

ERC Starting Grant 2023**Part B2¹****(not evaluated in Step 1)**

Sections (a) and (b) of Part B2 should not exceed 14 pages. **References do not count** towards the page limits.

Section a. State-of-the-art and objectives**1. Statement of the problem**

Many perceptual and learning mechanisms serving the acquisition of language are not specific to language nor to humans (e.g., Christiansen & Kirby, 2003; Yip, 2006; Fitch, 2018; Santolin & Saffran, 2018; Petkov & ten Cate, 2020). Then **why are humans the only species so far able to learn language?** Research at the crossroad of Cognitive Neuroscience, Biology and Linguistics has not yet fully identified the features and the mechanisms of language that make it easily learnable by humans. Animals are equipped with extraordinary capacities, spanning from basic auditory and visual scene analysis to complex cognitive abilities, such as theory of mind, remembering the past and planning for the future, reasoning, and few exceptional language-like communication skills (i.e., *Alex* the parrot's capacity to use human words and ask questions; Pepperberg, 2009; for review: Trestman, 2015). Amongst this set of abilities, there is the capacity to identify relevant processing units from structured inputs. For instance, songbird species organize their own songs into shorter segments (Comins & Gentner, 2013; Abe & Watanabe, 2011) that have been argued to remind of human syllables (Mann et al., 2021). Dogs detect word-like units from speech (Boros et al., 2021). Newborn chicks parse streams of visual objects into smaller units co-occurring in the input (Santolin et al., 2016; Santolin et al., 2020). Pigeons and rhesus macaques detect units of printed letters that form words (Grainger et al., 2012; Scarf et al., 2016).

Parsing the input into smaller processing units is one of the critical components of human language acquisition. Before six months of age, human infants are able to identify relevant units from speech, such as phonemes, syllables or words, which are the building blocks of subsequent linguistic processing (Friederici, 2005; Werker, 2018; Saffran & Kirkham, 2018). **Syllables have a primacy over all the other linguistic units.** Neonates and infants automatically and spontaneously process speech as units that roughly correspond to syllables (e.g., Bijeljac-Babic et al., 1993). Universal linguistic principles have been claimed to determine the structure of syllables (i.e., order of phonemes) across human languages, and to account for the advantage of syllables as preferred linguistic unit early in development.

The present project proposes that syllables are biological units of sounds that constitute an “entry-gate” to language, and it is aimed at characterizing the nature of the mechanisms defining the structure of such syllabic units. I will explore two alleged universal mechanisms that allow to parse the speech into syllable-sized units and modulate their internal structure (*Maximal Onset Principle* and *Sonority Hierarchy*) from ontogenetic and phylogenetic perspectives, with series of experiments with human neonates, infants, adults and nonhuman species. The use of a

¹ Instructions for completing Part B2 can be found in the ‘*Information for Applicants to the Starting and Consolidator Grant 2023 Calls*’.

comparative-developmental approach is a benchmark for this research goal: it allows to study species with distinct evolutionary histories, and cognitive and neural architectures, as well as humans from birth to early infancy. Such approach is pivotal to unveil the nature of the mechanisms that make syllables critical linguistic units at the onset of language acquisition. The GALA project will provide new, fundamental knowledge of the mechanisms operating at the emergence of language, and will bring us closer to understand why only humans are capable of learning language.

2. State-of-the-art

Extensive research has shown that a variety of species and taxa possess sophisticated cognitive capacities. Among others, apes and avian species show a sense of numerosity, being able to perform basic arithmetical operations (Call, 2000; Vallortigara, 2012) which has also been documented in honeybees and teleost fish (Agrillo & Bisazza, 2018; Giurfa, 2019). Corvids possess remarkable tool-use abilities, that allow to modify the shape of a tool to retrieve food (Weir et al., 2002; Emery & Clayton, 2009). Some parrot species process rhythmic information, synchronizing their body movements with a beat (Patel et al., 2009). These are just few examples of perceptual and cognitive abilities that enable different species to encode information from different sensory inputs, to best adapt to their ecological niches.

Language is perhaps the most prominent feature that distinguishes humans from the rest of the animal kingdom. Nevertheless, many components of language are processed and learned by nonhuman species. Dogs, rats and cotton-top tamarins notice that human languages sound different, discriminating Dutch and Japanese, and Spanish and English (Ramus et al., 2000; Toro et al., 2003; Cuaya et al., 2022). Different species process prosody, a salient aspect of speech that comprises intonation and rhythm. For instance, budgerigars, a parrot species, and rats, are sensitive to stress patterns (the emphasis of a given syllable in a word, e.g., *mamá*), discriminating words having stress in the final vs. initial syllable (Hoeschele & Fitch, 2016; Toro & Hoeschele, 2017; see also Spierings and ten Cate, 2014). Rodents, avian and nonhuman primate species are capable of performing statistical computations on the speech, such as detecting word boundaries or learning rudimentary syntactical structures mimicking human utterances (for review: Santolin & Saffran, 2018; Milne et al., 2018). Some nonhuman primates' communication signals have semantic- and combinatorial-like components, which can be found in distinct alarm calls produced to signal different types of predators (Marler et al., 1992; Zuberbuhler, 2000, 2001). Marmosets produce vocalizations that overlap with the frequency rate of human syllables (4-10 Hz), and synchronize oro-facial movements to the amplitude of the sounds being produced, which was thought to be a human prerogative among primates (Risueno-Segovia & Hage, 2020). All such evidence, along with others, highlights that a variety of animal species are equipped with cognitive and neural machineries to process many linguistic structures, spanning from perceptual (e.g., rhythmic differences across languages) to more complex features (e.g., syntax). Some species like songbirds and nonhuman primates also possess language-like components in their own vocalization systems. Yet, no animal species achieve the language skills of a young human infant.

The infancy literature is rife with evidence indicating that infants from 0 to 4-5 months of age parse the speech into syllable-sized units. Neonates spontaneously detect syllables from speech, automatically synchronizing brain activity to the syllable frequency rate (Fló et al., 2022) in a similar way as observed in adults (Poeppel, 2003). Neonates are sensitive to violations of the universal syllabic structure, being able to discriminate between syllables like *blif* and *lbif* (the latter violates the Sonority Hierarchy described below; Gomez et al., 2014). Infants preferentially process the speech signal as syllabic units rather than tracking individual phonemes (Jusczyk & Derrah, 1987; Bertoncini et al., 1988; Bijeljac-Babic et al., 1993) although some phonetic knowledge is available in the first months of life (Mersad et al., 2020; Gennari et al., 2021).

In linguistics, syllables are defined as abstract constituents with a hierarchical structure, whose main function is to group segments of speech into shorter strings ('phonological syllable', which will be

used in this proposal; e.g., Hockett, 1955; Halle & Vergnaud, 1980). The universal structure of syllable includes a nucleus (the vowel), which is the obligatory element, and onset and coda, optional elements located at the edges of the syllable, typically including consonants (e.g., Prince & Smolensky, 1993/2004; though see Ridouane, 2008 for alternatives). Such structure is mainly determined by the **Maximal Onset Principle (MOP)** which marks syllable boundaries based on where consonants are. MOP states that consonants are preferred at the onset of syllables rather than at codas. Thus, when consonants appear in the middle of a word, they are parsed with the following syllable (Pulgram, 1970; Goldsmith, 2011). For instance, as illustrated in Figure 1, the English word *prosper*, is parsed as *pro_sper*. To the best of my

MOP:

Itwasmeanttopro_sper

**Itwasmeanttopro_sper*

SH:

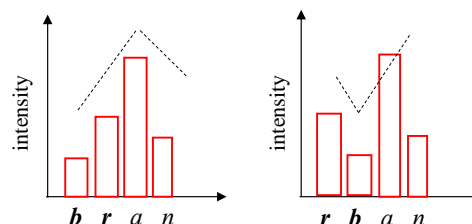


Figure 1. Schematic representation of the linguistic principles

knowledge, no research has investigated the use of MOP to identify syllable-sized units from speech in early infancy. One of the goals of the present research is to provide such evidence. Crucially, the segmented output supplied by MOP is subsequently modulated by the **Sonority Hierarchy (SH)**, an alleged universal phonological constraint that imposes restrictions on the order of consonants at syllable onset. According to SH, syllables' internal ordering of phonemes is based on their **intensity** (perceived loudness): it must increase towards the nucleus (vowel) and decline right after (Clements, 1990; Selkirk, 1984; Parker, 2008). For example, in the word *bran*, the intensity of *b* and *r* rises until *a* is reached, and subsequently fall (*n* has a lower intensity than *a*). In contrast, in *rban*, an unexpected intensity drop occurs before the vowel (since *b* has a lower intensity than *r*), violating the sonority hierarchy. These changes in intensity provide reliable cues to detect syllable-sized units from speech, as computational models trained on speech corpora have shown (Räsänen et al., 2018).

