SCUOLA NORMALE SUPERIORE

Classe di Scienze Umane

TESI DI PERFEZIONAMENTO (PHD) IN LETTERATURE E FILOLOGIE MODERNE

HARD TO SAY, HARD TO SEE?
SPEECH-IN-NOISE DISCRIMINATION AT DIFFERENT LEVELS OF SENSORIMOTOR PROFICIENCY

Relatore:	Candidato:
Prof. PIER MARCO BERTINETTO	IRENE LORENZINI

Anno accademico 2017/2018

Acknowledgements are well-hidden, after the bibliographic references

Chapter 1 DEVELOPING SPEECH PERCEPTION – PART ONE:

AUDITORY PROCESSING FROM BIRTH TO CHILDHOOD

2. Developmental milestones in the first year of life: Infancy

3. Perceiving a meaningful signal: From infancy to early childhood

4. Childhood: The long tail of perceptual development

Chapter 2 DEVELOPING SPEECH PERCEPTION – PART TWO: MULTISENSORIALITY

2. Geared to integrate? Audiovisual skills in the first year of life

3. Neuropsychological mechanisms underlying multisensory speech perception

4. Children: Multimodal perception and ongoing sensorimotor specialization

Chapter 3 DEVELOPMENTAL PRODUCTION DISORDERS AND THEIR PERCEPTUAL CORRELATES

2. Childhood Apraxia of Speech, or the uncontrolled movement

Chapter 4 SPEECH-IN-NOISE DISCRIMINATION AT DIFFERENT LEVELS OF SENSORIMOTOR PROFICIENCY

3. Methods - Perception task

4. Results - Perception tasks

5. Methods - Production tasks

6. Results – Production tasks

Chapter 5 GENERAL DISCUSSION

Appendix SENSORIMOTOR CONTRIBUTIONS TO EARLY LEXICAL PROCESSING? A CURRENT RESEARCH PROJECT

1. DEVELOPING SPEECH PERCEPTION – PART ONE:

AUDITORY SPEECH PERCEPTION FROM BIRTH TO CHILDHOOD

Foreword: On the language-specific nature of perception and its development

«I discover vision (…) as a gaze at grips with a visible world»

(Merleau-Ponty, 1945, p. 409)

Unlike any other linguistic skill, speech perception is already functional before birth. Perception provides infants with the first piece of the language acquisition's puzzle and its pivotal role has been proved by the many studies documenting the connection between perceptual abilities and later language outcomes in typical and atypical populations[1].

Despite the universality of this process, during development, the elaboration of linguistic sounds unavoidably passes through experience with one (or more) concrete form(-s) of the percept, i.e. the input language(-s). In this sense, and much as the other perceptual capacities, speech perception is only potentially owned as a birthright and rather developed and refined through ontogenesis. Such perspective is critical for this dissertation, focused on the accomplishment of those abilities during childhood years.

More in detail, if studies pertaining to the developmental course of speech perception have demonstrated a striking precocity during infancy, evidence also displays that some perceptual aspects reach maturation much later, following a long-tailed path which culminates at the end of childhood. This seems particularly true with reference to the topics addressed by this research, i.e. segmental processing, speech-in-noise elaboration and the ability to rely on the visual component of the signal.

The first two chapters aim at offering an overview on the developmental trajectory of speech perception which, going from infancy to childhood and describing auditory and audiovisual data (Chap. 1 and 2 respectively), will allow the statement of the problem.

---

[1] E.g. Benasich & Tallal, 2002; Tsao et al., 2004; Werker & Yeung, 2005; Kuhl et al., 2005.

1. Geared to listening: In-womb perception and the first days of life

In the mother's womb, the auditory system starts operating around the 25th gestational week. Thus, hearing represents the first direct contact with the world outside and, pervasive as communication is, language constitutes a substantial part of the prenatal acoustic landscape.

In this phase, the mother's voice directly reaches the fetus with its vibrations, even if covered with endogenous noise; while speech coming from the external environment is heard as filtered by the natural barriers of the mother's body and the amniotic fluid. As a consequence, the spared signal is attenuated in frequencies higher than 1000Hz, essentially cutting out much of the information characterizing the segmental level, but allowing suprasegmental perception (Griffiths et al., 1994; Langus & Nespor, 2013). Accordingly, studies investigating the fetal reaction through heart rate monitoring have shown that, in this period of the in-womb life, fetuses can detect changes between a female and a male voice (e.g. Lecanuet et al.,1993) and display learning processes over recurrent phonetic patterns, such as the frequencies characterizing their mother's voice but also specific speech passages regularly heard during a habituation period (Kisilevsky et al., 2003; DeCasper et al., 1994).

Soon after birth, some emergent abilities are indeed connectable to in-womb experience.

In this connection, neonates demonstrate a preference for their mother's speech over others (Mehler et al., 1977; DeCasper & Fifer, 1980) as well as for their native language as compared to an unknown[2] one or to backward speech (May et al., 2011; Peña et al., 2003; Vouloumanos & Werker, 2004 and 2007).

Finer-grained outcomes of *in utero* experience have also been shown. In this respect, a cross-linguistic experiment conducted with Sweden and American-English subjects have displayed that prenatal exposure sufficed to allow discrimination of native *vs* non-native vowels, inducing prototype effects[3] (Moon, Lagercrantz & Kuhl, 2013).

Overall, data on prenatal learning suggest cautiousness in considering birth as «a benchmark that reflects a complete separation between the effects of nature versus those of nurture», since «neonates' perception already reflects some degree of learning»[4].

This said, in the first days of life infants also make proof of abilities and behaviors that cannot be explicated on the basis of *in-utero* learning. Among these, they discriminate between

---

[2] Provided it differs in some fundamental prosodic characteristics, cf. page 3.

[3] I.e. better recognition of the central, prototypical phonological value V_1 as compared to non-prototypical ones and a tendency towards the assimilation to V_1 of similar sounds falling into its scope.

[4] Moon, Lagercrantz & Kuhl (2013), p. 160.

unfamiliar female and male speakers, which implies the processing of vocal-tract related information pertaining to previously unheard voices (Floccia, Nazzi & Bertoncini, 2000). Furthermore, electrophysiological and neuro-metabolic responses testify to a very early sensitivity to segmental contrasts and to cross-linguistically preferred phonotactic patterns (Dehaene-Lambertz & Peña, 2001; Gómez et al., 2014).

Research investigating the roots of such precocity have highlighted that the neurofunctional network underlying phonetic processing already displays a considerable efficiency at the beginning of life (cf. Section 1.1). This substrate would allow very young infants to rely on effective abilities of pattern extraction which, in turn, would represent a primary key to break the phonological code when applied to two macro-families of cues: the statistical and prosodic characteristics of the signal. On the one side, the regularities pertaining to the distributions of any kind of acoustic feature in the input; on the other, the acoustic configurations concerning phonological units larger than the phoneme (stress, pitch, length, pauses and their emerging property: rhythm) – a component of the signal with which the newborn is already familiar.

In this connection, during the first week, newborns have appeared sensitive to schemas of adjacent repetition (ABB structures like /mubaba/ and /penana/ – cf. Gervain et al., 2008) as well as to scale invariance, a basic property of natural sounds, consisting in exhibiting similar physical patterns at different levels of observation (Gervain et al., 2016). A large amount of evidence testifies the role played by statistical learning in later phonological acquisition and word segmentation (cf. Section 2.1).

