Body to Speech and Back: Considerations on Embodied Pronunciation

Josefina Carrera-Sabaté
Universitat de Barcelona, Spain, jcarrera@ub.edu

Abstract
While neuroscience research has increasingly demonstrated the effectiveness of embodied cognition, the teaching of pronunciation lags far behind. In this paper I connect three aspects of embodied cognition with the learning of pronunciation: 1) the use of senses of the body to improve pronunciation; 2) the physical properties of the body and posture as a basis for enhancing pronunciation; 3) emotion and the use of slow respiration and meditation to overcome anxiety when pronouncing.

Keywords: embodied pronunciation, senses, body posture, meditation, slow respiration.

1. Introduction
Following the Embodied Cognitive Theory, there is a strong connection between physiological experience and psychological states. Thus, the process of cognition is not an exclusively psychological process, but a process of understanding the external environment both psychologically and physiologically [1]. This paper deals with embodied cognition through three different axes that serve as a basis for connecting the results of neuroscience research to the teaching of pronunciation: 1) cognition, body, and senses, 2) cognition and physical properties of the body, 3) cognition, emotion, body, and meditation and respiration.

2. Cognition, body, and senses
Neuroscience has established two different types of senses (see, for example, [2]): exteroceptive senses (those that come from sight, sound, touch and taste and smell) and interoceptive senses (those that come from proprioception –signals from the skin and the musculoskeletal apparatus–, and visceroreception –signals from the inner organs, mainly generated by four systems: cardiovascular, respiratory, gastrointestinal and urogenital). In the literature, interoception includes visceroreception.

We know that our whole body is involved in the way we think, feel, and process information. We also know that the brain prioritizes interoceptive information (first visceroreceptive and then proprioceptive information) over exteroceptive information. Moreover, several studies have demonstrated that interoception plays a significant role in implicit memory processes and intuitive decision making [3], [4]. Furthermore, it seems that higher interoception is an indicator of greater attention allocated to both internally and externally relevant events, as well as self-focused attention [5]. Also, high interoception has benefits in memory for emotional material [6]. Following the Embodied Cognitive Theory, the brain captures information across the body’s senses and integrates all the representations, which are then stored in memory. Hence, when a person thinks about an experience or an idea, the brain re-enacts all the perceptual motor and introspective states that were stored during the time the body and the mind interacted with the physical world [7].

In parallel, new research has appeared on how to better understand the way illnesses and bodily states in general influence subjective well-being. [8] explained that each feeling state is likely to be characterized by a blend of interoceptive and exteroceptive inputs. They have mapped the organization of human conscious feelings, which are, at the same time, categorical, emotional, and embodied. They have clustered them into five major categories: positive emotions, negative emotions, cognitive functions, somatic states, and illnesses. (https://www.youtube.com/watch?v=PwU-pwgMw).

Furthermore, we know that the brain has a representation of the body in its somatosensory cortex (see Figure 1), and not all the parts of the body are represented in the same way or to the same degree. Some parts have more neuronal resources devoted to them by the brain. This difference is diagrammatically represented by drawing some parts larger than others.

Thus, we observe that the face (especially the mouth and the tongue), and the hands are larger than the rest of the body (see Figure 2 and [9]). All the neurons corresponding to the areas of the face and hands allow the brain to know what position they are...
adopting. The brain knows in real time how a person’s body is, and the sense that allows the brain to be aware of this is proprioception.