From a purely acoustic perspective, the envelope of syllables (determined by MOP and SH) mirror an arch-shaped 'melody' that is widespread amongst animal vocalizations. Calls emitted through exhalation (e.g., birdsong, cat meows, nonhuman primates alarm or mating calls, wolf howls) depict an envelope that is roughly similar to that of human syllables, with an initial rise of intensity followed by a decline. Such envelope is likely determined by biomechanical constraints on sound production that generates a rapid increase of subglottal pressure in the vocal tract, causing the initial energy burst (Tierney et al., 2011; Mol et al., 2017; Mann et al., 2021). These findings raise the tantalizing possibility that the advantage of syllables as natural linguistic unit and entry-gate to language, is tightly linked to the *nature* of its envelope.

I hypothesize that human syllables might be another biological instance of arch-shaped (acoustic) envelopes widespread amongst animal vocalizations. The main goal of the present proposal is **to investigate whether linguistic mechanisms at the emergence of language acquisition, allowing humans to segment the speech into syllable-sized units, derive from evolutionary-ancient sensory mechanisms that have presumably evolved to process different inputs in different species. I will propose a model that characterizes the nature of such mechanisms, through which humans begin to learn language.** I will explore Maximal Onset Principle and Sonority Hierarchy through a comparative-developmental approach aimed at studying species that have not developed language yet but have the biological predisposition for it (human neonates and infants) or will not develop language at all (Long-Evans rats).

To this end, I will implement two research lines:

Research Line (RL) 1: Investigating the use of Maximal Onset Principle as general mechanism to parse the speech signal. I will use segments of speech as inputs for rats and neonates to explore ontogenetic and phylogenetic origins of mechanisms that allow to detect syllables from speech. I will

also test 6-to-8 month-old infants to explore the role of experience with language for the emergence of language-specific constraints.

Research Line (RL) 2: Testing the sensory bases of Sonority Hierarchy in non-linguistic inputs. I will use segmented “syllables” implemented in different sensory modalities (auditory non-linguistic and vibro-tactile) as inputs for rats and neonates, as well as infants and adults, to investigate whether sonority constraints are based on general sensory processing.

Rationale. The comparative-developmental approach, using same materials across species and humans at different times in development, is optimal to implement these research lines. **Neonates** are ideal to explore biological dispositions towards language, since they have not yet accumulated substantial experience with language, but are endowed with cognitive and neural mechanisms to learn it. **Rats** do not experience nor develop language at all. Therefore, this is a straightforward comparison to explore purely sensory processing of speech. Contrary to humans, rats are non-vocal-learners, which produce only two types of ultrasonic vocalizations (Petkov & Jarvis, 2012; Brudzynski, 2013) however, plenty of research have shown that rats are able to perform many computations on the speech signal (for review: Toro, 2016). Rats thus represent an excellent animal model for the proposed research. This species would also allow to explore whether the computations investigated can be performed by means of an acoustic code only. Rats emit ultrasounds with flat prosody (no rising/falling of intensity), for which no significant movements of the articulators (e.g., mouth) are required. Therefore, it is plausible that rats would use acoustic information (rather than articulatory) to parse the speech, and track syllabic units. If this is the case, it would not imply that human infants do not use articulatory information to compute the same speech input (however, this goes beyond the scope of the proposal). **Infants** (4-5, 6-8 months of age) will be tested to explore the role of experience with language in modulating the universal structure of syllable. Infants also allow to assess preferences (e.g., for “correct vs. incorrect” syllabic structures), which are extremely difficult to measure in newborns. **Human adults** will be tested as control for rats, since very little research has been done with vibro-tactile inputs in early infancy in non-clinical research. Results of the adult experiment will be pivotal to provide solid ground for testing vibro-tactile perception of linguistic structures in neonates and infants. The rationale of the **experimental procedures** described in this proposal is to be able to draw comparisons across species, and across different stages of human development. Behavioral paradigms based on familiarization-discrimination are suitable for all infant ages and rats (with proper procedural adjustments to fit the capacities of each species). Newborns can provide very limited behavioral measures of discrimination, thus functional neuroimaging is highly recommended (Gervain et al., 2011; Benavides-Varela et al., 2011). Importantly, all the procedures allow the use of the same materials across species and ages, which is critical in comparative cognition research.

Section b. Methodology

In this section, I will describe the experiments I have designed (Section 3 for *Research Line 1*, Section 4 for *Research Line 2*), and the procedures I plan to use to implement the research (see Figure 2).

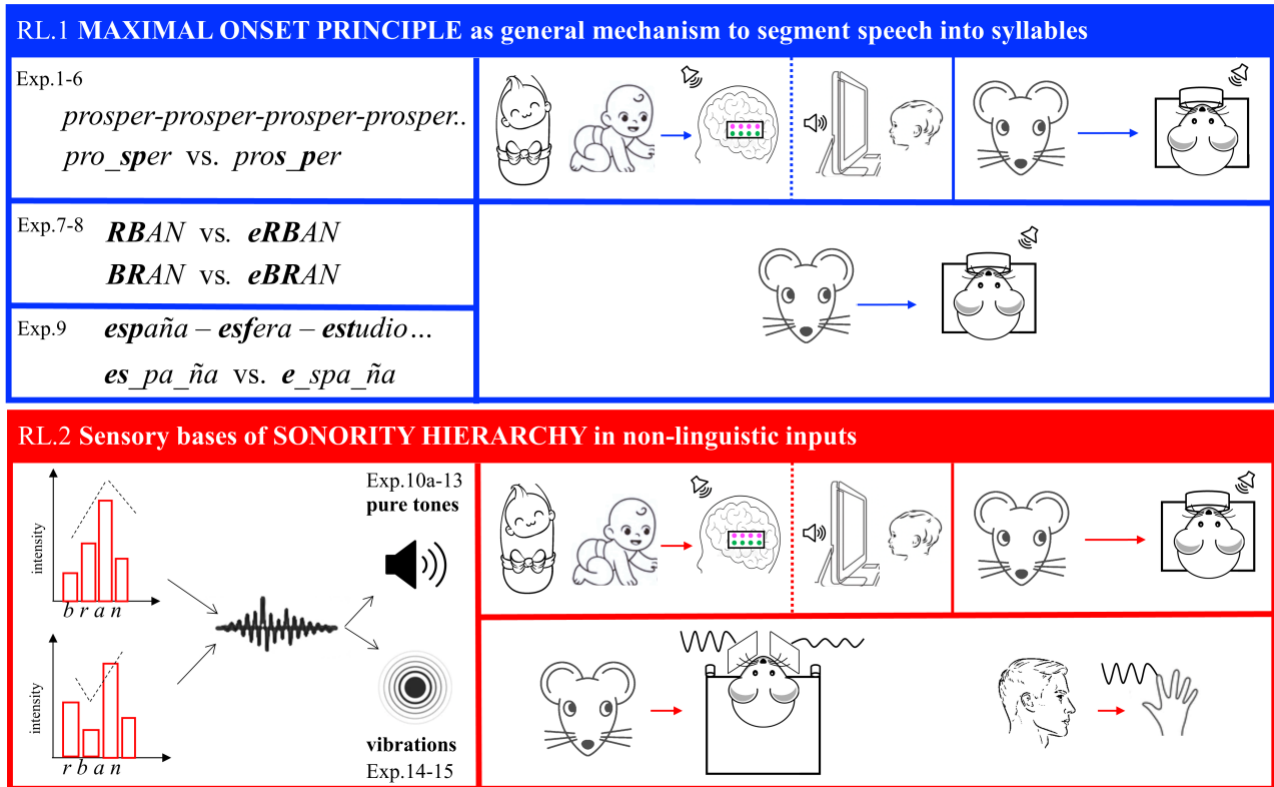


Figure 2. Schematic representation of research lines. **Left RL.1:** examples of speech stimuli to assess segmentation (Exp.1-6, above: familiarization, below: test), and language-specific phonotactics (Exp. 7-9). **Left RL.2:** schematic representation of sensory modalities implementing SH (Exp.10a-13; 14-15). **Right:** Populations and procedures.

3. Research Line 1: INVESTIGATING THE USE OF MAXIMAL ONSET PRINCIPLE (MOP) AS GENERAL MECHANISM TO PARSE THE SPEECH SIGNAL.

A series of experiments will be implemented to investigate whether human neonates and infants are sensitive to Maximal Onset Principle to identify syllable-sized units from speech (Exp. 1-3), and whether such constraints are tailored to learn linguistic structures or rooted into general acoustic processing (Exp. 4-9 with rats). Following previous research (e.g., Benavides-Varela et al., 2011; Celma-Miralles & Toro 2020; Crespo-Bojorque et al., 2022), I expect to test 50 neonates/infants per experiment (to obtain at least 30 valid datapoints), and 40 rats per experiment. Final sample sizes will be determined by appropriate power analyses from pilot data.