As to suprasegmental processing, Mehler et al. (1988) first demonstrated that four-day-old infants are able to distinguish their native language from another even if it is low-pass filtered and segmental information is cut off. Crucially, this ability does not depend on familiarity; on the contrary, discrimination is successful with unknown languages, provided they differ in rhythmical characteristics (Nazzi et al., 1998a). Though not exclusively human (Ramus et al., 2000), this strong sensitivity to rhythm is fundamental in the neonatal period, and not only in acoustic terms (Chap. 2, page 23).

Finally, very young infants also display detection of prosodic components in isolation, such as pitch, intensity and vowel length (e.g. Sansavini et al., 1997 and Nazzi et al., 1998b); a capacity that has been connected with the detection of word boundaries and of some grammar-relevant basic differences like the distinction between open-class and closed-class words (Shi et al., 1999; Ferry et al., 2016).

1.1 Biased to listening: Neural correlates

As described in the former paragraphs, newborns appear ‘biased’ towards speech processing, i.e. coming to life equipped with cognitive predispositions to process the signal that, *ceteris paribus*, guide them through language acquisition. In the last two decades, neuroimaging techniques have grounded such early skills in the morphological and functional neural endowment at birth.

Even if in-vivo research has surely not realized its full potential yet, it has already challenged previous standpoints understanding neurocognitive development as the passage from sheer sensory processing to abstract thought. In particular, a complex neurofunctional organization has been documented in the infant brain starting from the last trimester of gestation and, in this context, the participation of frontal areas and an early emergence of the basic left-right hemispheric asymmetries have been highlighted (cf. Dehaene-Lambertz & Spelke, 2015 for a review).
Overall, these findings indicate a close genetic-based continuity between infants and adults. However, the contemporary vision of inborn predispositions is remarkably different from older innatists views, and specifically in that «architectural and chronotopic innate constraints are kept, while representational innateness is discarded»[5] – a consideration which allows for a brief excursus.

The field of developmental psychology is still covered with the ashes of the debate opposing nature and nurture, inborn modularity and *tabula rasa*; yet, research has progressed enough to move beyond binary thinking and embrace more complex syntheses, in which both factors are acknowledged. Under this perspective, ontogenesis is nowadays understood as the result of a crucial dynamics between genetic endowment and relevant experience: the former equipping the individual with predispositions towards the *locus* and *modus* of elaboration of different stimuli; the latter providing the encounter between such dispositions and their corresponding objects in the world.

Relevant, in this connection, is evidence showing that even early biases are susceptible to experience-driven refinement.

An important contribution in this sense has been provided by Vouloumanos et al. (2010), showing that newborns do not display a preference between speech and rhesus monkey calls and prefer primate vocalizations to non-speech stimuli. This shared preference declines (rapidly) after approximately three months of language exposure. In the same direction, Dehaene-Lambertz and

---

[5] Karmiloff-Smith (2006), p. 9.

colleagues (2004) have documented the spontaneous recovery path of a subject struck by neonatal left sylvian infarct who, six weeks after, demonstrated satisfactory perceptual abilities supported by the right hemisphere. Though powerful in guiding development, innate predispositions are not the only factor at play: neuroplasticity, forming and reorganizing synaptic connections, plays a coprotagonist role.

2. Developmental milestones in the first year of life: infancy

During the first year of life, some well-known perceptual milestones are completed. In particular, infants tune into their native phonological system(-s) and become progressively proficient at segmenting the speech stream into units, thus paving the way to lexical storage.

2.1 Cracking the code: Prosody and statistics

As previously mentioned, remarkable precursors of pattern-extraction skills are observable since the first days of life (page 3). In the months following birth, young children effectively rely on these aspects. A notable exemplification of such mechanisms is represented by distributional learning, describable as the ability to compute the relative frequencies of phonetic tokens occurring in subregions of the acoustic space. E.g., while an English native hears one space of allophonic variation concentrating around a central area for the alveolar /d/, a Hindi one perceives two spaces of variation, revolving around the dental /d̪/ and the retroflex /ɖ/. The two infants subdivide the acoustic continuum differently: creating a unique category in the unimodal distribution, two in the bimodal one[6]. Several studies have proved the key role of this kind of processing in the extraction of phonological categories (e.g. Maye et al. 2002, 2008; Werker et al., 2007).

Another well-studied phenomenon is the infants' capacity to keep track of Transitional Probabilities, the forward and backward conditional probabilities to have a sound *x* when a sound *z* is encountered (so that, in an English sequence such as *pretty baby*, the forward transitional probabilities between the syllables /prɪ./ and /.tɪ/ are higher than those between /.tɪ/ and /beɪ/ and this creates a cue for the location of the word boundary). As documented by Saffran, Aslin & Newport in their landmark studies (1996, 1998), transitional probabilities represent a fundamental element for infants during speech segmentation, sufficient even in the absence of prosodic information and after a very brief time of exposure. Worth to be noted, different statistical strategies seem differentially successful when applied to cross-linguistically different languages (Saksida, Langus & Nespor, 2017).

---

[6] Example from Werker et al. (2012).

As described (page 3), another crucial source of perceptual information during these early stages is represented by the suprasegmental signal.

At 4 months of age, monolingual infants appear advanced enough in this respect to recognize their native language against a phonologically similar one on prosodic grounds (low-pass filtered speech, Bosch & Sebastián-Gallés, 1997). Besides, such ability to exploit prosody encompasses all its individual components. As representative examples, English 9-months-olds listening to word streams demonstrate a preference for words mirroring their native accentual pattern (cf. Jusczyk et al., 1993) and French 12-months-olds process disyllabic stimuli through a syllable-by-syllable segmentation, coherently with the French accentual patterns (cf. Nazzi et al., 2006).

Relevant in this respect, the issue of a priority order in resorting to either prosodic or statistical information during speech segmentation has been raised.

In particular, it has been reported that, at 11 months, infants ‘choose’ suprasegmental over distributional cues when the two are conflicting (Johnson & Seidl, 2009) and a similar behavior has also been observed in adults (Shukla, Nespor & Mehler, 2007). This suggested that prosody may intervene, in adulthood as in development, as an early processing filter discarding not-well-formed word candidates. Such interpretation is coherent with the results in Saksida et al. (2017), observing an association between the rhythmical profile of a certain language and the sub-type of statistical analysis which proves to be the most effective: an early analysis of the rhythmical patterns could somewhat orient infants in applying certain analytical mechanisms instead of others.

Not to be forgotten, the set of elements supporting the accomplishment of the segmentation problem in real life ecological settings is wide, ranging from the considered phonetic-perceptual type to the contextual-psychosocial one (e.g. Yeung & Werker, 2009).

2.2 Underneath prosody: Segmental processing and perceptual attunement

Very precocious abilities have been documented at the segmental level as well.

Starting from Eimas (1971), very young infants – ranging from 1 and 6 months of age – have been reported to succeed in consonant discrimination tasks (e.g. Bertoncini et al., 1987 and Dehaene-Lambertz & Dehaene, 1994).