![Representative of the body in its somatosensory cortex](https://www.alamy.es/imagenes/motor-homunculus.html)

**Figure 1: Representation of the body in its somatosensory cortex**

... and to design motor movements in relation to these points of tension. On the other hand, it is much easier to locate and define the positions of the tongue in relation to the palate or the teeth because the direction and position of the articulatory organs are difficult to pin down, and it is almost impossible to influence them directly so that they can produce appropriate sounds, an appropriate rhythm, and an appropriate intonation. By contrast, speech is the result of structured movements of the body understood as a whole. Each sound/syllable has its own points of concentration of tension, and for every syllable there exist specific movements that allow these points of tension to be reached. On the other hand, it is much easier to locate and define the points of concentration of tension of each sound and syllable in different parts of the body overall, and to design motor movements in relation to these points of concentration of tension. All the body movements that help in the production of articulations are considered macro-motricity, since they are produced by the macro-movements of the arms-hands and the legs-feet. These macro-movements correlate with the levels of tension in the articulatory organs, which are referred to as micro-motricity. Thanks to macro-motricity, one can successfully influence the micro-motricity of the articulatory organs and achieve an improved production of sounds ([12] and https://mvmt.uoh.univ-tlse2.fr/videos/mp4/275_procedures_de_correction_des_V_et_C.mp4).

The teachers and students’ training is body-centered, and while the strong points in the verbotonal method are proprioception and prosody, the haptic method pivots on touch and its impact on a person’s kinesthesia.

The theoretical bases of both methods are connected to voice and movement. Verbotonal proposals came after the Laban Movement Analysis (LMA). Following Laban [14], four components of movements are used: strength, duration, space, and development [15]. The body movements include fine muscle movements (especially in the face) that can express the affectivity and the meaning values (intellectual, social, or affective). On the other hand, haptics follows Lessac [16] in feeling the body energies in the production of sounds and being aware of the vibrations and sensations of sounds to explore and project the voice in different contexts [13].

### 2.1. Using the body to improve pronunciation

A complete view of pronunciation must consider the motility of each part of the body that influences speech, as well as the use of the body as an essential element in instruction. In addition, employing an embodied approach to pronunciation in a systematic manner has exciting potential in teaching and learning for fostering clear speech, fluency, and comprehensibility in communication [10].

An understanding of the human body as an instrument of speech and the employment of embodied pronunciation techniques can help language learners develop both perceptive and productive skills in auditory, tactile, kinesthetic, and proprioceptive senses. In the same vein, there is evidence of the benefits of embodied learning in education, especially considering hand gestures [11].

I present a review of two approaches that take embodied pronunciation methodologies into account: the verbotonal method and the haptic approach. As the goal of both approaches is to foster habitual sensory-motor automatization of speech, metalinguistic reflection is not necessary for the learners. It is indispensable, however, for the teacher [12]. Likewise, formal phonological rules are not taught [13].

![Homunculus of Penfield](https://neuromas.weebly.com/blog1/homunculus.html)

**Figure 2: Homunculus of Penfield**

... the strong points in the verbotonal method are proprioception and prosody, the haptic method pivots on touch and its impact on a person's kinesthesia.

The theoretical bases of both methods are connected to voice and movement. Verbotonal proposals came after the Laban Movement Analysis (LMA). Following Laban [14], four components of movements are used: strength, duration, space, and development [15]. The body movements include fine muscle movements (especially in the face) that can express the affectivity and the meaning values (intellectual, social, or affective). On the other hand, haptics follows Lessac [16] in feeling the body energies in the production of sounds and being aware of the vibrations and sensations of sounds to explore and project the voice in different contexts [13].

#### 2.1.1. The verbotonal method

Initially, the verbotonal method was applied to auditory pathologies. Later, taking the concept of phonological deafness [17] as a basis, it was adapted to the teaching-learning of foreign languages. See [18], among others, for more information.

