Foro Audrey 17.02.20

# Spectrographic interpretations:
# Can we "read" emotions from spectrograms of human vocalisations?

*Literature review*

## Table of contents

## *Introduction*

A large realm of research is oriented toward the study of speech in humans, however, the non-verbal aspects remains significantly less studied. Nevertheless, that aspect of daily vocal production represents a rich communicative signal which inform about the emotional and motivational state of the producer (Kreiman & Sidtis, 2011). Investigations in human vocal production and perception suggest that non-verbal vocalisations have been shaped by evolutionary processes and serve a communicative purpose benefiting both the transmitter and the receiver. Non-verbal vocalisations are distributed into different call types (e.g. laughter, screams, roars, cries…) that are shared by every culture and are well recognized by other humans when presented in playback experiments (Pisanski & Bryant, 2016). They mainly occur in interpersonal context and are associated with different social functions which involve most of the time emotional expression and the communication of motivational states (Anikin et al., 2018). As first suggested by Darwin (1872), human communication of emotions is deeply-rooted in a phylogenetic history. Current comparative research provides evidences of cross-specific connections in the expression of affective states, especially, clear acoustic similarities have been found among mammals for different affective categories (Briefer, 2012; Bryant & Aktipis, 2014; Pisanski & Bryant, 2016), for example, an agonistic or aggressive vocalisation produced during threat displays are typically characterized by a lower pitch, higher amplitude, and a higher proportion of nonlinear phenomena than vocalisations produced in an affiliative context. Thus, the study of human nonverbal vocalisations can provide new insights about the evolution of vocal behaviour and their communicative functions. Acoustic analysis and psychophysics experiments appear to be relevant methodologies to explore this topic.

In the first place, the research approach of affective communication related to the bioacoustic framework will be presented, a particular attention will be dedicated to nonlinear phenomenon in vocal production. Then, the focus will be oriented toward humans' vocalisations related to painful context with the example of babies' cries. Thirdly, the emphasis will be on visual representation of sounds and auditory-visual correspondences. Finally, the research question will be introduced while discussing the potential perceptive effects of a visual display of sounds and how it could be relevant in the study of pain expression.

## *The production and perception of affective vocalisation*

### *A bioacoustics framework*

The literature suggests that decoding processes of vocal information can be separated in two steps: the recognition of the type of vocalisation emitted and the interpretation of the emotion expressed (Anikin et al., 2018). Both vocal repertoire categorisation and emotional expression can be recognized by acoustic analyses and psychoacoustic discrimination in humans.

According to Russel's dimensional perspective (1980), affective states can be described relatively to their valence (negative or positive) and their arousal (degree of intensity). The investigation of affective vocalisations reveal that those two dimensions are actually coded in the vocal signal (Briefer, 2012).

Lima et al. (2013) investigated a corpus of non-verbal vocalisations related to various emotional expression. They asked listeners to categorize a set of vocalisations according to several emotions and rate the valence and arousal of those sounds. Besides, they conduct acoustic measures and analyses of those vocalisation. With statistical methods, the authors have been able to discriminate between emotion categories from acoustic cues and also predict listener's rate of those emotions. The acoustics cues they rely on were related to temporal aspects, intensity, fundamental frequency (f0) and voice quality of the records. These cues were sufficient in term of information to obtain an accurate categorization of emotions.

Recently, Taylor & Reby (2010) updated the "source-filter theory", a framework initially dedicated to study human speech, for the investigation of nonverbal vocal communication in mammals. This theory allows researchers to connect the vocal production to his acoustic structure. The vocal production is separated in two elements coming from distinct anatomical parts of the vocal apparatus: the *source* associated with the glottal wave generated in the larynx and the *filter* generated by the supralaryngeal vocal tract. Thus, the fundamental frequency of a sound, comes from the source (vibration rate of the vocal folds in the larynx) and the formant frequency comes from the filter (the vocal tract resonance). The perception of the voice pitch comes mainly from the mean f0 but can also be slightly influenced by the filter characteristics.

With the exploration of nonverbal vocalisation through the lens of the source-filter theory it becomes possible to predict certain acoustic parameter variations related to inner affective change in term of arousal and valence (Briefer, 2012). Changes in the affective states impact directly voice parameters. In mammals, the autonomic nervous system evolved along communicative behaviours, the vagus nerve is neuroanatomically linked to the cranial nerves where social and emotional regulation occur through vocal production (Porges, 2001). Indeed, the arousal and valence of an affective state tend to affect vocal production through short-term changes to the vocal apparatus. For instance, arousal tend to increase the muscles tension of the vocal fold which may lead to a higher pitched voice (Zeskind and Lester, 1978).