In 1991, Kuhl observed that 6-months-olds and adults, but not Rhesus monkeys, do display greater generalization when conditioned with the prototype of a category (the English vowel /i/) as compared with a non-prototypical instance. This led the author to extend her 'Perceptual Magnet Hypothesis'[7] to infants and to conclude not only that they display a cognitive tendency towards prototypical reasoning, but also that their speech sound representations might already have an internal structure. Yet, data exist nuancing such conclusions: toddlers' discrimination performances have been shown to depend also on physical-acoustic language-independent properties of the stimuli, e.g. Spanish infants do not succeed early in distinguishing their native [±voiced] bilabial stop contrast, while they recognize the English one, characterized by a larger VOT value (Lasky et al., 1975; cf. also Swoboda et al., 1976 and Bertoncini et al., 1988 with reference to vowels; Nazzi, 2006 for a review). Furthermore, studies conducted with older children have shown that, in terms of boundary precision, the refinement of phonological categories is a long-lasting process (Section 4).

A well-known phenomenon is indeed taking place at this stage: perceptual attunement. As classical studies have shown, while massively exposed to the input, young learners extract from the signal the sound differences which are relevant in their native system(s) and learn to discard the others, thus losing the ability to perceive not-functional subtle acoustic distinctions.

This well-documented process is largely accomplished during the first year. More precisely, infants begin to behave like adults in this respect around the 8th month of life (e.g. Werker et al., 1981; Werker & Tees, 1984; Werker and Lalonde, 1988; Kuhl et al., 1992; Best & McRoberts, 2003). Worthwhile to remark, perceptual specialization follows an analogous path in sign language acquisition, showing to be a general, multimodal cognitive mechanism (Baker-Palmer et al., 2012).

Initially described as the loss of purportedly universal discrimination capacities, the phenomenon has been reframed by subsequent research, showing that, while becoming less good in the elaboration of foreign sound categories, infants also display an enhancement in their reactions to native ones – both in behavioral and in neurofunctional terms (e.g. Werker, 1989; Kuhl

---

[7] I.e. Kuhl's fundamental description of the perceptual distortion characterizing phonological processing (much as other types of perception). Once knowledge of the native speech sounds has sufficiently progressed, humans process them not by paying attention to any physical difference, but by quickly and automatically assimilating the many different tokens they hear to prototypical categories (the 'perceptual magnets').

et al., 2006; Rivera-Gaxiola et al., 2005). Thus, today, the perceptual switch is rather seen as a developmental step marking the passage from an initial state of generalized listener to that of a neuroperceptual system committed to the sound patterns of a specific language and able to differentiate language-relevant and irrelevant information (also beyond the segmental level, cf. Kajikawa et al., 2006; Mugitani et al., 2007; Dietrich et al., 2007). Coherently with this vision, some studies have proved a correlation between the developmental rate of perceptual specialization and the quality of the input received (Elsabbagh et al., 2013), as well as with language outcomes (as assessed by means of the MacArthur-Bates questionnaire, cf. Kuhl et al., 2008).

At any rate, becoming a native listener does not mean losing any sensitivity to within-category distinctions.

In a Head-Turn Preference Procedure, McMurray & Aslin (2005) conditioned a group of English 6.5 to 9 months infants with words beginning with either /p/ or /b/; once exposed to intra-category stimuli during the test phase, participants displayed differential looking times to those items as compared to the original ones, thus demonstrating ability in detecting the difference (McMurray et al., 2002 for similar results with adult subjects). From another perspective, in a study of the Event-Related Potentials (ERPs) elicited by responses to native and non-native phonemes, Rivera-Gaxiola et al. (2005) reported that, while 11-months-olds show language-specific categorical perception as a group, sensitivity to non-native contrasts is detectable in some of them, once divided in subgroups accordingly to similarities in electrophysiological patterns.

In brief, albeit pervasive, specialization is a gradual rather than an absolute property.

When describing perceptual attunement, some finer-grained distinctions can also be considered. The process generally concerns the segmental-categorical aspect; yet, important differentiations exist at this level, namely that between vowels and consonants (or, more properly, vocoids – Vs and contoids – Cs). In this respect, studies focusing the developmental timeline of perceptual specialization suggest that it occurs earlier for Vs than for Cs (e.g. Polka & Werker, 1994) and at the latest for some kinds of Cs which are particularly demanding from the acoustic and motor point of view, i.e. fricatives and affricates (Polka et al., 2001; Tsao et al., 2006).

Some considerations can be put forward. To begin with, Vs and Cs differ significantly with respect to experience-related ontogenetic factors. Vs are much more hearable than Cs in the prenatal period and, as a consequence, the amount of exposure to the former is higher (e.g. Nishibayashi & Nazzi, 2016); moreover, Vs' production is mastered earlier than Cs' and, as it has been documented, the sensorimotor knowledge gained from production experience starts early participating to segmental perception (cf. Chap.2, Sections 3.3). Finally, the language systems considered in this review generally present a larger number of Cs as compared with Vs, so that more categorical distinctions have to be learned.

Future research is needed, to systematically examine the timeline of perceptual specialization as compared against experience-related differences of the types described.

2.3 Neural underpinnings

Overall, perceptual development in the first year of life is characterized by a fast harmonization of infants' capacities to those of adults. As described Section 1.1, a robust developmental continuity characterizes the structures and processes subserving linguistic perception since birth. Later in infancy, other strong precursors of the adult neurofunctional organization have appeared detectable through Functional Magnetic Imaging, Event-Related Potentials, Magnetoencephalography and Near-Infrared Spectroscopy.

From an anatomical point of view, Dubois and colleagues (2016) have recently shown that, around the fifth month, the dorsal and ventral pathways[8] already appear segregated and well-structured in terms of their macroscopic trajectory, asymmetry, and microstructure – an early organization that could represent the structural core allowing the fast and efficient creation of language-dedicated circuitry.

In agreement with such data, at three months of age, infants already display Mismatch Negativity Responses[9] to deviant phonological items in a stream of identical sounds; differential topographical activations to auditory-general as compared to linguistic stimuli (e.g. sinewave tones *vs* syllables) and – albeit in the context of additional activations in the right hemisphere – activity in left-lateralized regions in response to speech (e.g. Dehaene-Lambertz & Baillet, 1998; Dehaene-Lambertz, 2000; Dehaene-Lambertz et al., 2002; Dehaene-Lambertz et al., 2006a).

Finally (and coherently with newborn data, Section 1.1), studies in this field have also shown that not only perceptual areas, but also 'high-level' ones (i.e. integrative cortical structures) already display a certain degree of functionality in this period, including, among others, Broca's area (e.g. Dehaene-Lambertz, Dehaene, & Hertz-Pannier, 2002; Dehaene-Lambertz et al., 2010; Leroy et al., 2011; Mahmoudzadeh et al., 2013).

---

[8] I.e. the two macro-components of a currently leading model of language organization in the brain, first proposed by Hickok & Poeppel (2004, 2007) where, basically, the dorsal pathway supports the phonological mapping of sound and articulation and the ventral one the connection between sounds and meanings.

[9] Negative electrophysiological brainwaves which, in adult subjects, typically follow the detection of a deviant element after a stream of identical items.

In brief, even if far from the adult fine-grained, ontogenetically resulting competence, fundamental structures supporting auditory language processing appear to be operational since the first months of life.

3. Perceiving a meaningful signal: From infancy to early childhood

3.1 Recognizing familiar words

Once perceptual specialization has substantially progressed, the period going from infancy to early childhood is characterized by the mnemonic storage of linguistic elements, definitively marking the subject's entry into the socio-cognitive space of verbal communication.

