm won the first prize that same year); our 2002 seminal article on similarity is now the 2nd most-cited article in the community (source: Lee, Jones & Downie, 2009); on Google Scholar, the expression "bag-of-frames" now appears in more than 250 publications; this year still, I will be chairing the "music similarity" session at the field's main conference, ISMIR 2012.

The end of my PhD, and the beginning of my postdoctoral work, were marked by several discoveries which have come to question the paradigms that had by then become standard to MIR. Our 2004 JNRSAS study was the first to identify a structural "glass-ceiling" effect in the performance of bag-of-frames algorithms. With 270 citations, this is our most-cited work to date, and the expression "glass-ceiling" has become trademark for all research trying to bridge the semantic gap between purely physical representations of the signal and complexity of human musical judgements (e.g. Nanopoulos et al. "Looking through the glass-ceiling", in Proc. ISMIR 2010). In 2008, our Pattern Recognition study (35 citations) demonstrated that type-I errors in similarity algorithms were distributed as a power law, leading a small number of items (which we proposed to call "hubs") to concentrate the majority of errors. Following our characterization, and several replications (e.g. Flexer et al., in proc. ISMIR 2009), the "hub problem" quickly spread outside of the MIR community, and has now become an important topic of study in machine learning in general (e.g. most recently Radovanovic et al. "Hubs in space: popular nearest neighbors in high-dimensional data", Journal of Machine Learning Research, 2010).

With my subsequent affiliations as a postdoc, then a PI, my expertise in cognitive neuroscience grew and my contributions turned to focus on cognitive modeling and experimental studies: to be noted, our empirical study of self-organization in musical tuning systems (best paper award at the 2007 European Artificial Life Conference, 1 journal article, 1 book chapter); a neuro-mimetic model for rhythmic entrainment (2 journal articles, showcase in the Tokyo Apple Store with CNET media coverage); and most recently, a series of articles combining MIR technology (JASA, 2011) and behavioural techniques (Psychophysiology, under review) to explore emotional influences on cognition (Emotion, under review).

Through this early career, I have become internationally recognized as a bi-disciplinary leader in approaches combining audio signal processing with cognitive neuroscience. As an illustration, this Fall 2012, I have been both an invited speaker at an engineering conference, the International Conference of Music Information Retrieval, about "*the difficult dialog of MIR and cognitive psychology*" (a talk which received the conference's best presentation award) and at a psychology conference, the Auditory Perception SIG meeting of the French Acoustical Society, about "*new signal processing tools for music cognition*". This bi-disciplinary trajectory culminates in the present proposal, CREAM, which intends no less than to durably change both research fields and to establish a new interdisciplinary approach to solve one of the most enduring mysteries in modern cognitive neuroscience: how music creates emotions.

Publications:

My complete list of publications includes 4 journal manuscripts under review, **22 published journal articles** (incl. 18 as first author, 13 without PhD supervisor F. Pachet), 2 book chapters, **28 conference articles** (incl. 19, as first author, 19 without PhD supervisor). Available online on http://jjtok.io/publications.htm

Five Most-impactful: (Over 1600 citations, **H-index=18**)

- **270 citations:** Aucouturier, J.-J. and Pachet F. Improving Timbre Similarity: How high is the sky?. Journal of Negative Results in Speech and Audio Sciences, 1(1), 2004.
- **258 citations:** Aucouturier, J.-J. and Pachet, F. Representing Musical Genre: A State of the Art. Journal of New Music Research, 32(1), 2003.
- **218 citations:** Aucouturier, J.-J. and Pachet, F. Music Similarity Measures: What's the Use?. Proceedings of the International Symposium on Music Information Retrieval (ISMIR), October 2002, Paris, France **(Second most-cited ISMIR article of all times, 2009)**
- **95 citations:** .Aucouturier, J.-J., Defreville, B. and Pachet, F. The bag-of-frame approach to audio pattern recognition: A sufficient model for urban soundscapes but not for polyphonic music. Journal of the Acoustical Society of America, 122(2):881-91, 2007.
- **78 citations:** Aucouturier, J.-J., Pachet, F. and Sandler, M. Timbre Models For Analysis and Retrieval of Polyphonic Music Signals. IEEE Transactions of Multimedia, 7(6):1028-1035, 2005.