Concerning emotional valence, studies have shown that the formant frequency (especially F3 and F4) and the energy distribution in the spectrum which are cues related to the filter could help distinguishing between positive and negative emotions. (Briefer, 2012). Moreover, Scheiner et al. (2002) reported that in the case of infant vocalisations, the shift from positive to negative emotional state get along with an increase of the duration, the frequency range and the peak frequency of the power spectrum of the vocal signal.

### *Nonlinearities in the voice: which communicative function?*

The analysis of the different species vocal repertoire reveals the frequent occurrence of irregularities in the voice that are not related to pathological causes; these acoustic features are called nonlinearities. The associated calls are perceived as particularly harsh and seems to be associated with specific affective contexts (especially highly aroused negative states). The literature emphasizes that nonlinearities in vocalisations are correlated with the perception of vocal roughness. Quite common in the scream type of calls, roughness is observed in an acoustic space between 30 and 150 Hz and is characterized by strong temporal modulations (Arnal et al., 2015). It comes along with distressed states bringing a lot of physical tension, including in the vocal apparatus.

According to Fitch et al. (2002), roughness of voice result from an excessive airflow pushing through the vocal tract that creates irregularities in the vocal fold vibration. With an increasing subglottal pressure, the vocal capacities reach a certain threshold where nonlinear phenomenon can occur. These features, apparent in the spectral signal, result from "*abrupt transition between qualitatively different acoustic regimes". The authors point out two main acoustic nonlinearities as a characteristic of roughness: subharmonics and deterministic chaos. According to Fitch et al., theses nonlinearities are likely to appear successively in the vocal signal with increasing values of subglottal pressure and vocal fold tension parameters (subharmonics appear before deterministic chaos). *Figure 1* illustrate several calls recorded from rhesus macaque males in the study of Fitch et al. (2002): subharmonic formation on the 2nd and 3rd call can be identified; the 4th call display deterministic chaos.

- *Subharmonics*: they can be identified as an abrupt "frequency halving" in the spectral signal (parallel lines appear between the f0 and the first harmonic as in between harmonics). Mechanically, the subharmonics is formed when the tension between the vocal fold is not equally distributed: one could go through two periods while the other one is completing only one period.
- *Deterministic chaos*: it can be identified as an abrupt transition to a visual spectral noise and instability in the signal. Acoustically, the energy is spread at many different frequencies which create a sense of noisiness. However, the apparent random activity is actually deterministic and the signal keeps an actual spectral structure with traces of harmonics (periodic energy), which makes it distinctive from white noise. Mechanically, deterministic chaos is formed by irregular vibrations the vocal folds when they are desynchronised.

Besides, other non-linearities can be produced by the vocal apparatus. Less frequent, each vocal fold can vibrate independently at different frequencies and create a signal including *biphonations*. Another one is *frequency jumps (or pitch jumps)* characterized by an abrupt change of the vocal fold vibrations which suddenly produce a higher or lower f0 (Riede et al, 2007). A schematic spectral representation of the four non-linearities cited above is provided in Figure 2 (from Riede et al., 2007).

Finally, we will mention low frequency modulations (not included in nonlinearities) as an additional acoustic marker of roughness. They also signal a strong tension on the vocal fold and are notably observed in babies pain cries and (Koutseff et al., 2018; Raine et al., 2018).

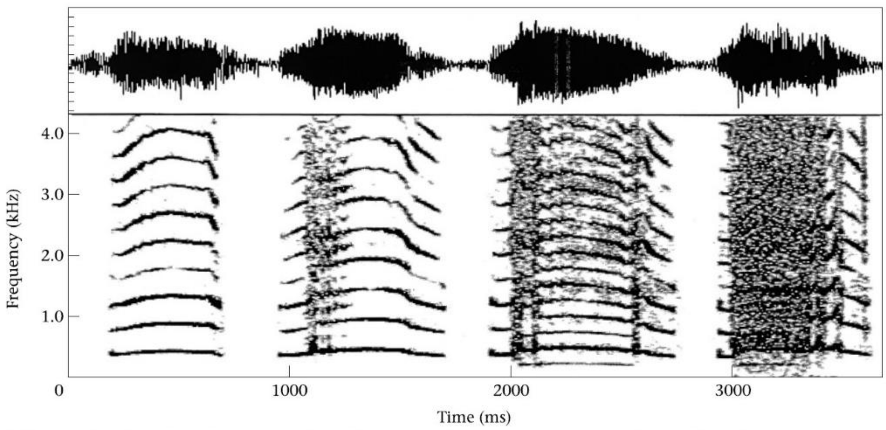


*Figure 1: Audiogram and corresponding spectrogram of four consecutive calls produced by a rhesus macaque male from Fitch et al., 2002. Intrusion of nonlinear phenomena are identifiable from call 2 to 4 (subharmonics on call 2 and 3, and deterministic chaos on call 4).*