According to Guberina ([19]: 178-179), articulation is not a simple succession of fixed positions of the tongue in relation to the palate or the teeth because the direction and position of the articulatory organs are difficult to pin down, and it is almost impossible to influence them directly so that they can produce appropriate sounds, an appropriate rhythm, and an appropriate intonation. By contrast, speech is the result of structured movements of the body understood as a whole. Each sound/syllable has its own points of concentration of tension, and for every syllable there exist specific movements that allow these points of tension to be reached. On the other hand, it is much easier to locate and define the points of concentration of tension of each sound and syllable in different parts of the body overall, and to design motor movements in relation to these points of concentration of tension. All the body movements that help in the production of articulations are considered macro-motricity, since they are produced by the macro-movements of the arms-hands and the legs-feet. These macro-movements correlate with the levels of tension in the articulatory organs, which are referred to as micro-motricity. Thanks to macro-motricity, one can successfully influence the micro-motricity of the articulatory organs and achieve an improved production of sounds ([12] and https://mvmt.uoh.univ-tlse2.fr/videos/mp4/275_procedures_de_correction_des_V_et_C.mp4).
Moreover, according to the verbotonal method, hearing and listening are possible thanks to proprioception and the sense of the body as they are intimately linked to proprioception, make external rhythmical-motor activity possible and they control it. In order to acquire the rhythm of a foreign language, the involvement of the whole body is proposed, through rhythmical body movements and the use of different musical compositions. Verbotonal experiences have determined that improvements in pronunciation are more quickly achieved and last longer if, at the same time, recourse is made to the rhythm of body movements. This is when the brain-body-brain cycle is at its most effective [20].

2.1.2. Haptic Phonetics and Pronunciation Instruction (HaPT)

In the field of pronunciation instruction, Haptic Phonetics integrates movement (with hand gestures), sight, sound, and touch synchronically to improve L2 pronunciation teaching. Hence, across the visual field, hands can be used to produce haptic gestures corresponding with syllables, stress, and rhythm. Moreover, hands can trace prosodic contours, illustrating pitch and intonation patterns, and they can also be used to facilitate the articulation of L2 vowels and consonants. Furthermore, hands touch each other, or one hand makes contact with a point on the body coinciding with lexical stress or discourse prominence [21]. Overall, using auditory, visual, tactile, and kinesthetic techniques, teachers can amplify students’ cognitive and perceptive sensitivities to consonants, vowels, and prosodic features. Recently, haptic pronunciation has been complemented with language chunks to facilitate pragmatic uses in the classroom in order to get students to respond more appropriately in social contexts [22].

There are around 24 haptic pedagogical movement patterns, which involve the hands moving across the visual field in several locations. Among those patterns, we find, for instance: 1) Touchinami, a systematic gesture that combines movement and touch to experience an intonational contour: https://vimeo.com/61197656; 2) Syllable Butterfly PMP to increase learners’ memory for phrases and longer words, and the ability to memorize texts, fluency in dialogues, and the intelligibility of spoken word grouping: https://vimeo.com/61190793; 3) Rhythm Fight Club, with controlled, boxing-like gestures to assist students in perceiving and producing word stress accurately when learning vocabulary.

These proposals on body movement link with the areas of the body that present a greater sensitivity, as was determined by [9], see Figures 1 and 2. Lastly, as far as embodied pronunciation is concerned, the role of movement in learning is important because it reduces stress and increases cognitive function, and enhances learning and memorizing [23]. Pronunciation understood as a proprioceptive/kinesthetic activity facilitates learning as it is known that human beings have a remarkable ability to remember the positions of their limbs quite accurately and for extended periods [24].

2.2. Effects of embodied pronunciation on second language acquisition

We have just given a general description of two of the most significant methodologies that consider the intervention of the body to be indispensable in improving second-language pronunciation. However, there has been little classroom experimental research devoted to analyzing non-verbal gesturing in the teaching of pronunciation features (See [25] for a state of the art). Some studies on gesturing in the teaching of pronunciation features show that hand gestures improve L2 pronunciation [26], and they seem to be more useful when they mimic phonetic features [27]. Embodied prosodic training improves non-native pronunciation of vowels [28]. Short training sessions with clapping highlighting the prosodic structure of words help improve pronunciation in a foreign language [29].

Visuospatial hand gestures increase voice onset time (VOT) values of aspirated plosives and facilitate an improvement in the overall pronunciation of the target words. Also, gestural training helped maintain word recognition accuracy after three days, while non-gestural training did not [30].