3.1. Do human neonates and infants use MOP to parse the speech into syllable-sized units? MOP is claimed to be a universal principle of speech parsing, stating that consonants are preferred at the onset of syllables. However, there seems to be no available evidence of the use of such principle within the first months of life. I will test the hypothesis that neonates and infants are sensitive to maximal onset constraints when detecting syllables from speech. If this is the case, I predict neonates and infants to respond differently to correctly-segmented syllables according to maximal onset constraints vs. syllables violating such constraints.

Experiment 1. Neonates' neural responses will be recorded with an innovative optical neuroimaging technique that combine functional Diffused Correlation Spectroscopy (fDCS) and functional Near-Infrared Spectroscopy (fNIRS). fDCS has been developed by the Institute of Photonic Sciences (ICFO, Barcelona; Durduran et al., 2010; Durduran & Yodh, 2014) and its feasibility in non-clinical infant research has been assessed in a collaborative project with Drs. Turgut Durduran, Nuria Sebastian-Galles and the PI (Zhang et al., 2022; *in prep*). Combining fDCS with fNIRS provides the following advantages: **(1) assessing both cerebral blood flow and blood oxygenation**, hence being able

to derive cerebral metabolic rate of oxygen extraction, important measures of hemodynamic activity in the brain in response to stimuli (fNIRS only allows to measure blood oxygenation), 2) deeper penetration of near-infrared light onto the head scalp, providing greater measure of brain activity, 3) covering bigger areas of the brain with respect to commercial fNIRS devices. The fDCS/fNIRS combined system has recently been extended to cover the whole infant brain with a high-density optode arrangement (H2020 TinyBrains 101017113, PI: Dr. Durduran). I will use modules from TinyBrains inserted in a headband that can easily be placed on the infant head. Channels will record hemodynamic activity from bilateral temporal regions, where linguistic and non-linguistic auditory processing occur (e.g., Dehaene-Lambertz et al., 2002; Friederici et al., 2002; Dehaene-Lambertz et al., 2008). A block design including a familiarization and a test phase will be implemented (Gervain et al., 2008; Benavides-Varela et al., 2011b). **Familiarization.** Neonates will listen to a sequence containing a whole word repeated several times (e.g., *prosper – prosper – prosper* etc.) for 10 blocks of 6 words each. Words within blocks will be separated by silence of varying duration (0.5-1.5s). Blocks will be separated by 20-22s of silence, a necessary time-window to restore hemodynamic response in the brain. Familiarization will last ~5 min. **Test.** Artificially segmented words will be presented as test stimuli, created by inserting a silence (~500 ms) in the middle of each word that will generate two types of segmentation options: e.g., *pro_sper* (correct segmentation according to MOP) and **pros_per* (incorrect segmentation). Test stimuli will be presented in 4 alternating test blocks, in which correctly-segmented and incorrectly-segmented syllables will be presented in alternation within each block, and 4 non-alternating blocks, in which only one syllable type (either correctly-segmented or incorrectly-segmented) will be presented within each block. A block could start with one of the syllable type, and order will be counterbalanced across participants. Test blocks will include 12 syllables (1 correctly-segmented and 1 incorrectly-segmented, for each of the 6 words presented at familiarization). Neonates will be tested individually within 1-5 days after birth, at BCNatal Fetal Medicine Research Center (Barcelona), while in their crib in a silent room assisted by medical staff. Two loudspeakers will be located at 1m from the infant's head, at an angle of 30°, and at the same height of the crib. A computer will operate the fDCS/fNIRS system, and deliver stimuli using a custom-made Matlab script. Data acquisition and analysis will be conducted using custom-made software and pipeline developed by ICFO (Zhang et al., 2022; *in prep*). Regions of interest (ROIs) will be determined based on previous research and pilot data. Main steps of the analysis include calculating blood flow index (BFI), and blood oxygen saturation (StO₂; ratio of oxygenated hemoglobin concentration to total) for all channels, normalizing the indexes by subtracting a baseline (~2-4s prior to the onset of stimuli). Relative BFI and StO₂ indexes will be either presented as 2D topographic map or averaged across all channels when homogeneous. Screening of the quality of the signal will be performed to control for artifacts (e.g., infant heartbeat, slow oscillations of blood flow). Data will be analyzed with repeated-measures Anova with blocks (alternating/non-alternating) and channels (left/right) as within-subject factors. Differences in hemodynamic activity in response to correctly-segmented and incorrectly-segmented syllables will indicate discrimination.

Experiment 2 and 3. Two experiments will be conducted with 6-8 month-old infants. If neonates in Exp. 1 do not show differences in response to the two parsing options, the present experiments will allow to test alternative explanations. Exp. 2 will test the sensitivity of fDCS/fNIRS to assess the process under investigation as well as to explore the role of experience with language. It could be the case that neonates need additional experience with speech to properly detect syllable boundaries. Exp. 3 will measure infant preferences for one of the parsing options; preferences are extremely difficult to assess with newborns. **Experiment 2** will be conducted with same materials, paradigm, apparatus used in Exp. 1, at UPF BabyLab, as in previous studies (Zhang et al., 2022; *in prep*). The fNIRS/fDCS system is suitable for testing infants of 6-8 months of age. The headband with the channels will be placed on the infant's head, who will be sitting on the caregiver's lap in a sound-attenuated room. A computer screen will present a 10-min silent cartoon movie to keep the infant engaged during the experiment. Differences in hemodynamic activity (blood flow and oxygenation) in response to correctly-segmented (e.g., *pro_sper*) and incorrectly-segmented (e.g., **pros_per*) syllables will

indicate discrimination, thus sensitivity to maximal onset constraints to identify syllabic units from speech. **Experiment 3** will be conducted on another group of 6-8 month-olds, with a behavioral procedure (Head-turn Preference) and same materials used in Exp.1-2. The infant will be sitting on the caregiver's lap in a sound-attenuated room equipped with three computer screens, one facing the infant, the other two on the left and right side of the infant. During a **familiarization** phase, the infant will listen to the sequence of words for 3min, while neutral images (i.e., moving clouds) will appear on the central screen. Visual stimuli are not related to auditory stimuli and only serve to keep the infant engaged with the task (e.g., Thiessen & Saffran, 2003; Santolin & Saffran, 2019, for similar procedures). Within the sequence, words will be separated by 1.5s-silence. At **test**, the two parsing options will be presented in 12 trials, divided in 3 blocks of 4 trials each. Within a block, two types of test stimuli will be presented: correctly-segmented syllables (e.g., *pro_sper*) and incorrectly-segmented syllables (e.g., **pros_per*). A trials will begin with a visual stimulus appearing on the central screen (e.g., blinking light). When the infant will fixate on it, an experimenter will extinguish the light, which will appear in one of the side screens. When the infant will turn the head towards the side light, one of the test stimuli will be presented. A trial will end when the infant will look away from the screen for more than 2s, or until max trial duration will elapse (usually between 18 to 24s). The procedure will be set up in WISP, a custom-designed Matlab software (Olson, 2017) which will be also used to code looking times, and pre-process raw looking time data. Equipment and softwares are currently available at UPF BabyLab. Looking times at screens when correctly-segmented and incorrectly-segmented test syllables are presented will be analyzed using paired-sample t-tests. If infants have a preference for the correct segmentation according to MOP, longer looking times for correctly-segmented syllables would be expected. Note that predicting direction of preference is not always trivial in infancy, as infants sometimes look longer at stimuli that are unexpected or surprising, rather than at familiar ones (Hunter & Ames, 1988; Aslin, 2007; Santolin et al., 2021). For example, if infants find **pros_per* unexpected (since it violates MOP), they may show longer looking times for it.

Results of this set of experiments will indicate whether MOP is actually in place at birth and within the first months of life, when infants begin to process speech.

3.2. Is MOP specific to language? Maximal onset constraints might derive from acoustic regularities of the speech signal, according to which a syllable unit must start with analogues of consonant-like sounds, possibly to facilitate a transition to subsequent sounds with higher intensity (vowel-like). This creates an arch-shaped pattern that simulates the envelopes of several animal vocalizations. To test this hypothesis, I will investigate a nonhuman species' (rats) sensitivity to MOP, with same materials described for neonates and infants. If rats pick up on such regularities, they should show similar sensitivity to correctly-segmented syllables following maximal onset constraints.