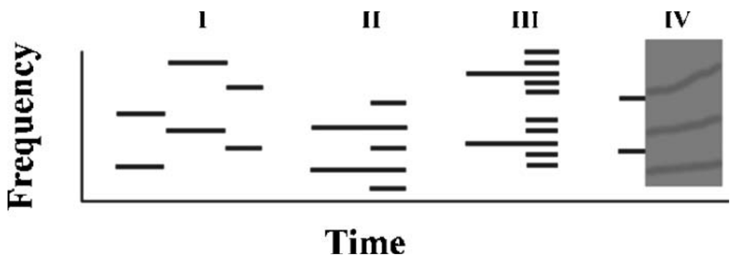


*Figure 2: Schematic narrow-band spectrograms illustrating frequency jumps (I), subharmonics (II), biphonation (III), and deterministic chaos (IV) by Riede et al., 2007.*

Nonlinearities have been observed in many species, however, the communicative function and adaptative value of such phenomenon is still in debate (Anikin, 2019; Fitch, 2002; Karp et al., 2014; Riede et al., 2007). Rough vocalisations are likely designed to trigger an affective response in the receiver since they are qualitatively alarming (and present acoustical characteristics used in the configuration of artificial alarms; Arnal et al., 2015).

One hypothesis is that the acoustic unpredictability of a rough vocalisation makes them difficult to ignore and difficult to habituate to (Reby & Charlton, 2012). Producing rough vocalisations could be a good strategy to grab the attention of a broad range of individuals. Karp and colleagues (2013) tested this hypothesis on meerkat with a playback experiment. They found that meerkats take longer to habituate to natural alarm calls containing subharmonics than calls without it.

Regarding human perception of nonlinearities, an fMRI study carried by Arnal et al. (2015) on human subjects revealed that the amygdala is especially sensitive to the temporal modulation of rough sounds. Moreover, Anikin (2019) found that chaos is especially associated with aversive experiences. He investigates the effect of different levels of nonlinearities in the vocal signal with the help of parametrical voice synthesis. By adding non-linearities in a vocal signal he showed that frequency jumps, subharmonics and chaos tended to be associated with pain. For instance, hearing a moan or a gasp initially related to pleasure tend to sound rather hurtful when nonlinearities are added into the signal.

## *The expression of pain in humans' vocalisations*

### *Pain assessment*

The investigation of pain communication through vocal signal reveal that actors are able to produce sounds matching the acoustic properties of an honest vocal signal of pain (comparable with authentic pain vocalisations of adults such as childbirth vocalisations). Indeed Raine et al. (2018) found that simulated levels (mild, moderate and severe) of pain were well assessed by listeners. With an increase of simulated pain intensity, the following parameters increases as well: the mean and range of voice f0, the amplitude of the vocalisation, the degree of periodicity of the vocalisation and the proportion of the signal displaying nonlinear phenomena.

Pain is defined by the International Association for the Study of Pain as "*an unpleasant sensory and emotional experience associated with actual or potential tissue damage, or described in terms of such damage*".

Specific clinical skills are required when it comes to assess pain on individuals who can't report their subjective experience. Indeed, a significant portion of patients can present communicative disorders or are inapt to speak (psychic disorders, aging, severe illness, preverbal age…). In that case, the measure of pain is dependent on an observer who can rely on different cues to make that judgment. Infant assessment of pain is a good example: clinicians usually rely on behavioural (e.g. facial expression, body gesture, vocalisation) and/or physiological measures (e.g. heart rate variability, blood norepinephrine concentration) to assess pain in neonates (Cong et al., 2013).

We can highlight two challenges related to the development of a good pain assessment tool: the method has to be able to detect when the individual actually experience pain (valence dimension), and capture a panel of different pain levels (arousal dimension). Applying the knowledge of bioacoustics to spontaneous pain vocalisations can bring insights about their communicative value and potentially improve pain assessment tool, knowing that vocal recording present clear practical advantages in its implementation.

### *Baby's cries: when are they in pain?*

Distress cries of neonates are caused by aversive internal or external stimulation (LaGasse et al., 2005). According to Lester et al. (1984) cited by LaGasse et al. the vocalization "*mediates development through parental intervention*". The ultimate cause of cries can be interpreted through a biosocial frame in which the helpless neonate assures his survival through communicative cries, so that his basic needs are met. The vocal signal is shaped in a way that elicit attention and trigger a fast response from the adult caregiver. Especially, the vocalisation has to be well detectable and perceived as aversive. Thus, when they are produced, cries elicit a high motivation in the caregiver to stop them (Pisanski & Bryant, 2018).