In this context, coherently with the efficient use of segmentation cues displayed in the preceding months (and, clearly, as allowed by the frequent and significant associations connecting referents and references in everyday life) throughout the 11th month infants begin to display recognition of familiar words as compared to unknown ones. In visual preference procedures, they devote more attention to familiar stimuli than to unfamiliar ones (e.g. Hallé & de Boisson-Bardies, 1994, as well as all the studies considered below); in ERPs analyses, they display fast and automatic detection responses (Thierry et al., 2003; Kooijman et al., 2005 in the same direction).

A more problematic aspect is constituted by the recognition of familiar words when contrasted with phonological neighbors, i.e. real words or non-words constituting minimal pairs with the familiar one. Considering their early discrimination and categorization skills, it was initially hypothesized that infants should succeed in the task, and early studies encouraged this vision (e.g. Jusczyk & Aslin, 1995). Subsequent research, however, has yielded contrasting results and raised a long-standing discussion.

The particular relevance of this topic stems from the fact that it calls into question the degree of phonological specificity characterizing early representations: if they are fully specified, i.e. memorized on the basis of distinctive segmental contrasts, infants should be able to detect any such difference. If they cannot, then how to interpret this 'failure', in a global picture of perceptual efficiency?

Three main methodological lines have been applied to the problem. The first is based on the measurement of listening preferences when subjects hear familiar words as contrasted to (i) phonologically dissimilar and (ii) phonologically similar unknown ones (i.e. altered versions of the former, constituted by non-words based on the substitution of one segmental feature).

In 1996, Hallé & De Boysson-Bardies showed that, in such experimental conditions, French 11 months-olds do not distinguish among altered and unaltered familiar word-forms. However, such tolerance to variation did not appear unconstrained, as participants assimilate the known word to the novel when the initial C was substituted, but not when it was cancelled (i.e. they tolerated substitution, but not cancellation).

A comparable effect was highlighted by Vihman and colleagues (2004). In this case, though, English infants could not recognize familiar lexical elements if the first C (always embedded in an accented syllable) was changed, while they still displayed recognition if the substitution targeted an internal, unaccented C. Strengthening such results, Swingley (2005) reported that Dutch 11-months-olds prefer listening to onset correct pronunciations than to mispronunciations.

Taken together, these results suggest that, at the end of the first year of life and while listening to familiar words *vs* phonological neighbors, infants display consonants phonological specification provided these falls into a stressed syllable. This hypothesis was recently confirmed by Poltrock & Nazzi (2015). Capitalizing on the fact that French words are characterized by a stronger accent on the final syllable, the authors compared French 11-months-olds' orientation times to well-pronounced familiar words *vs* non-words built upon the substitution of the onset-consonant in the final syllable. In such conditions, French participants paralleled English and Dutch subjects and oriented more towards the known *vs* the unknown word.

However, electrophysiological evidence failed to detect significant differences in the ERP signature to the same kind of stimuli at this early stage: in a passive listening paradigm, the evoked potentials to familiar words containing altered *vs* unaltered initial Cs differed for 20-months-old English infants', but not for 14-months-olds (Mills et al., 2004).

Beyond mere lexical-acoustic elaboration, a second line of investigation entails object/label matching procedures in which infants are tested for the association between a (verified) known word and its referent while they are presented together with physical and/or phonological distractors.

The results obtained with similar procedures have globally argued in favor of detailed knowledge of stressed Cs starting around the 14th month (Bailey & Plunkett, 2002; Fennel & Werker, 2003; Swingley & Aslin; 2000; 2002; Swingley, 2003).

Again, though, evidence is not simply unidirectional. In this connection, Swingley (2016) reported that, when a novel and a familiar object are simultaneously shown to 24-months-olds and coupled with a novel phonological neighbor, albeit detecting the phonological difference that underlies the task (as measured *via* eye-tracking), the participants seem more prone to associate the novel phonological neighbor with the familiar rather than with the novel object.

Such behavior seems to point to the relevance of considering the interaction between perceptual mechanisms and cognitive strategies.

A well-known interpretation in this sense has been proposed by Stager & Werker (1997), who underlined that, at the onset of lexical learning, linking a label with an object might be a cognitively demanding process, asking for a considerable amount of attentional resources which infants may subtract – among other things – to fine-grained phonological processing. In ontogenetic terms, giving the priority to the computation of the label/entity association could be advantageous, by reducing the information load and allowing for massive reference/referent memorization. This interpretation is supported by studies highlighting that phonological neighbors are rare in early receptive vocabularies (e.g. Caselli et al., 1995; Charles-Luce & Luce, 1995) and that bilingual subjects, faced with two lexical inputs, seem to rely longer on this possibly compensatory mechanism (until the 20th month of life, cf. Fennel et al., 2007).

In this line of thought, out-and-out segmental categories emerge through a partial, possibly long overlap between lexical and phonological acquisition. In the words of Yoshida et al. (2009): «younger infants may have accumulated enough experience with familiar words such as *ball* to know that it is never pronounced *doll*, but this specific knowledge has yet to be generalized to all b–d contrasts» (page 416). By storing more and more elements that differ in fine-grained characteristics, learners would progressively be 'forced' to take them into account, thus setting adult-like phonological boundaries.

The third research direction applied to the problem of phonological specificity in early word forms is centered upon a more complex activity: the simultaneous learning of two novel lexical neighbors.

Initiated by Stager & Werker (1997), results in this field displayed that successful completion of the task could be obtained starting from the 17th - 20th month of life (e.g. Werker et al., 2002; Fennell et al., 2007). Later investigations, however, have underlined the possibly high cognitive costs entailed by the methodology characterizing some of those studies (more precisely, on the 'Switch Task') and have shown that toddlers may demonstrate earlier the ability to learn two similar-sounding labels, provided a less demanding setting (Yoshida et al., 2009; Fennell & Waxman, 2011) or other facilitating elements such as cues of semantic and syntactic nature (Dautriche et al., 2015).

Overall, the disentanglement of perceptual and general learning processes constitutes a priority issue in connection with the degree of phonological specificity in older infants' representations.

To conclude, the extreme precocity and efficacy characterizing perceptual auditory abilities in infancy do not imply that, by the end of the first year, phonological categories are specified entirely and in any phonetic context; nor that infants and toddlers are fully able to apply their perceptual resources in challenging cognitive situations.

3.2 Vowels, consonants and their specificity

This said, research has also more specifically focused on the function of early segmental representations in linguistic development, showing that the fundamental division between the two main phonemic classes, vocoids and contoids, is not negligible in this context.

Investigations in this field have started from the hypothesis that Cs and Vs might have different functional roles in language processing and acquisition (cf. Nespor & Peña, 2003): Vs would carry prosodic information and everything that can be inferred on this basis, Cs would allow for word distinction (and, therefore, identification and memorization).

In support of this view, several studies based on word-learning and recognition have shown that Cs play a primary role in lexical acquisition, displaying – to put it in this way – that, when successfully distinguishing among phonological neighbors, infants and young children rely on Cs and not on Vs (e.g. Nazzi, 2005; Nazzi & New, 2007; Havy & Nazzi, 2009; Zesiger & Jöhr, 2011; Hochmann et al., 2011; Poltrock & Nazzi, 2015).