Experiment 4. Long-Evans rats (*Rattus norvegicus*) will be tested with a familiarization-discrimination procedure including 3 phases: shaping, familiarization, test. **Shaping.** Prior to the beginning of the experiment, rats will be familiarized with the setting and learn the nose-poking response, during a 10-days training phase. Animals will be placed individually in an experimental chamber for 30 minutes a day, during which a sugar pellet will be presented every minute, or when the animal will poke the feeder with its nose (the response is detected by an infrared detector placed in the feeder). Stimuli will be delivered through stereo amplifier and loudspeakers located adjacent to the chambers. A custom-made software (available at PRBB – Parc Recerca Biomedica Barcelona, where all the rat experiments will be conducted) will be used to present stimuli, record nose-poking responses, and deliver food. **Familiarization.** Rats will listen to the sequences of words (e.g., *prosper* – *prosper* – *prosper*) separated by 10-sec inter stimulus interval (ISI) for 20 days. Food reward will be delivered if nose-poking responses will be emitted. **Test.** Rats will be presented with two types of test stimuli: **correctly-segmented** syllables (e.g., *pro_sper*) and **incorrectly-segmented** syllables (e.g., **pros_per*), presented in 8 trials in which a single stimulus will be repeated (~5 times). No food reward will be given. To avoid response extinction, test stimuli will be interleaved with

familiarization stimuli, which will be rewarded. Discrimination between test stimuli will be measured as a function of number of nose-poking responses for each syllable type (correctly-segmented vs. incorrectly-segmented), and analyzed with paired-samples t-test. Differences in number of responses will indicate that rats recognize one of the syllable types (either *pro_sper* or **pros_per*) as better segmentation option for the familiarization words (e.g., *prosper*). This will suggest that rats are sensitive to maximal onset constraints defining syllable boundaries. Instead, if rats show no differences in responses to test stimuli, it may be because they process both types of segmented syllables as viable parsing options.

Experiments 5 and 6. Two experiments will be conducted to control for potential lack of differences in Exp. 4, which may hinder a lack of discrimination between test stimuli. Do rats process correctly- and incorrectly-segmented syllables as same instances of the words heard at familiarization? Two new groups of rats will be tested with the same familiarization-discrimination procedure described in Exp. 4 but will be **directly familiarized with the artificially-segmented syllables** (in **Experiment 5**, with correctly-segmented, e.g., *pro_sper*; in **Experiment 6**, with incorrectly-segmented, **pros_per*). Both types of syllables will be also used at test as familiar and novel stimuli (correctly-segmented ones will be familiar at test in Exp. 5 and novel in Exp. 6, and incorrectly-segmented ones will be novel in Exp. 5 and familiar in Exp. 6). If rats can discriminate between correctly- and incorrectly-segmented stimuli, they should show a greater number of nose-pokes for the familiarization ones.

Results of this set of experiments will reveal whether rats are sensitive to MOP, and whether such linguistic constraints is rooted into general mechanisms of acoustic processing of regularities of the signal.

3.3. Can nonhuman animals learn language-specific phonotactics on the segmented output?

These experiments will investigate whether rats are sensitive to language-specific constraints that allows to “repair” incorrectly-segmented syllables. The segmented output provided by MOP, in fact, is a unit of speech that do not always constitute a possible syllable for all languages. For instance, Spanish does not allow /sp/ and other consonant clusters as syllable or word onset. Japanese has even more strict rules on consonants, and would break all the consonant clusters by inserting vowels in between, e.g., *prosper* would be *Pu_RO_Su_PE_Ru* (Dupoux et al., 1999; Mazuka et al., 2014). As much as such “repairs” are language-specific (LaCharité & Paradis, 2005), from a purely acoustic perspective, the insertion of a vowel (before or after consonants, depending on the language) constitutes the addition of a sound, and thus could be simply perceived as such. Rats are an excellent model to investigate if linguistic repair strategies can be computed in the absence of a language system, or whether experience with language is necessary.

Experiment 7 and 8. Rats will be tested with the same familiarization-discrimination procedure described in Exp. 4. In **Experiment 7**, rats will be familiarized with **incorrectly-segmented syllables** with onset consonant clusters that are not accepted in many languages (e.g., *RBAN*). For example, according to MOP, the word *urban* could be parsed as *u_rban* because /rb/ would provide a maximal onset to the second syllable. However, /rb/ is not tolerated as syllable onset in many languages because it violates the sonority hierarchy. After familiarization, rats will be tested for discrimination between incorrectly-segmented syllables vs. **linguistically-repaired versions** of such syllables with the addition of a prothetic vowel (e.g., *RBAN* vs. *eRBAN*). In **Experiment 8**, instead, rats will be familiarized with **correctly-segmented syllables** (e.g., *BRAN*), having syllable onsets that match with sonority hierarchy, and tested with these same stimuli vs. **unnecessarily-repaired versions** (e.g., *eBRAN*). Shaping, Familiarization and Test will be conducted as in Exp. 4 as well as data analysis. If rats process the prothetic vowel as an additional sound, they should discriminate in both experiments (e.g., showing significantly different number of nose-pokes in response to *RBAN* vs. *eRBAN*, and in response to *BRAN* vs. *eBRAN*) since all syllables differ at the acoustic level. The contrast used in Exp. 8 is even more interesting because it compares linguistically-identical syllables: for instance, both

BRAN and *eBRAN* have well-accepted onsets across most languages, but they vary acoustically. Results of these experiments will provide critical information about the perceptual foundations of language-specific constraints operating on syllables.

Experiment 9. The goal of this experiment is to investigate whether language-specific phonotactic constraints can be learned in the absence of a language system. I will implement the Spanish prothetic vowel /e/ in a miniature language composed of real Spanish words that will include only possible consonant clusters in Spanish (e.g., *eSP*, *eST*, *eSF* as in *estudio*, *España*, *esfera*, as opposed to *study*, *Spain*, *sphere* in English). I will test rats using the same familiarization-discrimination procedure described in Exp. 4, with a longer familiarization phase. Given the complexity of the stimulus, the animals will listen to the language for 1 month. The language will contain ~20 words, separated by 1.5s silence-break to make words stand out. At test, rats will be presented with **correctly-segmented words according to Spanish** (e.g., *eS_PA_ÑA*) vs. **incorrectly-segmented words** (e.g., *e_SPA_ÑA*). Crucially, in the latter, the prothetic vowel is separated from the consonant cluster, creating an illegal syllable onset which is not experienced during familiarization. Shaping, Familiarization and Test will be conducted as in Exp. 4 as well as data analysis. If rats are able to learn the distribution of Spanish phonemes during familiarization, they should reject incorrectly-segmented words (e.g., *e_SPA_ÑA*) and show significantly higher number of responses for correctly-segmented words. If rats do not show differences at test, it could be because they may not be sensitive to language-specific phonotactics or the familiarization language is too difficult to learn from. To disentangle these options, an additional group of rats would be familiarized with a new miniature language containing syllable onsets that both match and violate Spanish phonotactics (e.g., words beginning with /*SP*/, /*ST*/, /*SF*/ as well as /*eSP*/, /*eST*/, /*eSF*/). If no differences occur at test (for example, rats process /*SP*/ and /*eSP*/ as identical), this would indicate that rats are not sensitive to language-specific phonotactics that regulate the structure of syllables. Either way, such results will complement results of Exp. 7-8, supplying new evidence of the interplay of experience with language and predispositions when learning linguistic constraints.

4. Research Line 2: **TESTING THE SENSORY BASES OF THE SONORITY HIERARCHY (SH).**

In this series of experiments, I will test the hypothesis that sensitivity to the Sonority Hierarchy (SH), that modulate and refine the structure of syllables, is based on general sensory processing which would be grounded on ancient mechanisms present in humans at birth, and in other animal species. The intensity pattern defining the universal syllabic structure (*rise* → *fall*) will be implemented in auditory non-linguistic (Exp. 10-13) and vibro-tactile (Exp. 14-15) inputs. Sensitivity to such pattern and its violations will be tested in humans (neonates, infants, adults) and non-human species (rats). Materials will consist of non-linguistic analogues of speech syllables implemented in the two sensory modalities. Vibro-tactile stimuli are highly ecological for rats, for which the vibrissal sensory system is central for perception (Diamond et al., 2008). My hypothesis would predict sensitivity to violations of the *rise* → *fall* intensity pattern (matching SH) in non-linguistic inputs across species. This would suggest that the sonority hierarchy emerges from general sensory processing that is not tailored specifically for learning linguistic structures. As for RL.1, I expect to test 50 neonates/infants (to obtain at least 30 valid datapoints) and 40 rats per experiment, and 30 adults. Final sample sizes will be determined by appropriate power analyses from pilot data.