There is no clear consensus about the development of emotional expression in neonates within the first months and whether they can express specific emotions. However, aversive emotional states and distress appear to be an emotional reaction that can be well identified from birth (Scheiner & Fisher, 2011). In a negative emotional context, the experience of pain is for instance expected to be correlated with the production of aversive vocalizations.

In general, high pitched cries of babies' are perceived as more painful by adults (Porter et al., 1986). However, f0 parameters haven't been found to reliably vary with different levels of pain and doesn't seem to be a good predictor of babies pain intensity despite the effect of high pitch on adults perception of babies cries (Koutseff et al., 2018). Particularly, the inter-variability of babies' pitch is quite high (Reby et al., 2016). Besides, Bellieni et al. (2004) reported an abrupt increase of the mean f0 when a threshold of pain is reached, that is, a DAN scores above 8 (Douleur Aiguë Nouveau-né, a scoring system of pain assessment for neonates), which result of a siren like cry.

A recent study of Koutseff et al. (2018) showed that from a mild discomfort (bath context) to the experience of pain (vaccine), parents can estimate the intensity of the pain expressed by their infant only by listening to their vocalisations. Surprisingly, the acoustic analysis allowed a discrimination between the two types of vaccines eliciting different levels of pain but parents failed to make this discrimination. Therefore, the acoustic information allowed a more precise discrimination of pain levels than did parents. The spectrographic analysis reveals different profiles of signals according to the painful context suggesting that variations of pain intensity are visually detectable on a spectrogram: bath cries tend to display a well-define harmonic structure while vaccine cries tend to display visual irregularities caused by nonlinear phenomenon (see Figure 3 from Koutseff et al., 2018). Indeed, from the sound analyses of this study, one major acoustic structure correlated with pain expression in babies was extracted: *the cry roughness*. With an increased pain, cries tend to have a more non-linear phenomena such as biphonation, subharmonics, deterministic chaos and vibrato-like frequency modulation. Hence, roughness properties in infant cries appear to be a key acoustic feature to assess pain and could potentially generalise to other types of pain vocalisations such as screams, moans, grunt…

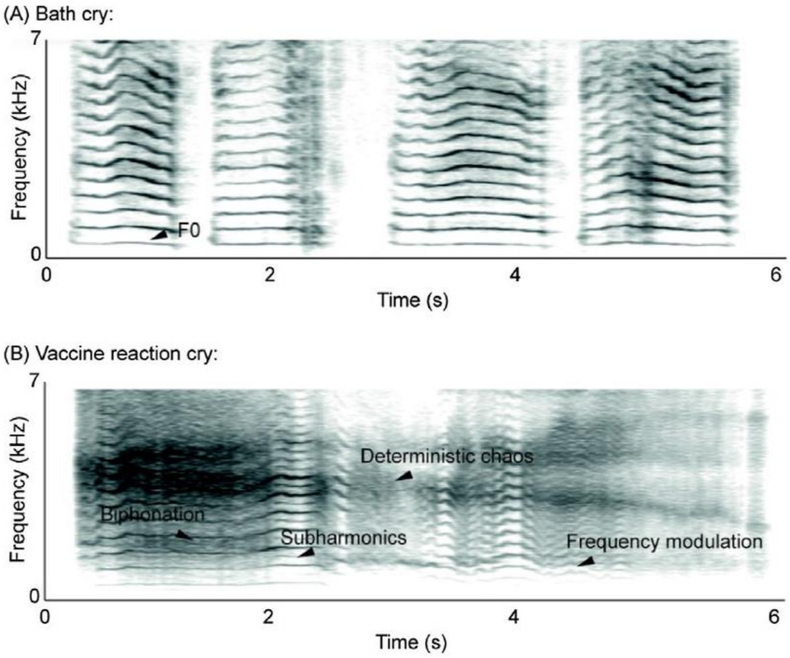


Figure 3: Spectrograms of a cry recorded during bath (A) and immediately after a first vaccine injection (B) from Koutseff et al., 2018.

## *Crossmodality and Sensory substitution*

### *Crossmodality: Can we switch from auditory to visual perception?*

Constantly, our perception of the world is dependent on multiple sensory inputs. We access the various stimulation through the common 5 sensory modalities (vision, audition, somatosensation, olfaction and gustation). Since early childhood, the developing brain can easily have access to an object through different perceptive modalities.

