UNIVERSIDADE FEDERAL DO PARANÁ
Departamento de Letras Estrangeiras Modernas

ISSN: 1980-0614

REVISTA X

SPEECH PERCEPTION AND PRODUCTION AS CONSTRUCTS OF ACTION: IMPLICATIONS FOR MODELS OF L2 DEVELOPMENT

Percepção e Produção da Fala como Construtos de Ação: Implicações para Modelos de Desenvolvimento de L2

Reiner Vinicius PEROZZO
Universidade Federal do Rio Grande do Sul
reiner.vinicius@ufrgs.br
https://orcid.org/0000-0002-7778-9690

Felipe Flores KUPSKE
Universidade Federal da Bahia
kupske@ufba.br
https://orcid.org/0000-0002-0616-612X

ABSTRACT: Speech production involves an intricate set of actions. Its underlying cognitive mechanisms are thus historically seen as distant from those of speech perception, usually assumed to be a passive process. However, dynamic perspectives on language congregate grammar and language use, approximate phonetics and phonology, and value the role of speech perception in language development. Recent studies argue that speech production and perception are overlaying or at least highly interacting. Some scholars claim that the link between these two processes surpasses the acoustics, as studies have revealed that action also has a role in language comprehension. Phonic gestures are not just mechanisms by means of which one experiences speech production, but are supporting to perception. In this perspective, models interested in L2 development face a twofold challenge: to amalgamate speech perception and production, and to consider that speech transcends the acoustics, since - in a dynamic frame of reference - phonetic-phonological representations are auditory, gestural and general. This paper aims at presenting evidence for a gesture-driven perspective to L2 speech development in which the gesture is a phonological primitive that pervades and connects speech perception and production. By emphasizing a gesture-driven point of view, this work presents congruent and incongruent tenets among some hegemonic models of L2 speech development and an ecological/dynamic account.
KEYWORDS: Speech production; Speech perception; Nonnative speech; Gesture-driven speech development.

UNIVERSIDADE FEDERAL DO PARANÁ
Departamento de Letras Estrangeiras Modernas

ISSN: 1980-0614

REVISTA X

RESUMO: A produção da fala envolve um conjunto intrincado de ações. Seus mecanismos cognitivos subjacentes são, portanto, historicamente vistos como distantes daqueles da percepção da fala, geralmente tomada como um processo passivo. No entanto, perspectivas dinâmicas para a língua(gem) congregam a gramática e o uso linguístico, aproximam a fonética e a fonologia e valorizam o papel da percepção da fala no desenvolvimento linguístico. Estudos recentes argumentam que a produção e a percepção da fala são sobrepostas ou, pelo menos, altamente interativas. Alguns estudiosos afirmam que a conexão entre esses dois processos extrapola a acústica, pois estudos revelaram que a ação também tem um papel na compreensão da linguagem. Os gestos fônicos não são apenas mecanismos por meio dos quais se experimenta a produção da fala, mas dão suporte à percepção. Nessa perspectiva, modelos interessados no desenvolvimento de L2 enfrentam um duplo desafio: amalgamar a percepção e a produção da fala e considerar que a fala transcende a acústica, uma vez que - em um quadro dinâmico - representações fonético-fonológicas são auditivas, gestuais e gerais. Este artigo tem como objetivo apresentar evidências para uma perspectiva dirigida pelo gesto para o desenvolvimento da fala em L2, em que o gesto é um primitivo fonológico que permeia e conecta a percepção e a produção da fala. Ao dar destaque a um horizonte gestual, este trabalho apresenta princípios congruentes e incongruentes entre alguns modelos hegemônicos para o desenvolvimento da fala em L2 e uma perspectiva ecológica/dinâmica.
PALAVRAS-CHAVE: Produção da fala; Percepção da Fala; Fala não nativa; Desenvolvimento da fala orientado por gestos.

INTRODUCTION

Speech production is a complex sensorimotor activity, as it involves a highly intricate set of actions regarding the coordination of various parts of the human body. This is one of the reasons why the cognitive mechanisms underlying speech production are historically seen as completely different or separated from those supporting speech perception (MCGETTIGAN; TREMBLAY, 2018), often considered a passive process in communication. Dynamic perspectives to language, such as the ones proposed by Albano (2001, 2020) and Beckner et al. (2009), congregate grammar and language use, approximate phonetics and phonology, and, as a natural consequence, value the role of speech perception in language development, variation and change. Bybee (2001), for instance, establishes that mental representations of languages are affected and driven by experience, and the use of forms and patterns, both in production and perception, will impact their storage in memory.

ISSN: 1980-0614

To Beckner et al. (2009), language development entails complex and probabilistic analyses of the language of the environment. For those who take language as a dynamic system, as we do in this work, language development rests on the estimation of patterns of a specific speech community by means of the experiences perceived by our cognitive machinery, psychomotor capacities, as well as by the dynamics of social interaction itself (BECKNER et al., 2009). Grammar thus results from dynamic cycles involving language production and perception (ELLIS, 2008).

As pointed out by Kupske, Perozzo and Alves (2019), more recent perspectives in linguistics have gathered evidence of phonological plasticity. Not even adult grammars are immune to the effects of environmental changes, such as in the case of first language attrition, a situation in which, for example, bilingual immigrants in L2-dominant contexts apply L2 sound patterns to L1 speech production (KUPSKE, 2017, 2019). The role of perception in language development is also evidenced by Evans and Alshangiti (2011). To these authors, in multidialectal scenarios, speakers tend to accommodate their linguistic behavior so that communication is facilitated. Evans and Iverson (2004) affirm this alignment may lead to changes in speech production and perception. In the same light, Pardo (2006) goes further and states that even short-time interactions are able to drive permanent sound changes. Data like these strengthen the argument that speech perception and production are at least complementary. To McGettigan and Tremblay (2018), for instance, speech perception and production are connected from the earliest childhood.

Even though the discussion about the foundations of speech perception and its connection to speech production gained some attention few decades ago, as can be seen from the well-known "debate" between Fowler (1996) and Ohala (1996), recent studies in psycholinguistics and in neurosciences are reviving the discussion and argue that speech production and perception are overlapping or at least highly interacting processes. These areas go further and claim that the link between speech perception and production goes beyond the acoustics. To McGettigan and Tremblay (2018, p. 02), "speaking requires learning to map the relationships between oral movements and the resulting acoustical signal, which demands a close interaction between perceptual and motor systems". A number of studies, immersed in what Albano (2020) names as the pragmatic turn in the study of language and mind, have revealed that action goes beyond the construction of the phonetic-phonological signifiers and has a role in language comprehension. In this

perspective, phonic gestures[1] - the actions in speaking - are not just mechanisms by means of which we experience speech production, but are supporting to perception.

In this perspective, models interested in L2 development - our focus in terms of this theoretical paper - face a twofold challenge: to amalgamate speech perception and production, and to consider that speech development goes beyond the acoustics, since phonetic-phonological representations, in a dynamic frame of reference, are auditory, gestural and general (non-linguistic). On this note, to approximate speech perception and production might be a challenge for gestural models of nonnative language development (e.g., BEST, 1995; BEST; TYLER, 2007) or might be at least easier for theories of language anchored solely in acoustic cues (e.g., FLEGE, 1995; FLEGE; BOHN, 2021), as it seems less intuitive to state that the gestural primitive permeates both speech perception and production. On the other hand, models that exclusively focus on the acoustics of speech might ignore the role coordinated actions (e.g., gestures) play in language development as evidenced in the past decades. A more comprehensive model for L2 speech development would ideally consider (coordinated) action in lead both for speech production and perception or, in other words, would consider articulatory, acoustic, and general dimensions[2]. A dynamic account of L2 speech development must integrate both processes instead of treating them as independent constructs, as well as should encompass the phonological grammar as more than a direct auditory development.