4.1. Are rats, human neonates and infants sensitive to Sonority Hierarchy in acoustic non-linguistic inputs? Experiments 10a and 10b. Rats will be presented with strings of **pure tones** implementing two patterns of intensity: *rise* → *fall*, in which tones will increase their intensity at onset and drop right after, simulating well-formed linguistic syllables (**well-formed tone strings**), and *fall* → *rise*, in which tones will decrease their intensity at onset and rise afterwards, mirroring ill-formed syllables uncommon across languages (**ill-formed tone strings**). Pure tones of the same

octave and length will be created using a sine-wave generator (Audacity), and will be combined into strings of four tones, mimicking the CVCC syllable structure (C=consonant, V=vowel). Experiments will be conducted with the familiarization-discrimination procedure described in Exp. 4. In **Experiment 10a**, rats will be familiarized with well-formed tone strings, in **Experiment 10b**, rats will be familiarized with ill-formed tone strings. For both experiments, test stimuli will be 8 new tone strings that either match or violate the intensity patterns experienced at familiarization. Discrimination between familiar vs. novel test strings will be measured as a function of number of nose-poking responses for each type of test strings. Discrimination within experiments will be analyzed with paired-sample t-tests. If sensitivity to SH defining strings of tones is in place, rats familiarized to well-formed strings should discriminate at test, showing a significantly higher number of responses to familiar strings matching the *rise* \rightarrow *fall* pattern (Exp. 10a). Rats familiarized with ill-formed strings, instead, may find the tone strings presented at test harder to discriminate (Exp. 10b). Overall performance at test (i.e., total number of nose-poking responses) between experiments will be analyzed with independent-samples t-test to see if rats would respond significantly more or less at test as a function of the familiarization stimulus. Being exposed to well-formed strings may have a processing advantage (e.g., more interesting for rats to learn) because such stimuli depict an arch-shaped envelope similar to that characterizing many sounds and calls of the wild.

Experiment 11. This experiment will be conducted with human neonates with similar neuroimaging procedure described in Exp. 1, and same non-linguistic materials (pure tone strings) described in Exp. 10a-b. Neonates will listen to **well-formed tone strings** and **ill-formed tone strings** presented in blocks. A block will include 8 strings of one or the other string type, presented in random order. As in Exp. 1, silent breaks will separate strings (0.5-1.5s) and blocks (20-22s). Ten blocks of well-formed tone strings and 10 blocks of ill-formed tone strings will be presented (no prior familiarization phase will be needed). Neonates' neural responses to the two types of strings will be recorded with the fDCS/fNIRS combined system. Data acquisition and pre-processing will be conducted similarly to Exp. 1 (following pipeline developed by ICFO; Zhang et al., 2022; *in prep*). Paired-samples t-tests will be computed to compare hemodynamic activity (averaged across channels) for the two types of tone strings. Differences in hemodynamic activity in response to well-formed and ill-formed tone strings will indicate discrimination, suggesting that newborns are sensitive to differences in sonority hierarchy in auditory non-linguistic inputs.

Experiment 12 and 13. Two experiments will be conducted with 4-5 month-old infants. Exp. 12 will be conducted with fNIRS/fDCS to test the sensitivity of the combined system to measure the type of processing under investigation with older infants, in case newborns do not show any differences in Exp. 11. Exp. 13 will be conducted with a behavioral procedure that will allow to assess preferences between stimuli that follow vs. violate SH. **Experiment 12** will be conducted as Exp. 11, with identical materials, design, data acquisition and analysis. Differences in cerebral hemodynamics in response to well-formed tone strings and ill-formed tone strings will indicate that 4-5 month-old infants discriminate between intensity patterns depicted by the stimuli. This result would complement results of Exp. 11, indicating that infants within the first few months of life are sensitive to SH in nonlinguistic acoustic signals. **Experiment 13** will be conducted with another group of infants to assess whether spontaneous preference well-formed tone strings or ill-formed tone strings. I will use the Infant-Controlled Preferential Looking procedure to compare looking times in conjunction with the string types, using automatic recording of eye movements with an eye-tracker (for similar procedures: Santolin et al., 2019; Zacharaki & Sebastian-Galles, 2022). This procedure adapts best to participants' young age. No familiarization phase prior to testing will be necessary. Infants will be sitting on a high-chair in a sound-attenuated room equipped with a computer screen placed in front of the chair (60cm distance), and two loudspeakers located at both sides of the screen. An eye-tracker (Tobii Pro Spectrum, 120 Hz sampling rate, available at UPF BabyLab) will be used to record eye-movements. After standard 5-point calibration phase, there will be 12 trials, 6 per each type of tone string. A trial begins with a visual attention getter appearing on the screen (e.g., a blue circle); when

the infant will look at it, another visual stimulus will pop up, either on left or right side of the same screen, along with the acoustic stimulus (either a well-formed or an ill-formed string). Both visual and acoustic stimuli will terminate when infant will look away for more than 2 s. A custom-made Matlab script will control the contingent presentation of visual and acoustic stimuli, and the timing of the experiment. Raw eye-tracker data will be processed in RStudio (RStudio Team, 2019) using the *eyetrackingR* package. Looking times will be computed only when infants look inside the Areas of Interests (i.e., areas on the screen occupied by the visual stimulus); looking elsewhere will be considered as looking away time. Looking times for well-formed tone strings and ill-formed tone strings will be analyzed with paired-samples t-test. Longer looking times for one of the string types will indicate preference.

This set of experiments will shed light into the perceptual bases of the human bias for sonority constraints that shape the structure of syllables (Gomez et al., 2014; Zhao & Berent, 2018). Such constraints could be apparent in other domains and species, thus rooted into general sensory processing mechanisms and preferences for natural patterns of intensity (e.g., arch-shaped envelopes widespread in nature).

4.2. Are rats and humans sensitive to Sonority Hierarchy in vibro-tactile inputs? **Experiment**

14. A group of rats will be tested with an innovative methodology that will allow to deliver **sinusoidal vibrations** to the animals' whiskers (for similar procedure: Adibi, Diamond & Arabzadeh, 2011). Prior research has shown that rats can detect vibrations presented to their whiskers differing in frequency and amplitude (Hutson & Masterton, 1986; Adibi & Arabzadeh, 2011), making this procedure suitable for the current research goals. I will assess rats' ability to discriminate between two types of stimuli: **well-formed vibration sequences**, that will implement the *rise* \rightarrow *fall* pattern (matching SH), and **ill-formed vibration sequences**, that will implement the *fall* \rightarrow *rise* pattern (violating SH). Stimuli will be generated in Matlab with an analog output (sampling rate of 44.1-kHz) using the intensity envelopes of the tone strings used in Exp. 10-13. Procedure and apparatus are illustrated in Figure 3. The animals will be placed individually in an experimental chamber equipped with a plexiglass wall with an aperture. Two mesh plates will be attached to the edges of the aperture, inclined toward each other such that, when the rat pokes its nose, the whiskers will touch the meshes. Meshes will be connected to piezoelectric ceramic bars that will deliver vibrations (sent through an amplifier) simultaneously to both whisker pads. **Shaping.** Rats will be familiarized with the setting and learn to nose-poke in the aperture (10-15 days). Nose-pokes will be recorded by an optical

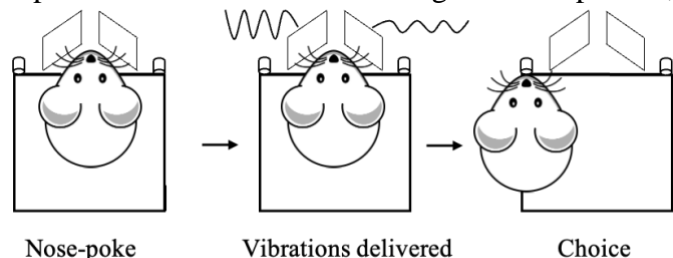


Figure 3. Illustration of 'whisker procedure'






sensor that will activate the feeders placed at both sides of the aperture to deliver food. **Test.** A nose-poke through the aperture will start the simultaneous presentation of vibration sequences. The sequence presented to the left whisker pad will be either well-formed or ill-formed, and the sequence presented to the right whisker pad will be the other one. For half of the rats, S+ (reinforced stimulus) will be the well-formed sequences and Sn (neutral) will be the ill-formed sequences; viceversa for the other half of the animals. Rats will respond by choosing the left or right associated feeder. Correct choices will be rewarded whereas incorrect choices (i.e., approaching the Sn feeder) will be associated with ~4s-delay for the beginning of the next trial. Eight well-formed vibration sequences and 8 ill-formed vibration sequences will be presented. Valid choices will be coded as the first attempt to get the reward from the corresponding feeder after test stimulus presentation (between 0.5s to 8s from stimulus onset). A Matlab script will control stimuli presentation and food delivery, and record rats' behavior. Data (correct choices for S+) will be analyzed with Chi-square tests. An infra-red camera will record the experiment. If rats are able to discriminate between well-formed and ill-formed vibration sequences, they should recognize the corresponding S+ (correctly identify the associated

feeder). This result will indicate that rats can differentiate between *rise* → *fall* and *fall* → *rise* patterns of intensity (that simulate SH of linguistic syllables) in vibration sensory modality.