We will focus on one related psychological effect of multimodal perception documented in humans: the cross-modal correspondence phenomenon. This effect refers to a "*compatibility effect between attributes or dimension of a stimulus (i.e. an object or event) in different sensory modalities*" (Spence, 2011). In general population, a strong cross-modal association effect between sounds and visual representations can be observed. For example, sounds of high pitch are intuitively associated with upper space, small objects and bright colour whereas the reverse is observed for lower pitched sounds (Spence, 2011). As well, the semantical field of height is systematically used to describe the pitch, also, musical notation represents higher pitched note higher in the musical stave. Moreover, the assessment of voice quality used to involve cross-modal processes. As Kreiman & Sidtis (2011) reported, one of the approaches to specify voice quality is to create a list of terms helping listeners to describe their impressions of the voice. They may describe a voice from another sensorial modality (for example visually by rating the extent to which a voice is brilliant or dark).

Other evidences of this phenomenon are provided in the realm of Gestalt psychology related to sound symbolism. From older studies, the associations between word's sounds and shapes have been replicated: Köhler (1929) initially found that people strongly agree to associate non-sense words with specific shapes. Represented in *Figure 4*, "*Baluma*" is associated with the curved shape (A) and the word "*Takete*" with the sharp angular shape (B). Later, Ramachandran and Hubbard (2001,2003) have replicated this experiment with another combination of stimuli and demonstrate the "*bouba/kiki effect*" (see *Figure 4*). They indeed obtain similar results where 95 to 98% of the population agreed on the association of the rounded shape (1) with the word "*Bouba*" and the angular one with the word "*Kiki*". Moreover, Ozturk, Krehm, and Vouloumanos (2013) found the same effect when words were presented aurally.
Crossmodal correspondence in the realm of sound symbolism is likely to involve high order processing. Ramachandran and Hubbard (2003) reported that lesion in the angular gyrus (temporal-parietal-occipital region) is related to a loss of the bouba/kiki effect. Sounds-shape associations might depend on the neural organization of the perceptual system. The authors suggest that the association between two sensory features could result from a close located neural processing (Ramachandran & Hubbard, 2001).

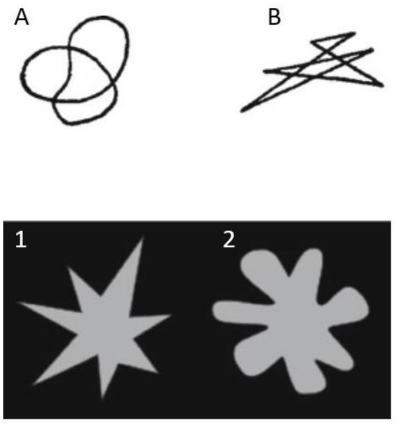


Figure 4: Schematic figure illustrating the kinds of stimuli used to demonstrate crossmodal association between word's sounds and shapes from Spence (2011). Shapes A and B refer to the work of Köhler (1929); Shapes 1 and 2 refer to the work of Ramachandran and Hubbard (2001,2003).

Another important point related to cross modal perception concern persons with sensory impairment. Brain plasticity cause rearrangements in the processing of sensory input: for instance, neural substrate of the auditory cortex of the deaf could be used for visual processing and even enhance their visual abilities (Lomber et al., 2010). This can call into question the perceptive abilities among sensory deprived person: does reported crossmodal effects in the general population apply the same way to sensory deprived persons? Here the role of sensory experience has been investigated to understand the mechanisms involved in crossmodal correspondences. Sound-shape association has been suggested to depend on visual experience.

For instance, the Bouba/kiki effect is significantly weaker in blind and partially-sighted participants (Fryer, Freeman, and Pring, 2014). Hamilton-Fletcher et al. (2018) investigated the role of visual experience in the emergence of auditory-tactile correspondences. In that purpose, they compared the strength of several correspondences in blind and sighted subjects. Their findings suggest that sensory input affect cross-modal correspondence in a complex way: depending on the correspondence studied, the association can be either dependant or independent of visual experience, or emerge with blindness. Different process might be involved in the emergence of cross-modal correspondence and the question of learned versus innate associations isn't clear (Spence, 2011), which encourage research pursuit on this topic.

Nevertheless, this field of research allowed the development of specific cognitive tools taking advantage of the cross-modal phenomenon to offer new possibilities of perceptive experiences.

### *Sensory substitution devices to compensate sensory loss*

Sensory substitution devices have been developed for sensory deficient persons, they "*aim at replacing or assisting one or several functions of a deficient sensory modality by means of another sensory modality*" (Auvray et al., 2009). For instance, in case of an audition-to-vision devices, the substituting modality (vision) will allow the access to certain information of the environment that is normally accessed through audition. Their efficacity depend on brain plasticity properties and so their use requires an initial training (Bach-y-rita & Kercel, 2003).