In this conceptual analysis, we claim that speech perception and production are indissociable in terms of L2 speech development. We therefore aim at presenting evidence from recent investigations for integrating phonic gestures in L2 development. In other words, we bring to light a theoretical framework that takes the gesture as the phonological primitive, and its likeliness to pervade and connect both speech perception and

[1] In this paper, we refer to "articulatory gesture" and "phonic gesture" as different constructs for didactic purposes. The former is related to what is proposed by Liberman and Mattingly (1985), Fowler (1986, 1996), Browman and Goldstein (1989, 1992), Best (1995) and Best and Tyler (2007). On the other hand, by using "phonic gesture" (FREITAS, 2012), we refer to the primitive of the phonological grammar as developed by Albano (2001), which includes both articulatory - as do the above mentioned authors - and acoustic information, even though this term only appears in Albano (2020).

[2] Since we draw from a dynamic view of language, we assume that grammar is rich and bound to the circumstances of use. Phonological grammar is thus viewed as the cognitive organization of individuals' experience with their language (BYBEE, 2001) and includes not only the representation of sounds and gestures, but also every other information individuals are able to perceive in their ecological experiences. "This information consists of phonetic detail, including redundant and variable features, the lexical items and constructions used, the meaning, inferences made from this meaning and from the context, and properties of the social, physical and linguistic context" (BYBEE, 2010, p. 14).

production. The present theoretical paper is mainly floated for the purpose of evidencing the shortage of models in a gestural perspective that integrate speech perception and production in L2 research and stimulating specialized feedback. It represents the first step we take in understanding and developing an integrative, gesture-driven, approach to L2 speech development. After presenting an outline of speech perception and production in linguistics, we focus on the topic of speech perception as a process and a product of action. After discussing the role of gestures in speech, we will defend a gesture-driven account to L2 speech development. In the same section, we also highlight congruent and incongruent tenets among some hegemonic models of L2 to speech development and a more ecological/dynamic view towards it.

AN OUTLINE OF SPEECH PERCEPTION AND PRODUCTION

Opposed to the notion that speech perception and production are relatively novel fields of inquiry in linguistics, especially in Brazil, we may trace their roots, to a certain extent, back to the works of linguists such as Henry Sweet, Daniel Jones, and Nicolai Trubestkoy. According to Tatham and Morton (2011), Henry Sweet was innovative enough to develop a transcription system that would foreshadow the system-oriented concept of the phoneme and its allophonic variants. David Jones was more focused on the "linguistic" description of pronunciation, and one of his greatest achievements was to propose the cardinal vowel system in terms of the different positions of the tongue. Jan Baudouin de Courtenay, a Polish linguist, discussed Jones' ideas in a number of writings in the twentieth century, and assumed that the phoneme was the psychic equivalent of the speech sound. Taking Courtenay's definition into account, Trubetskoy (1949) pointed out that:

> This definition was untenable because several sounds of language can correspond, as variants, to the same phoneme, and each of these sounds of language has its own "psychic equivalent" – namely the acoustic and motor representations which correspond to it (TRUBETSKOY, 1949, p. 41. Translated by the authors)[3].

[3] Original text: *Cette définition était insoutenable, car au même phonème peuvent correspondre, comme variantes, plusieurs sons du langage, et chacun de ces sons du langage possède un "équivalent psychique" propre – à savoir les représentations acoustiques et motrices qui lui correspondent* (TRUBETSKOY, 1949, p. 41).

ISSN: 1980-0614

Trubetskoy (1949) brought up the so-called extrinsic allophones, and tried to explain that realizations of the same phoneme could differ widely. For the Russian linguist, a particular speech sound, including an allophone, can only be defined by its relation to the phoneme, considering that if we depart from the speech sound to define the phoneme, we fall into a vicious circle. If there were as many speech sounds as psychic equivalents, all of the former had to be considered phonemes, and phonological operations from a formal perspective would be disregarded. Such a position is taken by Trubetskoy (1949) because he had a system-oriented view of phonology, and, according to him, "the phoneme can be defined satisfactorily neither in terms of its psychological nature nor in terms of its relation to the phonetic variants, but purely and solely with reference to its function in the language" (TRUBETSKOY, 1949, p. 44. Translated by the authors)[4].

Later on, Roman Jakobson expands the notion of phoneme and characterizes it as a bundle of distinctive features, since it can be decomposed into smaller units, which would capture the functional role they play in a language. Together with Gunnar Fant and Morris Halle, Roman Jakobson formulates a featural system that addresses an acoustic characterization of linguistic sounds and aims to provide a minimum amount of distinctive features in order to establish functional contrasts (JAKOBSON; FANT; HALLE, 1952). Conversely, in 1968, Noam Chomsky and Morris Halle propose an articulatory basis to distinctive features, which not only is thought to encompass functional contrasts, but also phonological rules and processes (CALLOU; LEITE, 1994).

The course units of speech took both in a psychological and in a linguistic account has had great impact on the theoretical and methodological development of speech perception and production. Cognitive mechanisms range from fundamental to irrelevant for some researchers, while other scholars understand that articulatory parameters – and not acoustic cues – are taken as phonological primitives. In what follows, we shall outline the recent development of the main trends in speech perception and production.

The first studies that concentrated on speech perception were conducted by researchers who tried to explain such a process by means of its acoustic nature. In this regard, works by Pisoni (1973, 1974), Cole and Scott (1974), and Kuhl and Miller (1975) stand out. They were largely influenced by the structuralist paradigm, through which speech perception was conceived as being fundamentally a matter of hearing, that is, perceiving speech meant listening to speech. As Perozzo (2017) states, the treatment

[4] Original text: *Le phonème ne peut être défini d'une façon satisfaisante, ni par sa nature psychologique, ni par ses rapports avec les variantes phonétiques – mais seulement et uniquement par sa fonction dans la langue* (TRUBETSKOY, 1949, p. 44).

ISSN: 1980-0614

of speech perception in terms of this first school is conceptualized in a psychoacoustic fashion, in which the perceptual primitives are acoustic cues apprehended indirectly through cognitive processing and mental representations. Best (1995) explains that, according to this view, the pieces of information infants initially perceive are general (not only linguistic) and the experience with the target language triggers the formation of prototypes, properties and models.

Nishida (2012) indicates that other inquiries concerning the substrates of speech information that promoted lexical oppositions started to take place so far as research on speech perception evolved. The new agenda brought about a conflict between the widely spread acoustic basis of speech perception and the challenging evidence of its articulatory foundations, which began to impose itself by the development of the Motor Theory of Speech Perception (LIBERMAN; MATTINGLY, 1985). This theory assumes the vocal articulation as responsible for speech perception and mediated by a linguistic apparatus to be perceived (and produced). Different from the psychoacoustic perspective, the motor account of speech perception relates to intended articulatory gestures as primitives – derived from neuromotor commands – whose mental representations are indirectly accessed (BEST, 1995). The theory also postulates that the information infants initially perceive is, in fact, linguistic and the native phonetic input tunes the speech module as experience with the target language increases.