Current research points out that such ‘consonant advantage’ (or ‘consonant bias’) is displayed quite early, since the 8th month of life (cf. Nishibayashi & Nazzi, 2016). However, it also highlights an important aspect: this behavior is not cross-linguistically uniform.

While a robust bias for Cs has been observed in toddlers learning linguistic systems like French and Italian, in which the consonantal repertoire is more complex (and, thus, functionally more informative) than the vowel one, learners of languages that differ in this respect revealed that the phenomenon is not universal. In particular, studies conducted in English, which is characterized by a finer-grained vowel system, and Danish, where Vs are more than Cs (and, in addition, the latter undergo extensive lenition), have detected very different patterns. English infants seem to display an essentially equal sensitivity to Cs and Vs (Mani & Plunkett, 2007; 2010; Floccia et al., 2014), whereas Danish infants a reverse bias for Vs (Højen & Nazzi, 2016).

Such results implicate a connection between the C/V biases and language-specific experience, which could stem from the different distribution of the two categories in lexical and/or phonological terms (e.g. number of Vs and Cs, phonetic reduction phenomena, degree of internal complexity of the vowel and consonant systems). Very relevant in this sense, recent results highlighted that French infants seem to display a passage from an early V- to a later developed C-bias between the 6th and 8th month of life (cf. Bouchon et al., 2015; Nishibayashi & Nazzi, 2016). A possible lexical origin for this phenomenon seems rather improbable, given the negligible size of 6-to-8-months-olds' vocabulary (even if a possible relationship between the preference for Cs and individual vocabularies could still be further evaluated). Rather, they seemingly point to the relevance that experience with the two different sound classes could have: prenatal exposure to vowels could bias infants towards relying upon these sounds at the beginning of acquisition; subsequently, the massive presence of consonants in the input could gradually capture their attention and lead to develop an adult-like priority for Cs (cf. Nishibayashi & Nazzi, 2016).

Heretofore, it seems that, during infancy, linguistic experience could shape the processing and representation of the different classes of sounds through different possible paths, which remain to be completely clarified. A pilot experiment aiming at contributing to this research direction is described in Appendix I.

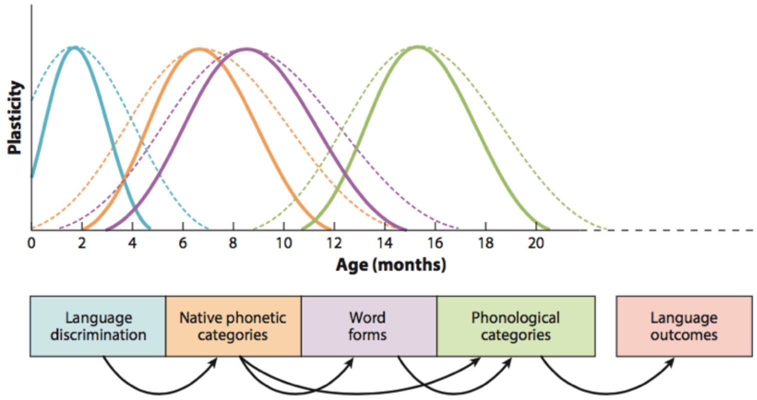
3.3 From birth to the second year in a picture

In order to conclude and sum up this overview from birth to the second year of life, the notion of critical periods turns out useful.

Definable as «windows, typically in early development, during which a [neurofunctional] system is open to structuring or restructuring on the basis of input from the environment» (Werker & Hensch, 2015, page 175), critical periods CPs lie at the foundations of our understanding of language acquisition's biology. Vastly investigated in their dynamics, they are currently explained as the result of the action of molecular agents triggering (and, later, inhibiting) plasticity in specific brain regions and whose action, in turn, is elicited by the input.

As shown in Image 1, summarizing the developmental stages until now described, the critical periods inherently participate to the cascading dynamics of linguistic development, where each ontogenetic achievement lays the foundations for the subsequent. This aspect is critically relevant to the topics introduced by Chapter 3, i.e. developmental speech disorders and their reflection on perceptual ontogenesis.

Figure 1 – Critical periods in perceptual-phonological acquisition (Werker & Hensch, 2015)



4. Childhood: The long tail of perceptual development

4.1 Age-dependent differences: A long-standing phenomenon

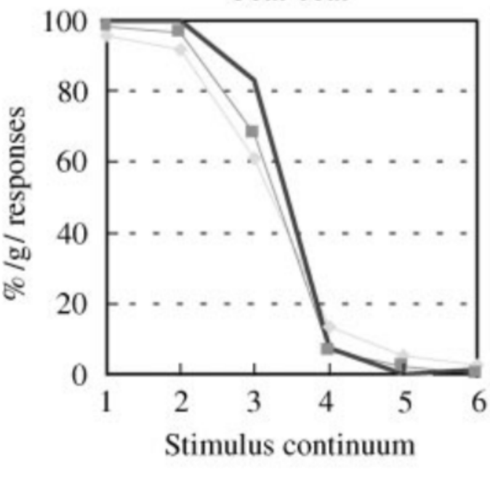
Available evidence converges in showing that the refinement of language-specific categorical perception is a long-term process, extending well into childhood years: the infants' striking progression towards phonological specialization do not result into an acquisitional 'final state' (cf. Zevin, 2012 for an overview).

In the footstep of what observed about phonological specification during the transition from infancy to childhood, a preponderant issue in childhood studies is segmental representation: when examined in depth, the latter proves to be immature until adolescence.

Such protracted development has been documented since the 1970s, the period marking the beginning of modern investigations on perception. In particular, the first studies highlighted long-lasting age-dependent differences in connection with discrimination and identification tasks over synthetized phonetic continua (cf. Zlatin & Koenigsknecht, 1975; Simon & Fourcin, 1978 and, later, Burnham et al., 1991; Slawinski & Fitzgerald, 1998). The identification task, as structured on phonetic continua ranging between two categories, offers a classical demonstration of categorical specialization in adults, who tend not to perceive the intra-phonemic ('cognitively irrelevant') distinctions and, as a consequence, to obtain very steep identification functions (Id *f*, i.e. the function given by the relationship between the identification answers and the stimuli placed on the two sides of the continuum). The more dichotomic the distribution of the answers, the steeper the function, hence the usage of the slope of the Id *f* to assess boundary precision in perception.

As Figure 2 displays (where adults' answers are marked in black and children's answers in grey), this steepness is not fully observed until the end of childhood.

Figure 2 – Example of identification function (Hazan & Barrett, 2000, p. 388).



Hazan & Barrett (2000) focused in detail the developmental timeline of the phenomenon, as measured over various acoustic continua (/g/-/k/; /d/-/g/; /s/-/z/; /s/-/ʃ/). In this connection, the authors showed a progression towards adult-like identification patterns between 6 and 12 years of age. 12-years-olds, tough, still appeared less consistent than adults in their mean judgements.
In another direction, Medina et al. (2010) pointed to the fact that precision in boundaries location and categorical perception are two distinct abilities: the former assessable through the steepness of the identification function, the latter given by the match between mean discrimination and identification answers. In their results, categorical perception did not vary developmentally after 9 years of age (and, interestingly, was not perfect throughout ages). Rather, the progression detected pertained to precision in boundaries location, suggesting that the late developmental tail could be a matter of prototype consolidation rather than of categorization ability.