Bach-y-rita & Kercel (2003) suggest that reading is the first sensory substitution system as it involves visual presentations of auditory information. Indeed, the perception of spoken words from signs perceptions remain a learned skill. Nowadays, new promising sensory substitution tools relate on human-machine interface that transduce the environmental information to an individual functioning sensory modality. The literature provides evidences of the successful exploitation of cross-modal phenomenon especially among visual, auditive and tactile modalities. Much of this technology had been applied to compensate visual deficiency (Auvray et al., 2009; Bach-y-rita & Kercel, 2003).

Many devices using tactile information as a sensory substitute prove their efficacity for different surface areas of the body. One is using the tongue interface by delivering electro-tactile stimulation converted from a head-mounted camera (Bach-y-Rita et al., 1998). The blind subject can then experience a stream of pulses giving him information about the visual scene and use this information to achieve localisation tasks. Promising visual-to-auditory substitution devices also emerged. Those systems can use the principle of echolocation to inform the blind user about the distance and direction of distant objects. For example, these dimensions can be respectively coded by pitch and inter-aural disparity (Kay, 1964, 1985). Moreover, with the idea to convert image into sound pattern, Meijer (1992) proposed a prototype able to construct an audio mapping of the scene from a video. His system associate height with pitch and brightness with loudness and also provide a left to right time scanning to encode horizontal position. The development of such devices goes along with technologic challenges related in part to portability and miniaturization

To compensate auditory deficiency, the amount of proposal is less consequent and research seems to be mainly oriented toward restorative strategies and neuroprosthetic devices such as cochlear implants. However, auditory-to-tactile devices have been developed and shown to be helpful to enhance auditory perception and especially speech perception. Cieśla et al. (2019), found that a finger tips tactile device could help to improve speech in noise perception. They developed a minimal sensory substitution tool relaying on the fundamental frequency of the speech signal. The system delivers a vibratory stimulation on two finger tips corresponding to a transduced low-frequency speech signal. Hence, in degraded auditory environment, this device significantly improves the understanding of speech without prior training (benefit of 6dB). In addition, Novich & Eagleman (2014, 2015) proposed an alternative aid for deaf populations who can't benefit from cochlear implant. A vibrotactile vest which transduce auditory information into vibratory stimulation in the torso. The authors investigate the possibility of providing spatio-temporal information with the goal to encode speech through tactile stimulation for the deaf.

## ***Research question & Hypothesis***

### *Does visual-to-auditory correspondence apply to spectrograms?*

Currently, it hasn't been tested yet if naive humans can interpret the affective states coded in a vocal signal only from a visual presentation of the vocalisation (supported by a spectrogram). A spectrographic representation informs visually about the relationship between three dimensions of a sound: the frequency (Y axis, in Hz), the time (X axis, in ms) and the energy of the soundwave (contrast in dB).

The investigation of pain vocalisations in babies showed that key features of pain communication such as roughness characteristics are visually noticeable in a spectrogram, however, there is no knowledge about whether sound visual representation can elicit cross-modal effects. Thus, we can address the following questions:

- *Does cross-modal correspondence occur between the auditory stimulus and its spectrographic representation?*
- *Are naive subjects (i.e. without voice analysis or bioacoustics knowledge) able to assess distress and pain intensity from spectrograms of human vocalisations (infant cries and adult both volitional and spontaneous pain vocalisations)?*
- *Which properties of the vocalisation (acoustic or spectrographic) predict listeners' pain rating? Does the level of auditory and visual roughness cause a higher pain rating?*
- *Can subjects assess variations of pain intensity over time from spectrograms?*
- *Can we train people to read pain intensity in spectrograms?*
- *Do deaf population respond similarly to spectrographic presentations than hearing population?*

We predict that naïve listeners that have no prior experience in reading spectrograms are able to deduce the pain intensity of humans' vocalisations using only the spectrogram of the sound. In other words, we assume that subjects won't need to be trained to interpret the affective information carrying by the visual signal. Supported by cross-modal phenomenon between visual and auditory perception, subjects could have an inherent and intuitive capacity to infer such affective cues in the spectrogram.

Indeed, the nonlinearities characterizing vocal roughness are likely to be perceived as chaotic both in the auditory and visual domain. From a Gestalt point of view, a higher-level processing effect could occur when a subject perceives a well-shaped harmonic vocal signal in the spectrogram versus a chaotic signal with abrupt transitions of the acoustic regime. Thus, a spectrographic representation of a vocalisation expressing pain is expected to produce a percept that will be associated with the aversiveness of the vocal signal conveyed by roughness. In other words, a disruption of the harmonicity of the signal is expected to be perceived as such both by the ear and the eye and therefore can be interpreted as an expression of pain.