Another school that operates with the articulatory gesture, but with divergent conceptualization, and that represents a third moment in the speech perception research is the Direct Realistic Theory of Speech Perception (FOWLER, 1986, 1996). The articulatory gesture, phonological primitive of Carol Fowler's theory, is seen as a real object, in opposition to an intended articulatory gesture, which alludes to a mental representation. Like Fowler (1986, 1996), Best (1995) understands that the direct realist construct of speech perception is based on distal articulatory gestures directly apprehended by our perceptual systems through the extraction of "affordances" and the active exploratory activity of speech events. Derived from the ecological perspective of perception (GIBSON, 1966, 1986) and linked to a dynamic conception of language, Fowler's theory contends that the information infants first perceive is general, but the experience with the target language presupposes the direct extraction of native gestural invariants, which engenders the knowledge of sound patterns (BEST, 1995).

With regard to speech production, we ought to discuss the Speech Learning Model (FLEGE, 1995), also known as SLM, as it is one of the best-known worldwide and widely present models in Brazil. To Bohn (2020), the SLM is a reaction to the

ISSN: 1980-0614

failures of Contrastive Analysis and the Critical Period Hypothesis to explain L2 speech development, and tries to elucidate nonnative accent. Basically, the SLM states that L2 learners - regardless of age - congregate the same mechanisms and processes applied for L1 development, including the capacity to create new phonetic categories. This ability would rest on three points: (1) the level/state of the previously developed languages; (2) quality and quantity of input in the target language; and (3) the relation among native and nonnative phonetic categories.

In a nutshell, the model considers that an L1 filter is in charge of the perceptual process, which might make sounds of an L2 to be perceived as similar to those of the L1, and to be then categorized this way. In this sense, to Flege (1995), speech production is strongly limited by the speaker's perceptual accuracy. The relation of equivalence between the L1 and L2 sound systems could, for example, hinder learners from developing adequate values for phonetic cue weighing required for the contrast between L1 and L2. These cues can potentially be used to discriminate between two sounds, and indicate whether they belong to the same category or not. If L2 sounds are identified as L1 ones, the formation of new categories of contrasts will be blocked. In other words, put in a simple way, Flege (1995) points out that the greater the perceived difference of an L2 sound compared to a closer L1 sound, the more likely it is that the separation of categories will be established for the L2. According to this model, as the perception of L1 sounds develops during childhood and adolescence, the assimilation of L2 sounds is more likely, as phonetic categories of L1 and L2 coexist in the same phonological space.

SLM was revised (SLM-r) in 2021 (FLEGE; BOHN, 2021). To Bohn (2020), the most important tenets of SLM remain the same in SLM-r, and they are: (1) there is no biologically-based limit to speech learning ability; and (2) learnability is a function of perceived cross-language similarity. The purpose of the new version of the model is to account for how sounds are learned across the lifespan. It therefore focuses on the effects of age and L2 experience, and on the input differences between monolinguals and bilinguals, and between children and adult learners. With this new emphasis, SLM-r investigates L2 development in a more ecological and ontological fashion, instead of focusing on the end state.

Actually, to Flege and Bohn (2021), there is no end state in speech development, no matter L1 or L2. The authors contest the fact that the accuracy of L2 perception would place an upper limit on L2 speech production, as very few correlations were found since 1995. In addition, new data reveal that there is a strong multidirectional link between speech production and perception. Thus, speech perception and production co-evolve

ISSN: 1980-0614

during lifespan. Another important aspect in the newer version of the model is that it focuses on the individual differences in L2 development, instead of focusing solely on the group level.

As we may observe, especially in relation to speech perception, the three most prominent schools attempted to explain such a phenomenon by describing their view both on the processes and on the primitives that could stand for speech. Not only are their epistemological references at play in each contribution they provided, but also their legacy remains. Concerning speech production in a nonnative language, SLM - and its extended version, SLM-r - is one of the most widespread models that can be used for speech perception and production, but still needs to adapt in certain ways in order to account for a gesture-driven approach to L2 development.

PHONOLOGICAL GRAMMAR

Since the objective of this paper is to argue that speech perception and production are indissociable in terms of L2 development, the theoretical framework we bring to light has a twofold approach: locating the phonic gestures in the arena of phonological grammar and discussing the likeliness of the gestural primitive to permeate both speech perception and production.

Perozzo (2019, p. 131) claims that, influenced by the Scientific Revolution, "modern linguistics began to incorporate mechanisms of analysis that reflected proposals from other sciences" (seen as prestigious), and highlighted the principles of reductionism[5], empirical observations, universality and closed linear systems. As pointed out by Berticelli (2010), we must not ignore the advances of classical science, but its successes are harmful, because the experimental dialog that it maintained with nature ended up making man a stranger in the world. This tragic finding is in line with the shedding of subjectivity and experience in modern science, which was constituted against nature, denying its complexity, graduality and flux in the name of a world governed by simple and immutable laws/rules/constraints.

The shedding of the ecology of language is indeed seen in the development of linguistics as an independent science. One of the most relevant notions of linguistic structuralism, for instance, is that a language corresponds to form (its structure), and

[5] The basic premise of analytical reductionism refers to the investigation of an object in its individuality and indicates that large phenomena (in general, complex ones) can be divided and modeled into smaller parts, that is, reduced. Thus, such parts are recombined in order to provide a description of the whole (COLCHESTER, 2016).

not to substance (the matter by which it manifests). Nevertheless, the same approach demands the need to analyze the substance in order to formulate hypotheses about the system that concerns it. This consideration brings up the proposition that a language should be studied on its own terms, implying that all "extralinguistic" variables (gender, age, social class, etc.) are nonessential, since they do not act on the internal relations of its units. This turns out to be evident in structuralist linguistics, which proposes a clear-cut distinction between phonetics and phonology, as Albano (2001) accentuates:

> This insistence on function is in stark contrast to the point of view of the phonetician who, as explained above, must carefully avoid considering the meaning of what is being said (in other words the meaning of the significant). This precludes the classification of phonetics and phonology under the same rubric, although these two sciences seem to be concerned with similar things. To use a striking comparison by R. Jakobson, the relationship between phonology and phonetics is the same between the national economy and the business directory or between financial science and numismatics (TRUBETSKOY, 1949, p. 12. Translated by the authors)[6].

Albano (2001) advocates that the discovery of discrete categories related to speech sounds with distinctive function – represented by the phoneme – could not be threatened by the gradience of the phonetic continua. Allowing linguistic contrasts to be inundated by continuous variations would result in the implosion of the signifier. Consequently, "a signifier based on conceivably infinite distinctions cannot establish differences that are univocally associated with differences in meaning" (ALBANO, 2001, p. 13)[7].

Not only does Albano (2001) refute the split between phonetics and phonology in linguistic structuralism, but also criticizes the way generative grammar tackles both fields. It is worth noting that, even though each period of generative linguistics has its specificities, one of its key properties is that linguistic knowledge is associated exclusively

[6] Original text: *Cette insistance sur la fonction s'oppose d'une manière très tranchée au point de vue du phonéticien qui, comme on l'a expliqué ci-dessus, doit éviter soigneusement de considérer le sens de ce qui est dit (autrement dit le sens du signifiant). Cela empêche de classer la phonétique et la phonologie sous une même rubrique, bien que ces deux sciences s'occupent apparemment de choses semblables. Pour reprendre une comparaison frappante de R. Jakobson, le rapport existant entre la phonologie et la phonétique est le même que celui qui existe entre l'économie nationale et l'annuaire du commerce ou entre la science financière et la numismatique* (TRUBETSKOY, 1949, p. 12).