Experiment 15. This experiment will be conducted with human adult participants. The acoustic strings used in Exp. 10-13 will be presented as sinusoidal vibrations through an amplifier connected to a computer, and to Oticon-A bone conduction vibrators (Oticon). Vibrators have a 3.8 cm² vibrating surface and will be placed in the dorsal side of the index fingers (below the nail) of participants' hands, secured with medical tape (Ruzzoli & Soto-Faraco, 2017 for similar procedure; see also Rantala et al., 2013). Vibrations will be presented at 200 Hz (that guarantee vibro-tactile perception in humans). Importantly, participants will be wearing noise-cancellation headphones to mask noise produced by the vibrators to avoid acoustic cues potentially interfering with the task. A two-alternative forced-choice procedure will be implemented, with a same-different design. There will be 16 trials, each one including well-formed and ill-formed vibration sequences, with 2-sec pause between sequences. The first sequence type will be used as 'reference' and the second one as 'comparison' (order will be counterbalanced across participants). Participants' task will be to report whether the reference stimulus is same or different than the comparison stimulus, by pressing designated keys on a computer keyboard. A binomial regression model will be used to analyze data, calculating the probability of correct responses as dependent variable. The probability of responding 'Same' when the two stimuli are identical will be scored as 1, and the probability of responding 'Different' when the two stimuli are different will be scored as 0. Order will be added as predictor of the model, and sequence type (well-formed vs. ill-formed) and participants as random effects (following Barr, 2013). Additionally, preference ratings of the stimuli will be scored at the end of the experiment. In addition, as a proxy of preference, participants will be instructed to rate well-formed and ill-formed sequences (presented as pairs, as described above) using a scale from 0 (unpleasant) to 10 (pleasant; 5 will indicate neutral). The scale will be presented visually, and participants will have to drag a cursor to select their rating, which could be any value in between 0 and 10 (including decimals), generating a continuous variable. The sequence type to be rated as first within each pair will be counterbalanced across participants. Data will be analyzed with repeated-measure Anova, with rating values as dependent variable, and the first sequence type as between-subject factor. Results of this experiment will be pivotal to develop a protocol for testing vibro-tactile perception of linguistic structures in neonates and infants. These results, along with findings on rats (Exp. 14), will provide new evidence about whether sonority constraints defining natural units of speech, emerge from general sensory processing that are not designed specifically for language.

All the experiments of the GALA project will provide new knowledge that will be pivotal to understand the nature of the mechanisms operating at the entry-gate of language, through which humans begin to discover linguistic structures.

Table 1. Summary of experiments, populations, materials and research goals for **RL. 1** (blue shades) and **RL.2** (red shades).

Experiment	Population	Stimuli	What is tested
1 2 3		Fam: <i>prosper</i> test: <i>pro_sper</i> vs. <i>pros_per</i>	Sensitivity to MOP (<i>pro_sper</i> vs. <i>pros_per</i>)
4 5 6 7 8 9		Discrimination <i>pro_sper</i> Discrimination <i>pros_per</i> <i>rban</i> vs. <i>erban</i> <i>bran</i> vs. <i>ebran</i> Training: ' <i>españa</i> '	control control Sensitivity to linguistic repair strategy Sensitivity to language-specific phonotactics
10a 10b		pure tones (well/ill-formed)	Sensitivity to acoustic SH (well- vs. ill-formed)
11 12 13			
14 15		vibrations (well/ill-formed)	Sensitivity to vibro-tactile SH (well- vs. ill-formed)

6. Risk management. The GALA project builds upon strong theoretical knowledge, and solid experimental background that I have developed during the past 7 years of research with animal and infant populations (see B1 for track record). Since the project includes novel experimental paradigms with different species, and humans at different timepoints in development, I have allocated the necessary amount of time in planning and implementing procedures and materials. The materials will be similar across experiments within research lines, and experimental procedures will be adapted to the capacities of each population. Conditioning procedures used with rats (Exp. 4-9, 10a-b) will be set up modelling similar procedures used in the past +10 years by Dr. Juan Manuel Toro. The ‘whisker’ procedure (Exp. 14) is relatively new; it comprises direct stimulation of rats’ vibrissae as well as several hardware and software pieces that must be coordinated. Given my expertise in animal research, the hiring of a technician-engineer, and the excellent infrastructures of the Host Institution (PRBB), I do not foresee unsurmountable difficulties. The fDCS/fNIRS system (Exp. 1-2, 11-12) will be used for the first time in non-clinical research. The system will be fully developed and tested with neonates by mid-2023 by the TinyBrains EU-H2020 Consortium (tinybrains.eu) led by Dr. Durduran, who will collaborate in the GALA project. Given the ambitious aspect of this procedure, I have prepared contingency measures and backup plans. Dr. Judit Gervain has agreed to be involved as *advisor*. Being one of the world’s specialists on fNIRS research with newborns and infants, she will assess the implementation of the procedure, and provide feedbacks on data analysis and results. It is worth noticing that GALA would start several months after the expected date the system will be operational (and likely, would start in 2024). Collecting infant data (and in particular neuroimaging data) is quite time-consuming therefore sufficient time for neonate and infant studies has been allocated in the project.

Despite being comparative, I designed this research project such that the animal and the human sets of experiments can also stand on their own, without diminishing the relevance of the results that will be obtained. Experiments with neonates/infants and rats can be conducted simultaneously.

This project includes a variety of populations and experimental procedures. I will be able to count on the support of 3 PIs of the Host Institution who have extensive experience with such populations and methodologies. Dr. Juan Manuel Toro is an expert on perceptual and cognitive research with rats, involving most of the procedures described here ([web](#)). Dr. Nuria Sebastian-Galles is recognized worldwide for her research on infant and adult language acquisition ([web](#)). Dr. Salvador Soto-Faraco

has +20 years of experience in multisensory research with adults, including vibro-tactile perception ([web](#)). All of them have been ERC grantees. Dr. Turgut Durduran (ICFO, [web](#)) and Dr. Judit Gervain (University of Padova, [web](#)) will assess the neuroimaging studies, being both leading experts on fDCS/fNIRS, and Dr. Elisenda Eixarch (BCNatal Fetal Medicine Research Center-Hospital Clínic/Hospital Sant Joan de Déu, Universitat de Barcelona, [web](#)) will provide access to neonates and assess the procedures. I have collaborated with most of them in the past, and they are thrilled to assess the research whenever necessary.

	Populations (<i>procedures</i>)	YEAR 1		YEAR 2		YEAR 3		YEAR 4		YEAR 5	
		1st sem	2nd sem	1st sem	2nd sem	1st sem	2nd sem	1st sem	2nd sem	1st sem	2nd sem
RL1	Neonates (<i>neuroimaging</i>)	Postdoc									
			<i>RA-M</i>								
				Predoc1							
	Infants (<i>neuroimag-behav</i>)					Postdoc (neuroimag)					
RL2						Predoc1 (neuroimag)		Predoc1 (behav)			
						<i>RA-M</i>					
	Rats (<i>conditioning</i>)					Predoc2					
								<i>RA</i>			
RL2	Neonates (<i>neuroimaging</i>)			Postdoc							
				<i>RA-M</i>							
				Predoc1							
	Infants (<i>neuroimag-behav</i>)							Postdoc (neuroimag)			
RL2								Predoc1 (neuroimag)		Predoc1 (behav)	
								<i>RA-M</i>			
	Rats (<i>whisker</i>)	Tech (-PI)	Predoc2								
				<i>RA</i>							
RL2	Adults (<i>vibrotactile</i>)										
								Predoc 2			

Figure 4. Timeline of GALA. Tasks of personnel include preparation of materials, implementation of procedures, piloting, testing, data analysis, writing. **Bold** indicate the main responsible of the research. **RA-M**: research assistant-manager. **RA**: research assistant. Some overlap in the timeline between personnel is based on the fact that neonate/infant testing does not occur with daily fixed schedule (as it mainly depends on birth frequency and families' availability). Rat experiments require less time of implementation, piloting, testing than neonate/infant research.

7. Research impact. The GALA project represents a new approach to the emergence of language in humans. Its novelty consists in exploring the biological nature of mechanisms at the 'entry-gate' of language, through comparative-developmental research involving human newborns, infants, adults and nonhuman animals. Most of the research investigating the uniquely-human features of language has focussed on the study of complex linguistic components (e.g., syntax, recursion). This project, instead, investigates the fundamentals of how humans *begin* to learn language. The project is highly interdisciplinary, bringing together theories and methodologies from Developmental Psychology, Animal Cognition, Linguistics and Neuroscience. The fDCS/fNIRS combined system and the 'whisker procedure' are innovative features of the proposal. fDCS/fNIRS provides simultaneous measures of brain activity that can only be assessed separately nowadays in non-clinical infant studies. This represents a break-through in psychological research. The 'whisker procedure' allows to use highly ecological signals for the animal model (rat), which is not always the case in comparative research exploring language evolution, in which animals are typically presented with speech inputs. This represents another innovative aspect of the proposal. The newborn-rat comparison is quite straightforward, as it allows to simultaneously explore ontogenetic and phylogenetic aspects of linguistic mechanisms and constraints at the entry-gate of language.