At any rate, processing differences between adults and children have also emerged *via* neurofunctional investigations. In an ERPs study based on automatic processing (participants were instructed to watch silent movies and to try to ignore the sound stimuli), Liu et al. (2014) reported immaturity in the electrophysiological responses to phonemic contrasts (affricates) in participants up to 8,5 years old.

Investigating the reasons behind these tendencies, it has long been underlined that children's usage of the phonetic auditory cues signaling phonological contrasts may diverge from that of adults (e.g. Nittrouer & Studdert-Kennedy, 1987; Sussman & Carney, 1989; Ohde et al., 1995; Holden-Pitt et al., 1995; Nittrouer & Miller, 1997; Sussman, 2001; Nittrouer, 2002). Specifically, Nittrouer & Miller (1997) hypothesized a 'Developmental Weighting Shift': adults would be able to differently weight phonetic cues as a function of the context in which they appear (i.e. by relying on the informativeness of the cue over the specific phonetic environment), while children could be less effective in this sense due to lesser knowledge of such patterns.

A reference to experience with language is here made, and this is in agreement with the results of a quite recent fMRI investigation. Conant et al. (2014) examined the activation patterns corresponding to a discrimination task in 7-to-12-years-olds and reported that the most skilled participants displayed a more significant involvement of the Posterior Middle Temporal Gyrus: a pattern which, in adults, is observed when processing sound categories learned through a former familiarization phase rather than native phonemes. In a nutshell, children could elaborate native contrasts by relying on mechanisms that adults deploy for not-overlearned sounds, thus displaying a lesser entrenchment of the native phonological categories.

Coherently with this perspective, some studies have highlighted correlations among phonological boundaries precision, lexical growth (e.g. Metsala & Walley, 1998; Walley, 2003) and literacy skills (e.g. Mayo et al., 2003; Conant et al., 2014). Overall, such investigations argue for substantial connections between the abilities considered, but their causal directions have yet to be fully clarified. Furthermore, as it will be described, other factors could be at play in this refinement process; notably, for what pertains to this dissertation, the attainment of sensorimotor knowledge through experience with production (Chap. 2, Section 4.3; Chap. 3, Section 1 and 2.3; Chap. 5, Section 1).

4.2 Speech processing in challenging conditions

Protracted age-related differences are all the more observed when the listening condition is non-optimal.

Re-examining Hazan & Barrett (2000), albeit the authors did not use signal degradation, their experiments consisted of conditions in which participants had to detect a phonological contrast by means of phonetic cues varying in their reliability and informativeness. Younger subjects displayed less categorization consistency in the latter situation, a behavior which points to a lack of flexibility in non-optimal listening conditions.

Coherently with this latter point (and in agreement with former literature – e.g. Neuman & Hochberg, 1983; Elliott, 1979), Johnson (2000) reported a very late maturation of perception skills in connection with substantially degraded stimuli. Participants ranging from 6 to 15 years of age were compared with adults in their ability to identify consonants and vowels embedded in CVCV non-sense structures, while listening in: (a) optimal conditions; (b) reverberation; (c) noise (multi-talker bubble), and (d) reverberation plus noise. The emerging maturational patterns were protracted and diversified, depending on both listening condition and stimuli: adult-like performance with vowels was reached well earlier than with consonants (the first being detectable already by the youngest participants, the second only by the oldest); while identification of consonants was successful at 14-15 years for the first three conditions, but still not equally proficient, at the same age, for the last and particularly challenging one (i.e. reverberation plus noise)[10].

Interpreting children's difficulty to deal with noise, the conceivable incidence of general cognitive factors such as the ontogenetic reduction in neuronal noise levels and maturation of the

---

[10] The superimposition of noise to the signal is not the only type of manipulation used to study the limits of children's auditory capacity. Two widely used alternative methods are represented by the manipulation of the speech temporal and spectral components. Albeit not directly addressed by this dissertation, such techniques have added important ontogenetic evidence (e.g. Dorman et al., 1998; Eisenberg et al., 2000; Bertoncini, Serniclaes & Lorenzi, 2009).

executive functions have been repeatedly hypothesized (e.g. Jones et al., 2015; Thompson et al., 2017) and systematic analyses in this direction will have to be conducted.

To sum up, speech perception during childhood is characterized by a residual categorical-perceptual immaturity and this aspect strongly emerges in noisy conditions. In this context, two main directions of inquiry remain to be clarified, i.e.:

(i) possible age-related differences in representational *content*, as given by linguistic experience of different types;

(ii) the impact of representational *processes*, i.e. the wide set of cognitive abilities influencing children's categorization skills.

This dissertation addresses the former of the two aspects. More specifically, the study presented in Chapter 4 investigates the degree of support that a specific source of phonological information can provide to speech-in-noise perception during childhood, i.e. sensorimotor knowledge. The literature reviewed in the next chapter will justify such perspective.

2. DEVELOPING SPEECH PERCEPTION – PART 2:
MULTISENSORIALITY

Foreword: On eyes that listen while ears see. Considerations over the nature of the speech signal

«This is not just the innocuous and obvious claim that we need a body to reason»
(Lakoff & Johnson, 1999, p.4)

Fundamental as they are, auditory perceptual skills are not everything that has to be considered about speech perception and its development. To introduce this topic, some considerations over the nature of the linguistic signal appear worthwhile.

In physical terms, speech is a multisensory phenomenon, made up of diversified and integrated information: a sequence of articulatory movements yielding auditory and visual consequences. Be this act considered under the production or the perception point of view, the sensory space where it takes place lies at the intersection of different processing modalities, namely the auditory, the visual and the sensorimotor one[11].

Yet, interestingly enough, not-technical definitions make exclusive reference to the dimension of sound (e.g. 'to speak': 'to use your voice to say something'[12]; 'word': 'a speech sound or series of speech sounds that symbolizes and communicates a meaning'[13]): in the folk concept, words flow carried by the voice and reach the brain of the listener through the ear, with no need for further psychophysical involvements.

One of the sources of this underestimation of multisensoriality is probably constituted by the fact that, for communicative purposes, sounds constitute particularly robust and prominent signals. As Corballis well conveys: «It is not only a question of being able to communicate at night. We can also speak to people when objects intervene and you can't see them, as when you speak to your friend in another room. All this has to do, of course, with the nature of sound itself, which travels equally well in the dark as in the light and wiggles its way around obstacles. A wall between you and the base drummer next door may attenuate the sound but does not completely block it. Vision, on the other hand, depends on light reflected from an external source, such as the sun, and is

---

[11] As to this latter aspect, speech production entails the planning and execution of motor commands, which are cognitively coupled with the sensory-proprioceptive correlates that they generate in the speaker; on the other side, this information is recruited by the listener during speech perception. Section 3 will offer a detailed description of this aspect.

[12] 'To speak', Cambridge English Dictionary – online version.

[13] 'Word', Merriam-Webster English Dictionary – online version.

therefore ineffective when no such source is available (…) the light reflected from the surface of an object to your eye travels in rigidly straight lines, which means that it (…) is susceptible to occlusion and interference (...)»[14].

Through the many daily-life advantages offered by the auditory channel (sufficient to consider, the telephone), sounds become the more remarkable constituents of the signal: the well noticeable tip of the iceberg of a more complex phenomenon.