[7] Original text: *um significante assentado sobre distinções potencialmente infinitas não pode marcar diferenças que se associem univocamente a diferenças de significado* (ALBANO, 2001, p. 13).

with abstraction. Grammatical units are available via Universal Grammar, a centralizing system that generates linguistic forms by means of transformations, derivations and/or the result of a hierarchy of universal constraints. Other relevant points about the theory are the binary/Cartesian aspect that characterizes several models and analyses (either in constituency, distinctive features, or even in the more general taxonomy) and the essentially top-down mechanism in which grammar is organized (PEROZZO, 2019).

Speech scientists such as Browman and Goldstein (1989, 1992) and Albano (2001, 2020) believe that there should not be a division between phonetics and phonology, and this would depend both on the way the phonological grammar is seen and also on the primitive that is associated with it. In this view, phonological systems would be developed from phonetics, in a bottom-up fashion, and the articulatory gestures would, therefore, be the basic primitive for distinctiveness. Browman and Goldstein (1989, p. 69) mention that the articulatory gestures are "units of action that are inherent in the maturation of a developing child and that therefore can be harnessed as elements of a phonological system in the course of development".

Perozzo (2017) emphasizes that, in order to capture what is at stake in the relationship among the phonic elements of languages from a perceptual perspective, we should recognize the advances proposed by Albano (2001) with regard to gestural phonology. The first factor concerns the symmetry with which Albano (2001) relates the abstract knowledge of sound patterns (mental) to its concrete reality (physical). The second aspect resides in the approximation of the phonic gesture to other units that operate in grammar (linguistic knowledge), such as the morpheme. Finally, the third contribution is exemplified by the relevance of acoustic information interwoven in the gestural primitive, which is established in terms of action (articulation) and representation (abstraction).

According to Albano (2020), phonology is not limited to the simultaneous or sequential combination of articulatory gestures and their sensitivity to the linguistic context. Similarly, phonology is the logic that encompasses their joint production and their realization through coordinated movements. In fact, "it is also the logic of their variability according to the situational context, which is sensitive to social and stylistic variables of the most diverse types" (ALBANO, 2020, p. 43. Our Translation)[8]. By questioning the boundaries between phonetics and phonology, the researcher posits that "phonology" should be used in a more general fashion, also covering phonetics and phonostylistics. This is due to the fact that it replaces the idea that there is a universal

[8] Original text: É ainda a lógica da sua variabilidade conforme o contexto situacional, que é sensível a variáveis sociais e estilísticas dos mais diversos tipos (ALBANO, 2020, p. 43).

mechanics (language independent) with the notion that phonetics is part of the language system. The same term would also replace the conception that style is restricted or idiosyncratic with the understanding that it is socially regulated and informs the context (as much as it is informed by it).

If we consider, then, that phonic gestures are the primitives of phonological grammars, either they should be the units of speech perception and production or, at least, relate to both to a large extent. We are fond of the former because it would avoid translation mechanisms from speech perception/production to the phonological knowledge of sound patterns. In this regard, Alves and Silva (2016) add that formal explanations for the description and analysis of sound systems can also be made possible and successful if we consider the articulatory gestures as the common currency between speech perception and production. Their perspective derives from and is supported by Goldstein and Folwer's (2003) assumption that perception and production must be addressed in terms of parity. In other words, as stated by Fowler and Galantucci (2005), listeners must characteristically perceive the language forms that speakers produce for speech to serve its public communication function. As it should be clear, these language forms are the phonic gestures, both concrete and abstract.

SPEECH PERCEPTION AS ACTION

The role of action is undeniably paramount in speech production, as discussed in the previous sections, but it is also important to phonological development and grammar in perspectives to language that approximate phonetics and phonology, grammar and language use, such as the Articulatory Phonology (BROWNMAN; GOLDSTEIN, 1992; ALBANO, 2001, 2020) or the Usage-based Phonology (BYBEE, 2001). However, as already mentioned in our introduction, more recent studies have revealed that coordinated actions transcend the construction of the phonetic-phonological signifers and support language comprehension.

With recent advances in neurosciences, according to MacWhinney (2010), linguistics and correlated areas are now capable of investigating human cognitive functions down to cells and cell assemblies. Studies, such as the ones conducted by Rizzolatti and colleagues (e.g., RIZZOLATTI et al., 1996), are bringing to light strong evidence for embodied cognition (or cognitive embodiment), in which the human brain would encode a full map of the body (MACWHINNEY, 2008). Contemporary investigations in this domain are concerned with the interaction between the sensory and motor systems (MCGETTIGAN; TREMBLAY, 2018).

ISSN: 1980-0614

From the perspective of embodied cognition, Rizzolatti et al. (1996) suggest that individuals learn how to comprehend actions performed by other individuals by imitating these actions in their own motor parts of the brain. For example, there is a robust body of evidence that areas of the brain near the appropriate motor cortex areas are activated when words related to bodily actions are presented to participants in a different range of tasks (GARNHAM et al., 2006, p. 13).

To Gullberg (2008), discussing hand gestures, gestural input captures attention, provides semantic redundancy, and engages more senses by transforming speech production in a concrete experience. A possible neurocognitive explanation is connected to mirror neurons, neurons that are activated both by action and perception. This hypothesis argues that the same areas in our brain engaged in the production of a given action will activate when we observe someone else performing that action, including gestures, as if we were performing them ourselves (e.g., RIZZOLATTI; CRAIGHERO, 2004). For example, Pulvermüller, Harle and Hummel (2001) revealed that being exposed to the verb "to walk" would activate areas associated with movements of the participants' legs. To Klatzky et al. (1989), the exposition to the word *doorknob* can activate the hand shape for clenching in our brains. On this note, studies on embodied cognition propose that comprehension would be grounded in action (GULLBERG, 2008; GLENBERG; KASCHAK, 2002), or, in other words, that speech perception would also rest cognitively on actions related to production.

Therefore, studies on speech perception and mirror neurons have revived the Motor Theory (MT) of speech perception (LIBERMAN et al., 1967), based on a linguistic module evolved for communication and that speech perception and production share a common neural code (LOTTO; HICKOK; HOLT, 2009), as well as proposed some weak version of the original Liberman's model. These weak models of motor theories posit that perception calls for some aspects of the motor system or at least demand access to speech production systems (LOTTO et al., 2009). At this point, we ought to stress that we will not endorse any of these models. In fact, from our perspective, language as a dynamic system, specialized modules for language processing are not considered and therefore is incongruent with the principles supporting the MT at least in its "strong" version. We agree with Lotto and colleagues (2009) that we must temper the debate about any motor theory and mirror neurons, since results are still contradictory. Nevertheless, the body of data already created on the issue is interesting and paves the way for a better understanding about the connection between speech perception and production in models that go beyond the acoustics.

ISSN: 1980-0614

Proponents of the mirror neurons have been finding evidence for the participation of motor and premotor cortices in speech perception. Garrod (1999) and Pickering and Garrod (2004) have pointed out that successful communication rests on speakers aligning themselves both in production and perception of the language in use. Consequently, to Pickering and Garrod (2004), speech production and perception become integrated. This is not a new perspective, and psycholinguistics and neurosciences have revealed an intimate connection between speech perception and articulation as there is strong evidence that articulators would activate in speech perception (FADIGA et al., 2002).

One of the key studies in the field was conducted by Wilson et al. (2004). The authors provided one of the first pieces of evidence that specific production areas of the brain were activated during the perception of syllables initiated by voiced stops. In a similar fashion, Pulvermüller et al. (2006) investigated participants listening to syllables with bilabial and coronal stops. On silent production tests, participants imagined themselves producing those stop sounds. By means of fMRI, the group concluded that just listening to syllables with the stops led to the activation of auditory receiving areas, such as the temporal lobes, but also the motor cortex, the frontal lobe. The group also pointed out that the responses were somatotopic, that is, that listening to the syllables would activate areas in the brain that correspond to very specific parts of the body, in this case, related to the specific articulators that would be engaged in the speech production of those sounds. These studies have shown that speech perception of specific segments (e.g., stop consonants) activate brain areas related to different articulators. Similar studies using transcranial magnetic stimulation demonstrate, for example, that the perception of words that require tongue movements is associated with more robust tongue motor-evoked potentials (FADIGA et al., 2002). Pulvermüller and Fadiga (2010) claim that perception-action networks support speech perception, from speech comprehension to semantic processing.