Furthermore, producing gestures helps the improvement of accentedness, comprehensibility, and fluency [31]. Despite these contributions, there are also studies that conclude that the use of gestures does not have a positive effect on the improvement of segmental phonology [32].

Finally, it is important to emphasize that many more studies are needed, especially considering that L2 teachers frequently use gestures to teach pronunciation (see e.g., [33], [34]) without substantial empirical support for this practice.

Therefore, future research should continue to explore the impact of gestures in classrooms: on the one hand, to validate the advantages of the verbotonal and haptic methods, and on the other hand, to collect experiences which use gestures as powerful instructional and learning tools.
3. Cognition, body, and senses

It has been stated that when people interact with their physical environment, their perceptions and actions are constrained by the attributes of their body (e.g., [35], [36]). Extending the embodied cognition framework, the body specificity hypothesis [37] indicates that each person’s body interacts with the world in a specific, unique way, and we can say that people with different body types are likely to think differently. In addition, body position interacts with cognition, suggesting that humans use their bodily states to interpret experiences. For instance, unconsciously smiling or frowning can influence how funny a cartoon seems [38], and slouching (maintaining a slumped posture) can generate feelings of impotence [39]. Furthermore, we remember words in a positive way when we are in an upright posture [40]. In addition, in stressful tasks, [41] found that adopting an upright sitting posture in the face of stress can maintain self-esteem, reduce negative mood, and increase positive mood compared to having a slumped posture. Moreover, more positive emotions are achieved when an expansive (vs. contractive) posture is adopted for two minutes [42].

The research is consistent with embodied cognition theories that assert that muscular and autonomic states influence emotional responding. In this sense, it has been observed that posture influences the nervous system. For this reason, taking care of one’s posture has benefits for one’s mind as well for one’s muscles. From a theoretical point of view, adopting a posture of power should increase one’s feeling of power through bottom-up processing ([43]); see [44] for more evidences).

3.1. Cognition and physical properties of the body – What interest does posture hold for improving pronunciation?

We have just seen that posture has a clear effect on cognitive and emotional aspects. From another perspective, posture is also fundamental in education, drama, and speech rehabilitation. Moreover, we know that imitation of posture and movements can smooth over interactions between interlocutors [45], and in conversing with one another there is an increased level of shared postural activity [46]. Posture and gestures do not only translate or express attitudes: they also create, manifest, and maintain them. Posture represents a continuously maintained attitude and marks, to some extent, a predisposition to act in a certain way. Postural awareness is fundamental in any educational exercise [47].

The implications of posture in the learning of pronunciation are clear: if, for example, one wants to produce a consonant that is less tense than a given sound (like an approximant), the interventions for improving pronunciation will have to begin with a balanced posture of the learner’s and teacher’s bodies, in which, for example, there is no tension in the neck or knees, which would hinder modifications of articulatory gestures.

When it comes to proposing exercises for improving pronunciation, however, there are few proposals that take a balanced posture of the body into account before starting (see, however, [48], [49]) and even fewer studies that analyse the impact of body postures in pronunciation.

Leaving that aside, and in order to find something of practical and ecological value for language teachers, I wonder if, for example, experiments could be carried out on pronunciations made in conjunction with different body postures, since, most likely, pronunciation will not be the same if learners are standing as opposed to sitting in chairs.

4. Cognition, emotion, body, meditation, and respiration

The theory of emotion [50] links somatic and viscerally-affecting feedback to subjective emotional experience (feelings). James stated that an emotive stimulus automatically initiates visceral, vascular, or somatic reactions (e.g., changes in blood pressure or heart rate), and it is the perception of these bodily reactions that crucially constitutes the emotional component of experience. Thus, interoception can shape our experience of emotions. As James said, I don’t cry because I feel sad, I am sad because I perceive that I cry. Refinements of this model, the somatic marker hypothesis, headed by Damasio [51], state the close relationship between physiological, emotional, and cognitive processes. Thus, in line with Damasio's somatic marker hypothesis, it can be argued that the initial perception of an emotional stimulus leads to bodily responses that are partially recorded and stored in the brain. Later, when cognitive processes require information about this emotional stimulus, the associated pattern of neural states is partially reactivated, stressing the link between modality-specific and conceptual systems [52].