This study is motivated by the possibility of using spectrogram visual cues as a new sensory input relevant in the assessment of babies' cries, indeed, acoustic analysis is easily accessible and could potentially be implemented in clinical setting or at home.

### *Perspectives: can we help deaf people having genuine acoustic information about the cries of their babies?*

For caregivers, perceiving and reacting to babies' communicative signals is necessary for the well-being of the infant and more broadly a question of survival. Vocal cues produced by babies came from a deeply-rooted adaptative process (Pysanski & Bryant, 2016) and the behavioural reaction of the caregivers is determinant for the baby development (Lester et al., 1995). Parent perception of cries has been shown to play a role in the development of the child (LaGasse et al., 2015). A good recognition of the aversiveness (increase f0) or not (normal f0) of a cry from mothers is related with a better cognitive development of the child. Lester et al. (1995) showed that toddlers of mothers who identified well the aversiveness of their cries had a higher Bayley mental score and language score at 18 months than toddlers of mothers who misperceived their cries. This result suggests that the perception of babies' cries could have a high adaptative value for humans.

Nonetheless, some parents can't perceive the vocal communication of their babies. Specific aids can be provided for hearing-impaired parents to help them understanding the vocal signals of their infants. Anderson and colleagues developed the application Chatteraby$^{TM}$ for that purpose. It uses algorithm and machine learning to process the sound data from the baby related to an already existing data base (cries in the context of vaccines or ear-piercing validated by veteran mothers) to access the expressed pain intensity from a 0 to 10 scale (no pain to worst pain possible). This application also assesses other states of the babies like hunger of fussiness (Alwan et al., 2019).

Another perspective of a well accessible aid for deaf parents could be using a direct sensory substitution tool displaying a visual information (spectrograms) of the vocal signal which could bring more flexibility in term of perceptive interpretation and let the parent form their own understanding of the vocal behaviour of their infant.

## *Bibliography*


Anikin, A., Bååth, R., & Persson, T. (2018). Human non-linguistic vocal repertoire: Call types and their meaning. *Journal of nonverbal behavior*, *42*(1), 53-80.

Alwan, B. S., Han, C., Bookheimer, S. Y., Eyer, S., Dapretto, M., & Zeltzer, L. (2019). Defining and distinguishing infant behavioral states using acoustic cry analysis: is colic painful?.

Anikin, A., & Lima, C. F. (2017). Perceptual and acoustic differences between authentic and acted nonverbal emotional vocalizations. The Quarterly Journal of Experimental Psychology, 1-21.

Arnal LH, Flinker A, Kleinschmidt A, Giraud AL, Poeppel D. 2015. Human screams occupy a privileged niche in the communication soundscape. Curr Biol. 25(15):2051–2056.

Auvray, M., & Myin, E. (2009). Perception with compensatory devices: From sensory substitution to sensorimotor extension. *Cognitive Science*, *33*(6), 1036-1058.

Bellieni, C. V., Sisto, R., Cordelli, D. M., & Buonocore, G. (2004). Cry features reflect pain intensity in term newborns: an alarm threshold. *Pediatric research*, *55*(1), 142.

Briefer, E. F. (2012). Vocal expression of emotions in mammals: mechanisms of production and evidence. Journal of Zoology, 288(1), 1-20.

Cieśla, K., Wolak, T., Lorens, A., Heimler, B., Skarżyński, H., & Amedi, A. (2019). Immediate improvement of speech-in-noise perception through multisensory stimulation via an auditory to tactile sensory substitution. Restorative neurology and neuroscience, 37(2), 155-166.

Cong X, McGrath JM, Cusson RM, Zhang D. 2013. Pain assessment and measurement in neonates: an updated review. Adv Neonatal Care. 13(6):379–395.

Darwin, C. 1872. The Expression of the Emotions in Man and Animals. London: John Murray.

Dessureau, B. K., Kurowski, C. O., & Thompson, N. S. (1998). A reassessment of the role of pitch and duration in adults' responses to infant crying. *Infant Behavior and Development*, *21*(2), 367-371.

Fitch, W. T., Neubauer, J., & Herzel, H. (2002). Calls out of chaos: the adaptive significance of nonlinear phenomena in mammalian vocal production. Animal behaviour, 63(3), 407-418.

Fryer, L., Freeman, J., & Pring, L. (2014). Touching words is not enough: How visual experience influences haptic–auditory associations in the “Bouba–Kiki” effect. Cognition, 132(2), 164-173.

Karp, D., Manser, M. B., Wiley, E. M., & Townsend, S. W. (2014). Nonlinearities in meerkat alarm calls prevent receivers from habituating. *Ethology*, *120*(2), 189-196.

Kay, L. (1964). An ultrasonic sensing probe as a mobility aid for the Blind. Ultrasonics, 2, 53.