As the investigation on the connection between speech production and perception in a neurolinguistic account is a recent development, at least considering the role of gestures, the area is still very efervescent and presents mixed claims. While some scholars hypothesize a crucial role for motor areas of the brain in speech perception, others yield more contained discussions and arguments. However, “even the most critical opponents of MT would not suggest motor and perceptual systems do not interact” (LOTTO et al., 2009, p. 5). In addition, neurosciences began to reveal “mirror-like” perception-production data (MCGETTIGAN; TREMBLAY, 2018). The link between speech perception and production is uncontroversial, and to Lotto and colleagues (2009,

ISSN: 1980-0614

p. 5), "given that we typically perceive the speech we produce, it seems unsurprising for there to be correlated neural activity corresponding to perception and production". Even those who argue against a paramount role of the motor parts of the brain in speech perception would agree that it may play a significant role in supporting the whole process (MCGETTIGAN; TREMBLAY, 2018). Studies like the ones presented in this section advocate that the speech gestures - if not intrinsically connected to - at least support speech perception (MCGETTIGAN; TREMBLAY, 2018). In this vein, both speech perception and production would be, in a way, products of action.

It is thus now known that perception-action neural networks provide support for language comprehension. Galantucci, Fowler and Tulvey (2006), for instance, point out that our general cognition indicates the importance of motor areas of the brain for perception. Even though we reject the idea of specialized modules in the brain, we claim that gestures are the primary objects of speech perception, corroborating Galantucci et al. (2006), who point out that the perception of speech demands the activation of corresponding gestures engaged in production. In this light, it is indeed possible to argue that the articulatory gesture has both an abstract and a concrete dimension, as posited by Browman and Goldstein (1992) and updated by Albano (2001)[9].

There is no debate that speech perception and production interact (LOTTO et al., 2009). We thus claim that, such as Guenther et al. (2006), speech perception and production are complementary. In this perspective, it is noteworthy that, in our opinion, these processes and/or representations do not belong to a linguistic system that is separate from other systems or perceptual processes. In fact, we agree with Lotto, Hickok and Holt (2009, p. 6), who point out that "speech production relies on speech perception and the shared representations are auditory and general (non-linguistic)". As a matter of fact, we complement such a statement by indicating that acoustic information and articulatory routines, by means of coordinated gestures, reflect the public units of the phonological grammar. This claim relates to Albano (2001) and to the view of language as a dynamic system (BECKNER et al. 2009), both pivotal for this work.

Although the mechanisms responsible for the relationship between speech perception and production in language development are not explicit, Flege (1995) points out that mechanisms and processes called for L1 phonetic-phonological development, including the formation of new categories, would remain intact throughout an individual's lifespan, and would also drive L2 development. In a similar vein, Best and Tyler (2007) posit that individuals continually refine their perceptions of speech sounds, including

[9] This update is addressed in the next section.

their own L1s. Also to these authors, as already pointed out by Flege (1995), both L1 and L2 phonetic-phonological categories would coexist in the same mental space and, then, influence each other. Thus, besides considering action in lead for speech production and perception or, in other words, considering both articulatory and acoustic dimensions, models of L2 speech development should at least try to integrate both processes instead of treating them as completely independent constructs.

GESTURE-DRIVEN L2 DEVELOPMENT

By discussing the traditional discrepancies posited by the status-quo approach to the phonological grammar and its relation to speech perception and production, Fowler and Galantucci (2005, p. 636) believe that "the elements of phonological competence have their primary home in the vocal tract, not in the mind". Such elements, in their point of view, correspond to linguistically significant actions of the vocal tract. This assumption is aligned with Albano's (2001) position, according to which we learn how to orchestrate phonic gestures by the action of doing them. We are by no means presuming that cognitive mechanisms responsible for abstract representations are not at play in speech perception or production. Actually, we argue that abstract representations are cognitively instantiated by gestural, coordinated actions of speech perception and production.

As illustrated in the beginning of the present article, SLM-r (FLEGE; BOHN, 2021) and SLM (FLEGE, 1995) could, at first sight, handle speech perception and production, since they settle on the notion that perception and production co-evolve during lifespan. Nevertheless, neither are such models able to account for the gestural primitive that encompasses an action-based approach to phonological elements, nor can they provide satisfactory answers to the relation between phonological knowledge and speech perception and production. A promising theoretical strategy would be to call for the Perceptual Assimilation Model for L2 Speech Learning (BEST; TYLER, 2007), also known as PAM-L2, as this perceptual model operates with a gestural unit and relates common and complementary aspects of nonnative and second-language[10] speech perception. However, based on Perozzo (2017), we disagree with PAM-L2 (BEST; TYLER, 2007) in respect to two tenets, which we shall outline below.

The first tenet is expressed by the nature of the phonic gestures. While Best and Tyler (2007) adopt the articulatory gesture developed by Browman and Goldstein

[10] For the purposes of the present article, we do not make any distinctions between the terms "nonnative" and "second-language speech perception".

ISSN: 1980-0614

(1989, 1992) and Goldstein and Fowler (2003), Perozzo (2017) understands that the acoustic-articulatory treatment of the phonic gestures, as established by Albano (2001), provides a clearer picture of a gesture-driven L2 development. One of Albano's (2001) theoretical advances is revising Browman and Goldstein's (1989, 1992) model and adapting it in terms of at least three circumstances, which permeate her work. At first, there is the notion of an auditory-acoustic bond intrinsic to the articulatory gesture, which incorporates quantum and adaptive dispersion criteria. The second one is based on the assumption that the symbolic projection of the gesture is stated by its borders. The third one designates that the realignment and the redimensioning of gestures are the means by which phonological regularities penetrate deeper levels of the grammar. Not surprisingly, these three circumstances are connected to coordinated actions in the vocal tract.

Albano's (2001) work showcases three fundamental factors that support the relation between phonic gestures and the perceptual event (with consequences for abstract representation). The first factor concerns the proportionality with which the researcher relates the abstract (mental) facet of the phonic gesture with its physical (motor) facet. The second factor lies in the approximation of the phonic gesture to other units that operate in grammar, such as the morpheme. Finally, the third factor is related to the relevance of acoustic information imbricated in the gestural primitive, which is established in terms of action (the result of articulation) and representation (symbolism).

Even though Goldstein and Fowler (2003) admit that there is abstraction with regard to articulatory gestures, their argument is, to a large extent, led to emphasize the physical character of such a unit. It is in this regard that we deem more convenient Albano's (2001) view, in which the abstract and physical facets of the phonic gestures seem to be more balanced. Put it differently, we consider that there seems to be a greater symbolism in Albano's (2001) perspective compared to that of Browman and Goldstein (1989, 1992). Moreover, by undertaking the construct of a phonological grammar, Albano (2001) creates, at the symbolic level, close links with other pieces of abstract knowledge.