On the other hand, Eastern traditions have been built upon an osmotic interaction between the body and emotions. Ancient mind-body practices bring us back to the body and allow us to orient our minds to observe emotional states while we experience them without being dominated by them. Several studies have demonstrated the positive effects of meditation and yoga on different aspects of cognition, such as
attention, memory, and executive function (see [53], [54]). It has also been shown that meditation has direct effects on breathing, since people who meditate have a slower rate of respiration than those who do not (see [55]).

Bearing respiration in mind, it must be pointed out that up until a few years ago neuroscience had not spoken of the influence of breathing on the brain. Now, it has been discovered that respiration allows us to mold our mental states, because our way of breathing influences our memory, attention, motivation, and responses to emotional states. It has been observed [56] that breathing through the nose modulates the brain stem (concretely, the piriform cortex, which is related with the detection of odors and with memory) and that every time we breathe, we change the rhythm of this piriform cortex.

Apart from the piriform cortex, it is known that there are two other structures that are directly influenced by respiration: the amygdala (responsible for the processing of emotions and the emotional content of our memories) and the hippocampus (responsible for the processing of memory) and it seems that every time we breathe through our nose, the amygdala and hippocampus change significantly. Moreover, we have a greater chance of memorizing and remembering when we inhale than when we exhale, and it seems that inhaling as deeply as possible is highly recommendable in learning processes [56].

4.1. Can improvement in pronunciation be fostered by proposals involving slow respiration and meditation?

As we have just seen, neuroscience has explored the impact of breathing on different aspects of life, especially to enhance psychological well-being, given the high prevalence of psychiatric disorders in society, and, specifically, among university students [57]. Moreover, both in second-language learning generally and in second-language pronunciation particularly, anxiety has been reported to be a very negative factor (see [58]). We also know that there is a positive correlation between respiratory rate and anxiety level [59], and that levels of individual anxiety especially affect expiratory time [60]. Apart from that, when it comes to speaking and reading in public, one of the main challenges is to manage breathing and stress, which can also reach the level of anxiety. Furthermore, it has been seen that controlled breathing exercises can be effective in inducing certain emotional states [61], and that muscle relaxation exercises can reduce anxiety (see [62] for a meta-analysis). These results suggest that changing bodily processes can influence emotion regulation.

On the other hand, yoga nidra is a profound meditation practice spanning thousands of years that is defined, in principle, as a method of physical, emotional, and mental relaxation and a tool for developing full attention. Through this practice, the impact of the senses is gradually disconnected from the mind and the only sense left open is that of hearing; the mind is focused on one’s inner experience and on the sensation resulting from placing attention on the different parts of the body that the facilitator enumerates. The parts of the body that attention is focused on correspond to the areas of the brain that have the highest neuronal load, as was stated by [9], see Figures 1 and 2. So, in the practice of yoga nidra, the cerebral cortex does not receive information from the motor organs, and the conscious mind passes on to a second plane, and, since there is no selection or judgement of the information that is being received, the mind is much more receptive (see [63]).

4.1.1. Experiment 1: meditation, speaking rate and silent pauses

To find out if the changes in breathing described after meditative practices can have an impact on reading aloud and on pronunciation, an experiment was designed in relation to the practice of yoga nidra. The corpus that was analyzed consists of the data obtained from two aloud readings of the text North wind and the sun by 19 university students (in Catalan, the text is 104 words long).