Indeed, this perspective is also connected with a major issue in our culture, i.e. its long tradition of dis-embodiment of the cognitive attributes and abilities. This theoretical stand has been the subject of a long scientific debate which, most lively in the last decades, has deeply changed the scientific – if not the common-sense – understanding of this relationship (e.g. Dove, 2016).

As regards language research, some decades have passed since the detection of various phenomena who encouraged to rethink the exclusive primacy of audition. To cite two milestone references: (a) the discovery that, in psychoacoustic terms, actual linguistic signals are not made up by a linear sequence of well-recognizable phonemes, but rather by a sequence of coarticulated sounds (Lieberman et al., 1967), and (b) that audiovisual integration is automatic, extremely rapid and involuntary, a process in which information coming by both sources shapes the percept (McGurk & MacDonald, 1976)[15].

By means of a description of the multisensory dynamics involved in speech perception and in its development, the goal of this chapter is to complete the ontogenetic picture and present some of the core topics of this thesis.

If linguistic perception is a complex cognitive construction, the object of this dissertation can be described as the development of such a construction.

---

[14] Corballis (2002), page 189.

[15] Cf. Rosenblum (2005) for a review treating the passage from the audition-centered vision of speech perception to the current multimodal understanding of the phenomenon.

1. The discovery of the talking face: Newborns

«Les parfums, les couleurs et les sons se répondent
dans une ténébreuse et profonde unité»
Charles Baudelaire, Correspondances

Multisensoriality is inherent to human cognition. As they come to life, newborns find themselves plunged into a reality where objects and events elicit the simultaneous reaction of different receptors.

Sights, sounds, tastes, smells and tactile sensations coexist in unitary entities, but sensory redundancy does not generate cognitive chaos. On the contrary, it supports the apprehension of invariance. From the examples provided by Bharick, Lickliter & Flom (2004): «the sights and sounds of a bouncing ball are synchronous, originate in the same location, and share a common rate, rhythm, and intensity pattern», «the face and voice of a person speaking share temporal synchrony, rhythm, tempo, and changing intensity»[16]. The repetition of this information allows the learner to discover patterns of recurrent association between modality-specific information (e.g. if the ball bounces higher, it also sounds louder), thus obtaining an holistic understanding of objects and events in the world.

Moreover (and easily imaginable by assuming an infant's perspective), the multisensorial stimuli populating the environment are not given in an unordered fashion during development.

In this connection, Fausey, Jayaraman & Smith (2016) have recently highlighted that the amount of exposure to faces and hands in the visual field is not equal during the first two years of life. Rather, the former prevail during an earlier period and their presence declines in time, while the latter follow the reverse path.

Indeed, in the first months of life, conspecific faces occupy a privileged role: faces that often talk.

---

[16] Bharick, Lickliter & Flom (2004), pp. 99-100.

Figure 3 - Detail from Giotto’s Natività, Cappella degli Scrovegni

At birth, and despite a weak visual system, humans appear phylogenetically attracted by faces[17] and, when those speak, they are better able to recognize them (e.g. Guellaï, Coulon & Streri 2011; Guellaï, Mersad & Streri, 2015). Indeed, the contextual presence of the voice helps them even in the visual recognition of the mother (cf. Sai, 2005).

Obviously enough, tough, the fact of being interested in something does not necessarily imply the ability to understand it.

The available evidence concerning audiovisual speech processing in the first moments of life agrees in suggesting that some foundational skills are active since the very beginning. Aldridge and colleagues (1999) showed that, a few hours after birth, newborns prefer looking at faces coherently articulating a vowel than to mismatching combinations, which is the sign of a precocious sensitivity to the multimodal association. Quite remarkably, thought, the combinations used in the task were characterized very sharp mismatches (e.g. face with rounded lips + [a] sound; face with open jaw + [u] sound). Furthermore, it later appeared that this precocious ability is not limited to conspecific stimuli, but also applicable to primates’ vocalizations, narrowing progressively during the second semester of life (Lewkowicz, Leo & Simion, 2010; Lewkowicz & Ghanzafar, 2006).

In synthesis, coherently with the whole picture of cognitive and linguistic development, genetically-encoded abilities to process multimodal speech signals are immediately available to newborns, but their functioning is broader and coarser-grained in comparison with the adult benchmark.

Given these premises, the precocious detection of the correspondence between sights and sounds has been attributed to the analysis of basic and macroscopic characteristics (such as the analogy in intensity and temporal synchrony connecting two perceptual events), rather than to an observation of the correlation between articulatory movements and the corresponding sounds: very

---

[17] E.g. Goren, Sarty, & Wu (1975).

young infants would perceive linguistic multisensoriality on the basis of domain-general information (e.g. Lewkowicz, Leo & Simion, 2010; Lewkowicz, 1996a; Murray et al., 2016).

At any rate, newborns are not exclusively able to cope with very simple stimuli such as vowels or vocalizations. As recently reported by Guellaï et al. (2016), their preference for coherent multimodal combinations holds with full sentences as well. Importantly, the stimuli used in this study were recorded in an infant-directed style, i.e. with both auditorily and visually accentuated prosody. In this respect, this result complements previous research stressing newborns' sensitivity to basic features of the signal (synchrony, intensity and the like), by pointing at the fact that those elements are harmonically encapsulated in a sort of holistic percept: the suprasegmental signal, flowing parallel to the segmental one in both the auditory and the visual dimension[18].

Suprasegmental features, in a nutshell, seem to represent a perceptual primitive which supports the inexpert learner in decomposing the stream, possibly independently from the modality of perception.

2. Geared to integrate? Audiovisual skills in the first year of life

2.1 The face-sound association

In continuity with what observed in newborns, infants prefer looking to correct (as opposed to mismatching) audiovisual combinations from the second month of life.

Familiarized with two silently articulating faces and tested while hearing the auditory counterpart of only one of them, participants look more at the correct than at the incorrect pairing. Such ability is not influenced by the gender of the speaker and is not weakened by ecologically valid stimuli (i.e. full heads including hair, neck and shoulders – cf. Patterson & Werker 2003 and 1999; Kuhl & Meltzoff, 1982; Walton & Bower, 1993). Accordingly, Bristow and colleagues (2009) described evidence of a precocious Mismatch Detection Response in 2-to-3-months-olds perceiving vowels as followed by a silently articulating, incongruent face.

However, the degree of refinement of these skills is not an undisputed topic.

Lewkowicz and colleagues (2015) evaluated 4-to-14-months-olds subjects through familiarization with two silent monologues, followed by a test phase in which the two were still visible but of only one of them audible: participants succeeded in the task only from the 12th month.

Such results support the authors' view (Section 1, pages 22-23 – e.g. also Hollick, Newman & Jusczyk, 2005 and Kubicek et al., 2014, in the same direction): until approximately the first birthday, infants would not be able to exploit articulatory information and would rely on more

---

[18] Cf. Kitamura, Guellaï & Kim, 2014 and Lewkowicz, 1998 for evidence of visual prosody detection in older infants.

general attributes connecting visual and auditory percepts. The latter would become less important as acquisition advances, then acting as facilitatory rather than fundamental elements. However, the complexity of the stimuli and of the experimental design in Lewkowicz et al. (2015) call into question the full extendibility of these results. Specifically, infants were habituated with silent visual articulations of long speech passages[19], which is the most difficult linguistic-perceptual condition across the life span (e.g. Taitelbaum-Swead & Fostick, 2016).