As for the relevance of the acoustic aspects in terms of phonic gestures, not only do they support the conception of an acoustic-articulatory phonology, but they also relate to two extremely important theoretical considerations: (1) although acoustics and articulation are linked to a causal relation in natural productions, since the acoustic signal of speech derives from constrictions involving the variables of the vocal tract, they can result in different consequences for the auditory interpretation of gestures (affrication vs. palatalization, for example); and (2) visualizing the articulation of a given gesture

and simultaneously hearing its acoustic properties enables the perceiver to approach the communicative event in a multimodal fashion[11], as complex as communication can be.

The second tenet on which we disagree with PAM-L2 is its philosophical foundations. Instead of adopting a direct realist view of L2 speech perception, we align with indirect realism (JACKSON, 1977, 2010; LOWE, 1981), as argued by Perozzo (2017). Indirect realism predicts that the external world exists (including the phonic gestures), but our perception of it is mediated by the perception of intermediate and subjective abstract objects, such as, for example, sensations (BROWN, 2009). As explained by Dancy (1985), the indirect aspect of realism assumes that we never access physical objects directly, since we directly access an intermediate apparatus - which can be ideas, images, impressions, sensations or sense data (MOUND, 2003). In other words, the distal object (objective, public) is perceived indirectly due to the direct perception of a proximal object (subjective, private). As an example, the existence of something between the perceiver and the object to be perceived can be associated with language attrition. Kupske (2016), for instance, points out that Brazilian first-generation immigrants in an English-dominant context yield Voice Onset Time (VOT) values for both L1 and L2 voiceless stops that are intermediate between the short lag Brazilian Portuguese and the long lag English. Data like these meet the hypothesis that the perception of L2 phonological elements is filtered by the L1 phonological knowledge (which refers to one of several possible intermediate objects), conjecture in favor of which we position ourselves.

The existence of an intermediate apparatus between the public/real object and the perceiver suits the notion that our perception is driven by cognitive mechanisms, which has recently entailed a great body of research in neurosciences and cognitive psychology. To start with, Beckner et al. (2009, p. 2) emphasize that "the structures of language emerge from interrelated patterns of experience, social interaction, and cognitive mechanisms". Additionally, Kandel (2014) explains that perception essentially corresponds to a process of cognitive construction that depends not only on the external stimulus but also on the mental apparatus of the subject who experiences the perceptual event, that is, perception is largely dependent on sensory and motor systems in the brain. At the same time, Gazzaniga et al. (2012) mention that the perception of the world does not operate as a camera or as an audio recorder, which faithfully and passively grasps the properties of the stimuli

[11] Multimodality in speech perception is initially explored in a seminal work by MacDonald and McGurk (1978), and brings thoughtful considerations about the subject.

we have access to. According to the scholars, what we taste, hear, see, touch or smell, results from brain processes that construct perceptual experiences. It is clear, therefore, that the perception of the objects of the world is only capable of being realized if there are cognitive resources that give support to the construction of the perceptual phenomenon.

As Perozzo (2017) points out, Gazzaniga et al. (2012) explain that the brain does not process raw stimuli, implying that they are necessarily translated into chemical and electrical signals so that the brain can interpret them. This way, different properties of the physical world are codified or translated by different patterns of neural impulses, an operation called sensory encoding (GAZZANIGA et al., 2012; GARDNER, JOHNSON, 2014). Sensory encoding begins with transduction, where sensory receptors - specialized neurons - produce neural impulses at the moment they receive physical or chemical stimulation. The information at play is then transmitted to the brain in the form of neural impulses.

Being able to mentally operate on the information to which we have access does not imply that external stimuli are inadequate, flawed or poor: it only designates our cognitive ability to process and interpret such pieces of information based on the ecology of our experiences. We thus assume that the environment around us is capable of providing a vast source of multimodal information about objects and facts; however, these only have meaning if they are understood from the perspective of our previous experiences, which pervade the perceptual phenomenon. According to Haugen (2001), languages are constantly being redesigned by the interactions of their speakers in order to reflect the communicational experiences of the past, and to project current and future ones. Thus, any behavior of a speaker is the result of a range of competing factors, including physical, cognitive and social motivational factors (SCHERESCHEWSKY; ALVES, KUPSKE, 2017). As we mentioned before, we highlight that abstract representations are cognitively instantiated by gestural, coordinated actions of speech perception and production. Those gestures find their roots on the vocal tract and, by means of their action, they integrate our phonological knowledge.

We believe that it is quite challenging for a model of L2 speech development to try to integrate both perception and production. Therefore, when it comes to modeling speech production in the same sphere of speech perception and from the perspective of phonic gestures, one should consider theoretical issues that go all the way along the coordination and action of the tract variables. Besides, it is mandatory to capture how spoken forms (through action) connect physically to stored forms (through perception) in a cognitive environment, so that the former can be recognizable to the latter and the

ISSN: 1980-0614

latter can make sense of the former - the parity principle (LIBERMAN; WHALEN, 2000; FOWLER; GALANTUCCI, 2005).

Should these two points we have just made be not enough for accomplishing the task of addressing both L2 perception and production within the same model, other crucial parameters cannot be overlooked. Any model that aims to account for L2 speech production is expected to reckon with coarticulation, coproduction, timing, and the metrical organization of speech, to name a few attributes. These variables gather some inceptive production parameters that are intimately related to speech perception, and, therefore, should be rationalized in an integrated model.

We suppose that an auspicious model of L2 speech perception and production is the one that successfully operates with phonic gestures in a way that these can be common units to perception, production, and representation, and reflects the role of the learner's experience with the world and the most diverse situated manifestations.

FINAL REMARKS

In this paper, we have argued that L2 abstract representations are cognitively instantiated by coordinated actions of speech perception and production. Such actions correspond to phonic gestures that find their roots on the vocal tract and, by means of their orchestration, they integrate our phonological knowledge. An outline of some theories and models of speech perception and production have been tackled, as well as some points have been made in terms of phonological grammar.

Additionally, we have claimed that speech perception and production are indissociable. We have also presented a theoretical framework that takes the phonic gesture as the phonological primitive, as well as its likeliness to permeate both L2 speech perception and production. However, by means of an analysis of some hegemonic models of L2 speech development, it is clear the incorporation of phonic gestures is indeed a challenge. On the one hand, some models of L2 speech solely rely on the acoustic information and ignore the role of coordinated action in language development. On the other hand, models that adopt speech gestures might fail in accommodating the interaction between speech perception and production, the acoustics, and the dynamic and still symbolic nature of L2 grammar. In this light, this paper marks the starting point of a mission aimed at understanding and developing an integrative, gesture-driven, approach to L2 speech development.

ACKNOWLEDGEMENTS

We would like to thank the two anonymous reviewers whose comments helped improve and clarify this manuscript.

FUNDING

Felipe Flores Kupske would like to acknowledge all the support received from the National Council for Scientific and Technological Development (CNPq, Brazil) - Process no. 432396/2018-7.

REFERENCES

ALBANO, E. **O gesto e suas bordas:** esboço de fonologia acústico-articulatória do português brasileiro. Campinas, SP: Mercado de Letras, 2001.

ALBANO, E. **O gesto audível:** fonologia como pragmática. São Paulo: Cortez, 2020.

ALVES, U.; SILVA, A. Implicações de uma perspectiva realista direta para o PAM-L2: Desafios teórico-metodológicos. **Revista do GEL**, v. 13, n. 1, p. 107-131, 2016.

BECKNER, C.; BLYTHE, R.; BYBEE, J.; CHRISTIANSEN, M.; CROFT, W.; ELLIS, N.; HOLLAND, J.; KE, J.; LARSEN-FREEMAN, D.; SCHOENEMANN, T. Language is a Complex Adaptive System: Position Paper. Language is a complex adaptive system: position paper. **Language Learning**, Michigan, v. 59, p. 1-26, Dec. 2009.