The text was read on the same day in two different contexts: after explanations by the teacher and after a practice of yoga nidra. The readers recorded themselves using their mobile devices. The total duration of the 38 recordings of the story was 22 minutes. The readers were university-level, Catalan-speaking readers from different dialect areas of Catalan who had no auditory or pronunciation problems and were between the ages of 21 and 24. The words were analyzed with Praat ([64]) and a modified script ([65]). See [66].

After carrying out a Poisson linear mixed model comparison of both readings, there were significant differences in: 1) the average total duration of the texts: the post-yoga readings were significantly longer (38.3 seconds) than the pre-yoga ones (33.7 seconds); 2) the number and duration of pauses: in the post-yoga reading, more pauses were produced (a total of 441 pauses) than in the pre-yoga one (412 were counted). Of the 19 readers, 13 increased their number of pauses between one reading and the other. In addition, the overall length of the silent pauses also was different between the two readings.
4.1.2. Experiment 2: meditation and vowels

In a second analysis of both readings (pre-yoga and post-yoga), pronunciation of stressed vowels among 16 women from the same dialect area was analysed. After the linear mixed models analysis, it was also found that the duration of vowels was significantly longer in post-yoga reading than in pre-yoga reading (see Figure 3). Similarly, the intensity of vowels was lower in post-yoga reading than in pre-yoga reading (see Figure 4).

![Figure 3. Vowel duration in pre-yoga (T1) and post-yoga (T2) (d = 0.254, p = 0.018)](image)

![Figure 4. Vowel intensity in pre-yoga (T1) and post-yoga (T2) (t = 0.726, p = 0.021)](image)

As for the vowel characterization, there was not significant difference between both readings. Nevertheless, it was observed that in post-yoga reading, F0 was significantly lower than in pre-yoga reading (see Figure 5). Furthermore, preliminary results show a tendency to have lower F1 in post-yoga reading. This suggests a slight tendency to produce closer vowels after the practice of yoga (see Figure 6).

![Figure 5. F0 in vowels in pre-yoga (T1) and post-yoga (T2) χ²(1) = 22.700, p < 0.001](image)

4.1.3. Discussion and conclusions: experiment 1 and experiment 2

The practice of yoga nidra among university students caused changes in the segmenting of the story, and in the duration of the pauses. These changes can be explained by the physiological and respiratory changes that meditation and full attention bring about ([55]). And, given the correlation existing between anxiety and breathing, the practice of yoga nidra could be considered a useful tool for reducing the anxiety of second-language students.

Preliminary results obtained after the analysis of vowels allow us to consider that there is a physiological impact of the practice of yoga nidra on vowel production, because F0 was significantly lower in post-yoga. The pitch of vowels (F0) from the post-test might be the result of a different biomechanical force involved in vowel articulation between pre-yoga and post-yoga (this reminds us of the correlation observed between intrinsic F0 and the position of the tongue or the tongue and the jaw – [67]). As F0 of vowels is significantly modified by yoga nidra practice, which are the articulatory mechanisms that yoga seems to modify? Furthermore, can yoga training definitely impact pronunciation? We need more evidences. (In the same vein, there has been experience in modifying pronunciation when people were starting to sleep. See [68]).

These exploratory results leave the door open to future research to discover new pathways for improving pronunciation and managing anxiety.

5. Conclusions

While neuroscience research has increasingly demonstrated the effectiveness of embodied cognition, teaching pronunciation is still far away from this. In this paper I connected three aspects of embodied cognition with the teaching and learning of pronunciation in a second language to be further explored in the future: 1) senses of the body to improve pronunciation, 2) the physical properties of the body and posture as a basis to enhance
pronunciation, and 3) emotion and slow respiration and meditation to alleviate anxiety and modify pronunciation.

6. Acknowledgements
I would like to thank James McCullough for discussion. This research was supported by 2019PID-UB/030, and was part of PID2020-113971GB-C21 project, supported by MCIN/AEI/10.13039/501100011033.

7. References