The GALA project will produce invaluable new knowledge of the mechanisms operating at the emergence of human language, and will provide fundamental contributions to the unsolved question of why are humans the only species so far able to learn language.

References

- Abe, K., & Watanabe, D. (2011). Songbirds possess the spontaneous ability to discriminate syntactic rules. *Nature neuroscience*, 14(8), 1067-1074.
- Adibi, M., & Arabzadeh, E. (2011). A comparison of neuronal and behavioral detection and discrimination performances in rat whisker system. *Journal of neurophysiology*, 105(1), 356-365.
- Adibi, M., Diamond, M. E., & Arabzadeh, E. (2012). Behavioral study of whisker-mediated vibration sensation in rats. *Proceedings of the National Academy of Sciences*, 109(3), 971-976.
- Agrillo, C., & Bisazza, A. (2018). Understanding the origin of number sense: a review of fish studies. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1740), 20160511.
- Aslin, R. N. (2007). What's in a look?. *Developmental science*, 10(1), 48-53.
- Benavides-Varela, S., Gómez, D. M., & Mehler, J. (2011). Studying neonates' language and memory capacities with functional near-infrared spectroscopy. *Frontiers in Psychology*, 2, 64.
- Benavides-Varela, S., Gómez, D. M., Macagno, F., Bion, R. A., Peretz, I., & Mehler, J. (2011b). Memory in the neonate brain. *PLoS One*, 6(11), e27497.
- Bertoncini, J., Bijeljac-Babic, R., Jusczyk, P. W., Kennedy, L. J., & Mehler, J. (1988). An investigation of young infants' perceptual representations of speech sounds. *Journal of Experimental Psychology: General*, 117(1), 21.
- Bijeljac-Babic, R., Bertoncini, J., & Mehler, J. (1993). How do 4-day-old infants categorize multisyllabic utterances?. *Developmental Psychology*, 29(4), 711.
- Boros, M., Magyari, L., Török, D., Bozsik, A., Deme, A., & Andics, A. (2021). Neural processes underlying statistical learning for speech segmentation in dogs. *Current Biology*, 31(24), 5512-5521.
- Brudzynski, S. M. (2013). Ethotransmission: communication of emotional states through ultrasonic vocalization in rats. *Current opinion in neurobiology*, 23(3), 310-317.
- Call, J. (2000). Estimating and operating on discrete quantities in orangutans (*Pongo pygmaeus*). *Journal of Comparative Psychology*, 114(2), 136.
- Celma-Mirallès, A., & Toro, J. M. (2020). Non-human animals detect the rhythmic structure of a familiar tune. *Psychonomic Bulletin & Review*, 27(4), 694-699.
- Christiansen, M. & Kirby, S. (2003). Language evolution: Consensus and controversies. *Trends in cognitive sciences*, 7(7), 300-307.
- Clements, G. (1990). The role of the sonority cycle in core syllabification. *Papers in Laboratory Phonology*, 1, 283-333.
- Comins, J. A., & Gentner, T. Q. (2013). Perceptual categories enable pattern generalization in songbirds. *Cognition*, 128(2), 113-118.
- Crespo-Bojorque, P., Celma-Mirallès, A., & Toro, J. M. (2022). Detecting surface changes in a familiar tune: exploring pitch, tempo and timbre. *Animal Cognition*, 1-10.
- Cuaya, L. V., Hernández-Pérez, R., Boros, M., Deme, A., & Andics, A. (2022). Speech naturalness detection and language representation in the dog brain. *NeuroImage*, 248, 118811.
- Dehaene-Lambertz, G., Dehaene, S., & Hertz-Pannier, L. (2002). Functional neuroimaging of speech perception in infants. *science*, 298(5600), 2013-2015.
- Dehaene-Lambertz, G., Hertz-Pannier, L., Dubois, J., & Dehaene, S. (2008). How does early brain organization promote language acquisition in humans?. *European Review*, 16(4), 399-411.

- Diamond, M. E., Von Heimendahl, M., Knutsen, P. M., Kleinfeld, D., & Ahissar, E. (2008). 'Where and what' in the whisker sensorimotor system. *Nature Reviews Neuroscience*, 9(8), 601-612.
- Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C., & Mehler, J. (1999). Epenthetic vowels in Japanese: a perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance*, 25, 1568-1578.
- Durduran, T., Zhou, C., Buckley, E. M., Kim, M. N., Yu, G., Choe, R., ... & Licht, D. J. (2010). Optical measurement of cerebral hemodynamics and oxygen metabolism in neonates with congenital heart defects. *Journal of biomedical optics*, 15(3), 037004.
- Durduran, T., Turgut, & Yodh, A. G. (2014). Diffuse correlation spectroscopy for non-invasive, micro-vascular cerebral blood flow measurement. *NeuroImage*, 85, 51.
- Durduran, T., & Yodh, A. G. (2014). Diffuse correlation spectroscopy for non-invasive, micro-vascular cerebral blood flow measurement. *Neuroimage*, 85, 51-63.
- Emery, N. J., & Clayton, N. S. (2009). Tool use and physical cognition in birds and mammals. *Current opinion in neurobiology*, 19(1), 27-33.
- Fló, A., Benjamin, L., Palu, M., & Dehaene-Lambertz, G. (2022). Sleeping neonates track transitional probabilities in speech but only retain the first syllable of words. *Scientific Reports*, 12(1), 1-13.
- Fitch, T. (2018). The biology and evolution of speech: a comparative analysis. *Annual review of linguistics*, 4, 255-279.
- Friederici, A. D., Steinhauer, K., & Pfeifer, E. (2002). Brain signatures of artificial language processing: Evidence challenging the critical period hypothesis. *Proceedings of the National Academy of Sciences*, 99(1), 529-534.
- Friederici, A. (2005). Neurophysiological markers of early language acquisition: from syllables to sentences. *Trends in cognitive sciences*, 9(10), 481-488.
- Gervain, J., Macagno, F., Cogoi, S., Peña, M., & Mehler, J. (2008). The neonate brain detects speech structure. *Proceedings of the National Academy of Sciences*, 105(37), 14222-14227.
- Gervain, J., Mehler, J., Werker, J. F., Nelson, C. A., Csibra, G., Lloyd-Fox, S., ... & Aslin, R. N. (2011). Near-infrared spectroscopy: a report from the McDonnell infant methodology consortium. *Developmental cognitive neuroscience*, 1(1), 22-46.
- Goldsmith, J. (2011). The Syllable. *The handbook of phonological theory*, 164.
- Gomez, D., Berent, I., Benavides-Varela, S., Bion, R., Cattarossi, L., Nespor, M. & Mehler, J. (2014). Language universals at birth. *Proceedings of the National Academy of Sciences*, 111(16), 5837-5841.
- Gennari, G., Marti, S., Palu, M., Fló, A., & Dehaene-Lambertz, G. (2021). Orthogonal neural codes for speech in the infant brain. *Proceedings of the National Academy of Sciences*, 118(31), e2020410118.
- Giurfa, M. (2019). An insect's sense of number. *Trends in Cognitive Sciences*, 23(9), 720-722.
- Grainger, J., Dufau, S., Montant, M., Ziegler, J. C., & Fagot, J. (2012). Orthographic processing in baboons (*Papio papio*). *Science*, 336(6078), 245-248.
- Halle, M., & Vergnaud, J. (1980). Three dimensional phonology. *Journal of Linguistic Research*, 1(1), 83-105.
- Hockett, C. (1955). *A manual of phonology*. Chicago, University of Chicago Press.