BERTICELLI, I. **Educação em perspectivas epistêmicas pós-modernas**. Chapecó: Argos, 2010.

BEST, C. A direct realist view of cross-language speech perception. In: STRANGE, W. (Ed.). **Speech perception and linguistic experience:** issues in cross-language research. Timonium, MD: York Press, 1995, p. 171-204.

BEST, C.; TYLER, M. Nonnative and second- language speech perception: commonalities and complementarities. In: BOHN, O.; MUNRO, M.. **Language Experience in Second Language Speech Learning:** In honor of James Emil Flege. Amsterdam: John Benjamins, 2007, p. 13-34.

BOHN, O-S. **Core aspects of the revised Speech Learning Model (SLM-r)**, 2020, 64 slides. Disponível em: https://snuling.com/materials/Bohn-SNU-2020/SNU_3_Core_aspects_of_the_revised_Speech_learning_Model_(SLM-r)_hdt.pdf. Acesso em: 23 Mai. 2021.

BROWMAN, C.; GOLDSTEIN, L. Articulatory gestures as phonological units. **Haskins Laboratories Status Report on Speech Research**, v. 100, p. 69-101, 1989.

BROWMAN, C.; GOLDSTEIN, L. **Articulatory Phonology** - an overview. Haskins Laboratories Status Report on Speech Research, SR-111/112, p. 23-42, 1992.

BROWN, D. Indirect perceptual realism and demonstratives. **Philosophical Studies**, v. 145, n. 3, p. 377-394, 2009.

BYBEE, J. **Phonology and Language Use**. Cambridge: Cambridge Univ. Press, 2001.

BYBEE, J. **Language, Usage and Cognition.** Cambridge: Cambridge Univ. Press, 2010.

CALLOU, D.; LEITE, Y. Iniciação à fonética e à fonologia. Rio de Janeiro: Zahar, 1994.

COLCHESTER, J. **Systems + Complexity:** an overview. Londres: CreateSpace Independent Publishing Platform, 2016.

COLE, R.; SCOTT, B. Toward a theory of speech perception. **Psychological Review**, v. 81, n. 4, p. 348-374, 1974.

DANCY, J. **Epistemologia contemporânea**. Cambridge: Harvard University Press, 1985.

ELLIS, N. The dynamics of language use, language change, and first and second language acquisition. **Modern Language Journal,** v. 41, n. 3, p. 232-249, 2008.

EVANS, B.; ALSHANGITI W. Regional Accent Accommodation in spontaneous speech: evidence for long-term accent change? **Proceedings of ICPhS XVII**, Hong Kong, Japan, p. 224- 227, 2011.

EVANS, B.; IVERSON, P. Vowel normalization for accent: An investigation of best exemplar locations in northern and southern British English sentences. **J. Acoust. Soc. Am.,** v. 115, n. 1, p. 352– 361, 2004.

FADIGA, L.; CRAIGHERO, L.; BUCCINO, G.; RIZZOLATTI, G. Speech listening specifically modulates the excitability of tongue muscles: A TMS study. **European Journal of Neuroscience,** v. 15, n. 2, p. 399–402, 2002.

FLEGE, J. Second language speech learning: Theory, findings, and problems. In: STRANGE, W. (Ed.). **Speech perception and linguistic experience:** Theoretical and methodological issues in cross-language speech research. Timonium: York Press, 1995. p. 233-272.

ISSN: 1980-0614

FLEGE, J.; BOHN, O-S. The revised speech learning model (SLM-r) 2021 In: WAYLAND, R. (Ed.). **Second language speech learning:** theoretical and empirical progress. Cambridge: Cambridge University Press, 2021. p. 3-84.

FOWLER, C. An event approach to the study of speech perception from a direct-realist perspective. **Journal of Phonetics,** v. 14, p. 3-28, 1986.

FOWLER, C. Listeners do hear sounds, not tongues. **Journal of the Acoustical Society of America,** v. 99, n. 3, p. 1730-1741, 1996.

FOWLER, C.; GALANTUCCI, B. The relation of speech perception and speech production. In: PISONI, D.; REMEZ, R. (Eds.). **The handbook of speech perception**. Padstow: Blackwell Publishing, 2005. p. 633-652.

FREITAS, M. **O gesto fônico na aquisição "desviante":** movimentos entre a produção e a percepção. Tese de doutorado. LAFAPE-IEL-UNICAMP, 2012.

GALANTUCCI, B.; FOWLER, C.; TURVEY, M. T. The motor theory of speech perception reviewed. **Psychon Bull Rev**, v. 13, n. 3, p. 361–377, 2006.

GARDNER, E.; JOHNSON, K. A codificação sensorial. In: KANDEL, E.; SCHWARTZ, J.; JESSEL, T.; SIEGELBAUM, S.; HUDSPETHET, A. (Eds.). **Princípios de neurociências**. 5. ed. Porto Alegre: Artmed, 2014. p. 392-413.

GARNHAM, A.; GARROD, S.; SANFORD, A. Observations on the past and future of psycholinguistics. In: TRAXLER, M.; GERNSBACHER, M. A. (Eds.). **Handbook of Psycholinguistics**. London: Academic Press, 2006. p. 1-18.

GARROD, S. The challenge of dialogue for theories of language processing. In: GARROD, S.; PICKERING, M. (Eds). **Language processing**. Hove, UK: Psychology Press, 1999. p. 389–415.

GAZZANIGA, M.; HEATHERTON, T.; HALPERN, D.; HEINE, S. **Psychological Science**. Nova Iorque: W.W. Norton & Company, 2012.

GIBSON, J. **The senses considered as perceptual systems.** Boston: Houghton Miffin, 1966.

GIBSON, J. **The ecological approach to visual perception.** Nova Iorque: Psychology Press, 1986.

GLENBERG; A.; KASCHAK, M. Grounding language in action. **Psychonomic Bulletin & Review,** v. 9, n. 3, p. 558–565, 2002.

ISSN: 1980-0614

GOLDSTEIN, L.; FOWLER, C. Articulatory Phonology: a phonology for public language use. In: MEYER, A.; SCHILLER, N. (Eds.). **Phonetics and phonology in language comprehension and production:** Differences and similarities. Berlim: Mouton de Gruyter, 2003. p. 159-207.

GUENTHER, F. H., GHOSH, S. S., TOURVILLE, J. A. Neural modeling and imaging of the cortical interactions underlying syllable production. **Brain Lang**, v. 96, n. 3, p. 280–301, 2006.

GULLBERG, M. Gestures and second language acquisition. In. ROBINSON, P.; ELLIS, N. C. (Eds.). **Handbook of cognitive linguistics and second language acquisition**. New York: Routledge, 2008. p. 276-306.

HAUGEN, E. The ecology of language. In: FILL, A; MÜHLHÄUSLER, P. (Org.). **The ecolinguistics reader**. London: Continuum, 2001. p. 57-66.

JACKSON, F. **Perception**. Nova Iorque: Cambridge University Press, 1977.

JACKSON, F. Representative realism. In: DANCY, J.; SOSA, E.; STEUP, M. (Eds.). **A companion to epistemology**. 2. ed. Malden: Blackwell, p. 702-705, 2010.

JAKOBSON, R., FANT, G., HALLE, M. Preliminaries to speech analysis: the distinctive features and their correlates. Cambridge: MIT Press, 1952.