Recent work:

- Caballé, J., Mercadié, L., Aucouturier, J.-J. and Bigand, E., Effect of shared music listening during osteopathic treatment: an EEG study, **Psychophysiology** (under review, 2012)
- Lienhardt, N., Aucouturier, J.-J., Johansson, P. and Hall, L. The sad experimenter effect: unvoluntary emotional cues in experimenter's voice influence participant responses, **Emotion** (under review, 2012)
- Aucouturier, J.-J., Fujita, M. and Sumikura, H. Psychological effect of participatory music consumption: Case-study of Nine Inch Nails's Lights In The Sky, **J. Consumer Behaviour**. (accepted, 2012)
- Aucouturier, J.-J. and Bigand, E. Mel Cepstrum and Ann Ova: the difficult dialogue between Music Information Retrieval and **Cognitive Psychology**, J. Intelligent Inform. Systems (accepted, 2012)
- Aucouturier, J.-J., Nonaka, Y., Katahira, K. and Okanoya, K. Segmentation of expiratory and inspiratory sounds in baby cry audio recordings using hidden Markov models, **Journal of Acoustical Society of America**, 2011, Volume 130, Issue 5, pp. 2969-2977

Patents

- US2008040362 Hybrid Audio-visual Categorization System and Method (2008), with F. Pachet
- JP2006106754 Mapped Meta-data Sound-Reproduction Device (2006), with F. Pachet
- WO2006037786: A Content Management Interface (2006) with F. Pachet, P. Roy

Awards and invitations:

- **Best presentation award**, Int. Conf. on Music Information Retrieval (ISMIR), Porto, Portugal, 2012
- **Best paper award**, European Conference on Artificial Life (ECAL), Lisbon, Portugal, Sept. 2007.
- **First prize**, Music Similarity Contest, MIR Evaluation eXchange (MIREX, 2006) (with E. Pampalk)
- **Conference keynote,** French Artificial Intelligence Society Meeting (RFIA), Amiens, Fr., Jan 2008
- **Invited (funded) plenary presentations**: French Acoustical Society Meeting (Dec. 2012); Int. Conf. on Music Information Retrieval (Oct. 2012); 10th Anniversary, Center for Digital Music, Queen Mary Univ. of London (Sept. 2011); Japan Music Week (Nov. 2009); Kansai Music Industry Conference (Sept. 2009); European Acoustics Association meeting (Jan. 2009); 10th Int. Conf. on the Simulation of Adaptive Behavior (Jul. 2008); Workshop on Intelligent Sound Processing (Jun. 2006); IEEE Workshop on Model-Based Processing and Coding of Audio (Nov. 2002).

Media coverage and public dissemination:

- Traditional and audiovisual media**: CNN** (*Social media in Japan*, 2011); Japan Inc (*Real music for a virtual world*, 2009), Associated Press (*Japanese Robot Dances to IPod Music*, 2007); **Wired** (*Sony researcher has duet with his computer*, 2003).
- More than 30 mentions on internet magazines, incl. Wooshi.com (*A Really Cool Crowdsourced Music Video Project*, 2011); The NextWeb (*11 apps + 4 minutes = 1 incredibly cool real-time music video*, 2010); CNET-Japan (*ZMP and University of Tokyo collaborate to make a robot move impulsively*, 2007)
- Art/Science Exhibitions: *Kansei Design*, Axis Gallery, Tokyo, Avril 2010; *The Future of Music*, Design Festa Gallery, Tokyo, Avril 2009; *Sound/MindScapes*, Art Center Ongoing, Tokyo, Nov. 2008; *Miuro robot showcase*, Apple Store, Tokyo 2007