Kay, L. (1985). Sensory aids to spatial perception for blind persons: Their design and evaluation. In D. Warren & E. Strelow (Eds.), Electronic spatial sensing for the blind (pp. 125–139). Dordrecht, The Netherlands: Martinus Nijhoff.

Kreiman, J., & Sidtis, D. (2011). Foundations of voice studies: An interdisciplinary approach to voice production and perception. John Wiley & Sons.

Köhler, W. (1929). Gestalt psychology. New York: Liveright.

Koutseff, A., Reby, D., Martin, O., Levrero, F., Patural, H., & Mathevon, N. (2018). The acoustic space of pain: cries as indicators of distress recovering dynamics in pre-verbal infants. *Bioacoustics*, *27*(4), 313-325.

LaGasse LL, Neal AR, Lester BM. 2005. Assessment of infant cry: acoustic cry analysis and parental perception. Ment Retard Dev Disabil Res Rev. 11(1):83–93.

Lima, C. F., Castro, S. L., & Scott, S. K. (2013). When voices get emotional: a corpus of nonverbal vocalizations for research on emotion processing. *Behavior research methods*, *45*(4), 1234-1245.

Lomber, S.G., Meredith, M.A. and Kral, A. (2010) Crossmodal plasticity in specific auditory cortices underlies visual compensations in the deaf. Nature Neuroscience 13: 1421-1427.

Meijer, P. B. L. (1992). An experimental system for auditory image representations. IEEE Transactions on Biomedical Engineering, 39, 112–121.

Novich, S. D., & Eagleman, D. M. (2014, February). [D79] A vibrotactile sensory substitution device for the deaf and profoundly hearing impaired. In 2014 IEEE Haptics Symposium (HAPTICS) (pp. 1-1). IEEE.

Novich, S. D., & Eagleman, D. M. (2015). Using space and time to encode vibrotactile information: toward an estimate of the skin's achievable throughput. Experimental brain research, 233(10), 2777-2788.

Ozturk, O., Krehm, M., & Vouloumanos, A. (2013). Sound symbolism in infancy: Evidence for sound–shape cross-modal correspondences in 4-month-olds. *Journal of experimental child psychology*, *114*(2), 173-186.

Pisanski, K., & Bryant, G. A. (2016). The evolution *of voice perception*. Oxford, UK: Oxford University Press.

Porges, S. W. 2001. "The Polyvagal Theory: Phylogenetic Substrates of a Social Nervous System." International Journal of Psychophysiology 42 (2): 123–46.

Porter FL, Miller RH, Marshall RE. 1986. Neonatal pain cries: effect of circumcision on acoustic features and perceived urgency. Child Dev. 57(3):790–802.

Raine, J., Pisanski, K., Simner, J., & Reby, D. (2019). Vocal communication of simulated pain. *Bioacoustics*, *28*(5), 404-426.

Ramachandran, V. S., & Hubbard, E. M. (2001). Psychophysical investigations into the neural basis of synaesthesia. Proceedings of the Royal Society of London. Series B: Biological Sciences, 268(1470), 979-983.

Ramachandran, V. S., & Hubbard, E. M. (2001). Synaesthesia—A window into perception, thought and language. Journal of Consciousness Studies, 8, 3–34.

Ramachandran, V. S., & Hubbard, E. M. (2003, May). Hearing colors, tasting shapes. Scientific American, 288, 43–49.

Reby, D., & Charlton, B. D. (2012). Attention grabbing in red deer sexual calls. *Animal cognition*, *15*(2), 265-270.

Reby D, Levrero F, Gustafsson E, Mathevon N. 2016. Sex stereotypes influence adults' perception of babies' cries. BMC Psychol. 4(1):19.

Russell, J. A. 1980. "A Circumplex Model of Affect." Journal of Personality and Social Psychology 39: 1161–78.

Scheiner, E., & Fischer, J. (2011). Emotion expression: The evolutionary heritage in the human voice. In Interdisciplinary anthropology (pp. 105-129). Springer, Berlin, Heidelberg.

Scheiner, E., Hammerschmidt, K., Jürgens, U., & Zwirner, P. (2002). Acoustic analyses of developmental changes and emotional expression in the preverbal vocalizations of infants. *Journal of Voice*, *16*(4), 509-529.

Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. Speech communication, 40(1-2), 227-256.

Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics*, *73*(4), 971-995.

Taylor, A. M., & Reby, D. (2010). The contribution of source–filter theory to mammal vocal communication research. Journal of Zoology, 280(3), 221-236.

Zeskind PS, Lester BM. 1978. Acoustic features and auditory perceptions of the cries of newborns with prenatal and perinatal complications. Child Dev. 49:580–589.