KANDEL, E. Das células nervosas à cognição: As representações internas de espaço e ação. In: KANDEL, E.; SCHWARTZ, J.; JESSEL, T.; SIEGELBAUM, S.; HUDSPETHET, A. (Eds.). **Princípios de neurociências**. 5. ed. Porto Alegre: Artmed, p. 327-344, 2014.

KLATZKY, R.; PELLEGRINO, J.; MCCLOSKEY, B.; DOHERTY, S. Can you squeeze a tomato? The role of motor representations in semantic sensibility judgments. **Journal of Memory & Language**, v. 28, n. 1, p. 56–77, 1989.

KUHL, P.; MILLER, J. Speech perception by the chinchilla: voiced-voiceless distinction in alveolar plosive consonants. **Science**, v. 190, n. 4209, p. 69-72, 1975.

KUPSKE, F. **Imigração, Atrito e Complexidade:** A Produção das Oclusivas Surdas Iniciais do Inglês e do Português por Sul-Brasileiros Residentes em Londres. Tese (Doutorado em Letras). Porto Alegre: Universidade Federal do Rio Grande do Sul, 2016.

KUPSKE, F. A complex approach on integrated late bilinguals' English VOT production: a study on South Brazilian immigrants in London. **Ilha do Desterro**, Florianópolis, v. 70, n. 3, p. 81-93, set./dez. 2017. Disponível em: https://doi.org/10.5007/2175-8026.2017v70n3p81. Acesso em: 1 Jun. 2021.

ISSN: 1980-0614

KUPSKE, F. The impact of language attrition on language teaching: the dynamics of linguistic knowledge retention and maintenance in multilingualism. **Ilha do Desterro**, Florianópolis, v. 72, n. 3, p. 311-329, 2019. DOI: 10.5007/2175-8026.2017v70n3p81.

KUPSKE, F.; PEROZZO, R.; ALVES, U. Sound change as a complex dynamic phenomenon and the blurriness of grammar stability. **Macabéa:** Revista Eletrônica do Netlli, Crato, v. 8, n. 2, p. 158-172, jul./dez. 2019.

LIBERMAN, A.; COOPER, F.; SHANKWEILER, D.; STUDDERT-KENNEDY, M. Perception of the speech code. **Psychological Review**, v. 74, n. 6, p. 431-461, 1967.

LIBERMAN, A.; MATTINGLY, I. The motor theory of speech perception revised. **Cognition**, v. 21, n. 1, p. 1-36, 1985.

LIBERMAN, A.; WHALEN, D. On the relation of speech to language. **Trends in Cognitive Sciences**, v. 4, p. 187-196, 2000.

LOTTO, A.; HICKOK, G.; HOLT, L. Reflections on mirror neurons and speech perception. **Trends Cogn Sci.**, v.13, n. 3, p. 110–114, 2009. doi:10.1016/j.tics.2008.11.008.

LOWE, J. Indirect perception and sense data. **The Philosophical Quarterly**, v. 31, n. 125, p. 330-342, 1981.

MACDONALD, J.; MCGURK, H. Visual influences on speech perception processes. **Perception & Psychophysics**, v. 24, p. 253–257, 1978.

MACWHINNEY, B. How mental models encode embodied linguistic perspectives. In KLATZKY, R.; MACWHINNEY, B.; BEHRMANN, M. (Eds). **Embodiment, Ego-Space, and Action**. Mahwah NJ: Lawrence Erlbaum Associates, 2008. p. 369–410.

MACWHINNEY, B. A tale of two paradigms. In. KAIL, M.; HICKMANN, M. (Eds.). **Language Acquisition across Linguistic and Cognitive Systems**. New York: John Benjamins, 2010. p. 17-32.

MCGETTIGAN, C.; TREMBLAY, P. Links between perception and production: examining the roles of motor and premotor cortices in understanding speech. In. GASKELL G.; RUESCHEMEYER S. (Eds.). **Oxford Handbook of Psycholinguistics**. Oxford: Oxford University Press, 2018. p. 306-334.

MOUND, B. **Perception**. Durham: Acumen, 2003.

NISHIDA, G. **Sobre Teorias de percepção da fala**. 2012. 174 f. Tese (Doutorado em Letras)—Setor de Ciências Humanas, Letras e Artes, Universidade Federal do Paraná, Curitiba, 2012.

ISSN: 1980-0614

OHALA, J. Speech perception is hearing sounds, not tongues. **Journal of the Acoustical Society of America**, v. 99, n. 3, p. 1718-1725, 1996.

PARDO, J. On phonetic convergence during conversational interaction. **Journal of the Acoustical Society of America**, v. 119, n. 1, p. 2382-2393, 2006

PEROZZO, R. **Sobre as esferas cognitiva, acústico-articulatória e realista indireta da percepção fônica não-nativa: para além do PAM-L2**. 2017. 225 f. Tese (Doutorado em Letras: Estudos da Linguagem)—Instituto de Letras, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2017.

PEROZZO, R. Interseções entre ciência e linguística: do reducionismo analítico à complexidade. **Estudos Linguísticos e Literários**, n. 64, p. 130-154, 2019.

PICKERING, M.; GARROD, S. Toward a mechanistic psychology of dialogue. **Behavioral and Brain Sciences**, v. 27, n. 2, p. 169-226, 2004.

PISONI, D. Auditory and phonetic memory codes in the discrimination of consonants and vowels. **Perception and Psychophysics**, 13, 1973. p.253-260.

PISONI, D.; TASH, J. Reaction times to comparisons within and across phonetic categories. **Perception and Psychophysics**, 15, 1974. p. 285-290.

PULVERMÜLLER, F.; FADIGA, L. Active perception: sensorimotor circuits as a cortical basis for language. **Nat Rev Neurosci.**, v. 11, n. 5, p. 351-60, 2010. doi: 10.1038/nrn2811.

PULVERMÜLLER, F.; HÄRLE, M.; HUMMEL, F. Walking or talking? Behavioral and neurophysiological correlates of action verb processing. **Brain Lang.**, v. 78, n. 2, p. 143-68, 2001. doi:10.1006/brln.2000.2390.

PULVERMÜLLER, F.; HUSS, M.; KHERIF, F.; MARTIN, F.; HAUK, O.; SHTYROV, Y. Motor cortex maps articulatory features of speech sounds. **Proc Natl Acad Sci USA**, v. 103, n. 20, p. 7865-70, 2006. doi: 10.1073/pnas.0509989103.

RIZZOLATTI, G.; CRAIGHERO, L. The mirror-neuron system. **Annual Review of Neuroscience**, v. 27, n. 1, p. 169–192, 2004.

RIZZOLATTI, G.; FADIGA, L.; GALLESE, V.; FOGASSI, L. Premotor cortex and the recognition of motor actions. **Cognitive Brain Research**, v. 3, n. 1, p. 131–141, 1996.

SCHERESCHEWSKY, L.; ALVES, U.; KUPSKE, F. First language attrition: the effects of English (L2) on Brazilian Portuguese VOT patterns in an L1-dominant environment. **Letrônica**, v. 10, n. 2, p. 700-716, 2017. DOI: 10.15448/1984-4301.2017.2.26365.

ISSN: 1980-0614

TRUBETSKOY, N. Principes de phonologie. Paris: Klincksieck, 1949.

TATHAM, M.; MORTON, K. **A guide to speech production and perception**. Edinburgh: Edinburgh University Press, 2011.

WILSON, S.; SAYGIN, A.; SERENO, M. I.; IACOBONI, M. Listening to speech activates motor areas involved in speech production. **Nat Neurosci.**, v. 7, n. 7, p. 701-2, 2004. doi: 10.1038/nn1263.

Received: 01 June 2021.
Accepted: 01 August 2021.