- Hoeschele, M. & Fitch, T. (2016). Phonological perception by birds: budgerigars can perceive lexical stress. *Animal cognition*, 19(3), 643-654.
- Hunter, M. A., & Ames, E. W. (1988). A multifactor model of infant preferences for novel and familiar stimuli. *Advances in infancy research*.
- Hutson, K. A., & Masterton, R. B. (1986). The sensory contribution of a single vibrissa's cortical barrel. *Journal of neurophysiology*, 56(4), 1196-1223.
- Jusczyk, P. & Derrah, C. (1987). Representation of speech sounds by young infants. *Developmental Psychology*, 23(5), 648.
- LaCharité, D., & Paradis, C. (2005). Category preservation and proximity versus phonetic approximation in loanword adaptation. *Linguistic inquiry*, 36(2), 223-258.
- Mann, D. C., Fitch, W., Tu, H. W., & Hoeschele, M. (2021). Universal principles underlying segmental structures in parrot song and human speech. *Scientific reports*, 11(1), 1-13.
- Marler, P., Evans, C. S., & Hauser, M. (1992). Animal signals: motivational, referential, or both. *Nonverbal vocal communication: Comparative and developmental approaches*, 66-86.
- Mazuka, R., Cao, Y., Dupoux, E., & Christophe, A. (2011). The development of a phonological illusion: a cross-linguistic study with Japanese and French infants. *Developmental science*, 14(4), 693-699.
- Mersad, K., Kabdebon, C., & Dehaene-Lambertz, G. (2021). Explicit access to phonetic representations in 3-month-old infants. *Cognition*, 213, 104613.
- Milne, A., Wilson, B. & Christiansen, M. H. (2018). Structured sequence learning across sensory modalities in humans and nonhuman primates. *Current Opinion in Behavioral Sciences*, 21, 39-48.
- Mol, C., Chen, A., Kager, R. & Ter Haar, S. (2017). Prosody in birdsong: A review and perspective. *Neuroscience & Biobehavioral Reviews*, 81, 167-180.
- Olson, R. (2017). Wisconsin Infant Studies Program. Retrieved from <https://bitbucket.org/rholson1/wisp/src/default/>.
- Parker, S. (2008). Sound level protrusions as physical correlates of sonority. *Journal of phonetics*, 36(1), 55-90.
- Patel, A. D., Iversen, J. R., Bregman, M. R., & Schulz, I. (2009). Studying synchronization to a musical beat in nonhuman animals. *Annals of the New York Academy of Sciences*, 1169(1), 459-469.
- Pepperberg, I. M. (2009). *The Alex studies: cognitive and communicative abilities of grey parrots*. Harvard University Press.
- Petkov, C. I., & Jarvis, E. D. (2012). Birds, primates, and spoken language origins: behavioral phenotypes and neurobiological substrates. *Frontiers in evolutionary neuroscience*, 4, 12.
- Petkov, C. I., & Ten Cate, C. (2020). Structured sequence learning: Animal abilities, cognitive operations, and language evolution. *Topics in Cognitive Science*, 12(3), 828-842.
- Poeppel, D. (2003). The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. *Speech communication*, 41(1), 245-255.
- Prince, V. Smolensky, 1993/2004. *Optimality Theory: Constraint Interaction in Generative Grammar*.
- Pulgram, E. (1970). Syllable, Word, Nexus, Cursus. The Hague, Paris: Mouton.

- RStudio Team. (2019). *RStudio: Integrated development environment for R*. RStudio, Inc. <http://www.rstudio.com/>
- Ramus, F., Hauser, M., Miller, C., Morris, D. & Mehler, J. (2000). Language discrimination by human newborns and by cotton-top tamarin monkeys. *Science*, 288(5464), 349-351.
- Räsänen, O., Doyle, G., & Frank, M. C. (2018). Pre-linguistic segmentation of speech into syllable-like units. *Cognition*, 171, 130-150.
- Rantala, J., Salminen, K., Raisamo, R., & Surakka, V. (2013). Touch gestures in communicating emotional intention via vibrotactile stimulation. *International Journal of Human-Computer Studies*, 71(6), 679-690.
- Ridouane, R. (2008). Syllables without vowels: Phonetic and phonological evidence from Tashlhiyt Berber. *Phonology*, 25(2), 321-359.
- Risueno-Segovia, C. & Hage, S. R. (2020). Theta synchronization of phonatory and articulatory systems in marmoset monkey vocal production. *Current Biology*, 30(21), 4276-4283.
- Ruzzoli, M., & Soto-Faraco, S. (2017). Modality-switching in the Simon task: The clash of reference frames. *Journal of Experimental Psychology: General*, 146(10), 1478.
- Saffran, J. & Kirkham, N. (2018). Infant statistical learning. *Annual review of psychology*, 69, 181.
- Santolin, C., Rosa-Salva, O., Vallortigara, G., & Regolin, L. (2016). Unsupervised statistical learning in newly hatched chicks. *Current Biology*, 26(23), R1218-R1220.
- Santolin, C. & Saffran, J. (2018). Constraints on statistical learning across species. *Trends in Cognitive Sciences*, 22(1), 52-63.
- Santolin, C., Russo, S., Calignano, G., Saffran, J. R., & Valenza, E. (2019). The role of prosody in infants' preference for speech: A comparison between speech and birdsong. *Infancy*, 24(5), 827-833.
- Santolin, C., & Saffran, J. R. (2019). Non-Linguistic Grammar Learning by 12-Month-Old Infants: Evidence for Constraints on Learning. *Journal of Cognition and Development*, 20(3), 433-441.
- Santolin, C., Rosa-Salva, O., Lemaire, B. S., Regolin, L., & Vallortigara, G. (2020). Statistical learning in domestic chicks is modulated by strain and sex. *Scientific Reports*, 10(1), 1-8.
- Santolin, C., Garcia-Castro, G., Zettersten, M., Sebastian-Galles, N., & Saffran, J. R. (2021). Experience with research paradigms relates to infants' direction of preference. *Infancy*, 26(1), 39-46.
- Santolin, C., Toro, J. M., & Sebastian-Galles, N. (2021). Abstract processing of syllabic structure in early infancy.
- Scarf, D., Boy, K., Uber Reinert, A., Devine, J., Güntürkün, O., & Colombo, M. (2016). Orthographic processing in pigeons (*Columba livia*). *Proceedings of the National Academy of Sciences*, 113(40), 11272-11276.
- Selkirk, E. (1984). On the major class features and syllable theory. In M. Aronoff & R. Oehrle (Eds.), *Language sound structure* (pp. 107-136). Cambridge Mass.: MIT Press.
- Spierings, M. & ten Cate, C. (2014). Zebra finches are sensitive to prosodic features of human speech. *Proceedings of the Royal Society B: Biological Sciences*, 281(1787), 20140480.
- Thiessen, E. D., & Saffran, J. R. (2003). When cues collide: use of stress and statistical cues to word boundaries by 7-to 9-month-old infants. *Developmental psychology*, 39(4), 706.
- Tierney, A., Russo, F. & Patel, A. (2011). The motor origins of human and avian song structure. *Proceedings of the National Academy of Sciences*, 108(37), 15510-15515.

- Toro, J. M., Trobalon, J. B., & Sebastián-Gallés, N. (2003). The use of prosodic cues in language discrimination tasks by rats. *Animal Cognition*, 6(2), 131-136.
- Toro, J. M. (2016). Something old, something new: Combining mechanisms during language acquisition. *Current Directions in Psychological Science*, 25(2), 130-134.
- Toro, J. M. & Hoeschele, M. (2017). Generalizing prosodic patterns by a non-vocal learning mammal. *Animal Cognition*, 20(2), 179-185.
- Trestman, M. (2015). Clever Hans, Alex the parrot, and Kanzi: what can exceptional animal learning teach us about human cognitive evolution?. *Biological Theory*, 10(1), 86-99.
- Vallortigara, G. (2012). Core knowledge of object, number, and geometry: A comparative and neural approach. *Cognitive neuropsychology*, 29(1-2), 213-236.
- Werker, J. (2018). Perceptual beginnings to language acquisition. *Applied Psycholinguistics*, 39(4), 703-728.
- Weir, A. A., Chappell, J., & Kacelnik, A. (2002). Shaping of hooks in New Caledonian crows. *Science*, 297(5583), 981-981.
- Yip, M. (2006). The search for phonology in other species. *Trends in cognitive sciences*, 10(10), 442-446.
- Zacharaki, K., & Sebastian-Galles, N. (2022). Before perceptual narrowing: The emergence of the native sounds of language. *Infancy*.
- Zhang, F., Monzon, J. C., Zacharaki, K., Eken, A., Santolin, C., Pagliazzi, M., Colomer, M., Sebastián-Gallés, N., & Durduran, T. (2022). Combined functional diffuse correlation spectroscopy and electroencephalography as a new tool in developmental neuroscience. fNIRS 2022 International Conference (Boston, Massachusetts, USA)
- Zhang, F., Monzon, J. C., Zacharaki, K., Eken, A., Santolin, C., Pagliazzi, M., Colomer, M., Sebastián-Gallés, N., & Durduran, T. *in prep.*
- Zhao, X., & Berent, I. (2018). The basis of the syllable hierarchy: Articulatory pressures or universal phonological constraints?. *Journal of psycholinguistic research*, 47(1), 29-64.
- Zuberbühler, K. (2000). Referential labelling in Diana monkeys. *Animal behaviour*, 59(5), 917-927.
- Zuberbühler, K. (2001). Predator-specific alarm calls in Campbell's monkeys, *Cercopithecus campbelli*. *Behavioral Ecology and Sociobiology*, 50(5), 414-422.