More generally, the great reliance upon domain-general *vs* linguistic-articulatory visual information that infants display should not be understood in absolute terms since, if so, it would not be consistent with data concerning perceptual specialization and statistical learning in the audiovisual domain (cf. Section 2.4), nor with evidence displaying early sensorimotor perceptual contributions (cf. Section 3.3). While not denying this above perspective, the latter directions of research document a more graded cognitive progression, in which the elaboration of speech movements can be detected earlier, depending on the task proposed and on the aspects considered.

2.2 Early integration abilities

The above reviewed ability, i.e. audiovisual matching, pertains to the process of associating mouth movements and sounds: it is not equivalent to the integration of the two components in a unique percept, modulated by concomitant information.

In exploring this domain, research has typically capitalized on the McGurk effect. Allowing for the creation of both fusional percepts (e.g. visual [ga] combined with auditory [ba], which yields perceptions such as [da]) and mismatching percepts (e.g. visual [ba] combined with articulatory [ga], producing 'odd' perceptions such as [bga]), this well-known phenomenon constitutes a manageable instrument to assess the presence of both automatic integration (as in the former case) and incongruency detection (as in the latter).

In this connection, 4.5-to-5-months-olds have been reported to display integration in behavioral procedures (habituation technique, cf. Rosenblum et al., 1997; Burnham & Dodd, 2004) and electrophysiological experiments, where differential responses follow audiovisually inconsistent stimuli that either can or cannot be fused into a single percept (Kushnerenko et al., 2008).

Nevertheless, the robustness of precocious multimodal integration has been challenged by a series of experiments led by Desjardins & Werker (2004) which, characterized by the repetition of

---

[19] E.g. from the scripts used in Lewkowicz et al. (2015), page 153: 'English monologue 1': «Good morning! Get up! Come on now, if you get up right away we'll have an hour to putter around. I love these long mornings, don't you? I wish they could last all day. Well at least it's Friday».

the task with different audiovisual combinations, has evidenced detectable but unstable patterns at this early stage. The hypothesis of an early full maturation of automatic integration is also inherently contradicted by studies targeting the developmental trajectory of this aspect from childhood to adulthood (and since the foundational paper of McGurk & MacDonald, 1976, cf. Section 4).

Ultimately, these contrasting findings may fit together, once a developmental point of view is assumed. In this perspective, inconsistent findings could not be surprising, but rather the natural reflex of an ongoing specialization; the expression of a transient developmental state in which both inter-individual and intra-individual variability is still very high (tendencies that future research will have to clarify both causally and temporally).

2.3 Neural underpinnings

Irrespectively of the cognitive mechanisms triggering the process (i.e. be these limited to domain-general integration or encompassing some analysis of articulatory movements), electrophysiological research addressing multimodal perception in infancy reveals that, from the second month, audiovisual binding occurs early in perception – at least for isolated vowels. Moreover, the sources of the signal appear located in the left hemisphere, while non-linguistic audiovisual features such as gender or other physical attributes of the speaker elicit right activations (cf. Bristow et al., 2009). Coarse-grained as it may be, at this stage a differentiation seems already to exist between the multimodal elaboration of language as opposed to other entities.

More in detail, by means of functional Near-Infrared Spectroscopy (fNIRS), Altvater-Mackensen & Grossmann (2016) observed that the administration of matching as opposed to mismatching audiovisual vowels recruits the inferior-frontal regions in 5.5-to-6-months-olds. As it will be further described (cf. Section 3.4)[20], such topography is fundamental in the adult coupling of multisensory information and its involvement in development has been repeatedly documented. Worthwhile to underline, the involvement of those areas is normally enhanced in adults when processing conflicting audiovisual stimuli (Ojanen et al., 2005) while, intriguingly, a reverse enhancement is observed by Altvater-Mackensen & Grossman. This different reaction can be seen as evidence of the fact that participants are tested during a period of intense learning about audiovisual integration, so that the task attracting more cognitive resources is the normal audiovisual pairing for them. Consistently, this pattern also appeared stronger in infants who focalize more on

---

[20] The neuromorphology of early multimodal perception will be described in Section 3.4, adding important information to this ontogenetic picture.

the mouth than on other facial regions while looking at the stimuli: a phenomenon which, as described in the next session, plays a key role in audiovisual perceptual specialization.

2.4 Audiovisual perceptual specialization

Perceptual narrowing is not a unique feature of auditory processing (cf. what previously remarked with reference to sign language acquisition, Chap. 1, page 7); nor is it unique of language. Rather, environment-specific specialization is a large developmental phenomenon through which infants adapt to their social groups (cf. Pascalis, de Haan & Nelson, 2002).

It takes place in face and language processing with interesting commonalities: as human infants lose efficacy in processing foreign phonological contrasts, they experience growing difficulty in the facial discrimination of human ethnic groups they do not see in their environment (Walker & Tanaka, 2003); as their perception of primates' vocalizations declines, they become worst at discriminating primates' faces (Pascalis, de Haan & Nelson, 2002).

Coherently, the same is true in the audiovisual domain, where the ability to match primates' mouths to vocalizations is progressively lost in favor of species-specific attunement (Lewkowicz & Ghanzafar, 2006).

However, linguistic exposure also affects multisensory speech perception in more specific ways. In this connection, the discrimination of two different languages on visual-only bases is only proficient in native speakers of those languages (Weikum et al., 2007 and 2013), highlighting «infants' selectivity for retaining only necessary perceptual sensitivities»[21]. Similarly, Spanish learners stop matching English syllables that are homophonous in their mother-tongue during the second semester of life (/ba/ and /va/, Pons et al., 2009; Kubicek et al., 2014). Thus, visual/sensorimotor information proves to contribute in linguistic-perceptual narrowing alongside auditory one. Indeed, as for the auditory modality, the degree of audiovisual specialization in infants appears correlated to both the input's quality (the mother's interactive behavior and degree of 'infant-directedness') and later language outcomes (receptive and expressive vocabulary size, e.g. Altvater-Mackensen & Grossmann, 2015).

For what pertains to the mechanisms at play in this transition, as mentioned at the end of the preceding section, a very relevant aspect is constituted by infants' fixation patterns.

A passage from an eyes-centered looking tendency to an increasingly mouth-centered one is observed among the 4th and the 8th month in monolingual infants (cf. Lewkowicz & Hansen-Tift, 2012) and, later, it appears accentuated when processing a foreign language. On these bases, mouth-

---

[21] Weikum et al., 2007, page 1159.

looking could constitute a cognitive strategy supporting the extraction of phonological knowledge through the observation of speech movements. Interestingly, as to Pons, Bosch & Lewkowicz (2015), bilingual infants display mouth-centered looking patterns earlier and longer in development. Such behavior is interpretable as a confirmation of the relevant role played by mouth-observation: due to the additional need to construe two different phonological systems, bilingual learners would devote more attention to articulatory movements than monolingual ones.

If this perspective is correct, mouth-lookers should be subjects who are engaged in such learning process at the moment of the observation. Conversely, infants who do not exhibit a preference for the mouth could be either more advanced, free to leave their gaze capture other information (such as the speaker's attitude, as understandable by observing the eyes' and the head's movements) or, on the contrary, less committed individuals[22]. Hence, the amount of mouth-looking time should act as a predictor of efficiency in multimodal processing. This vision is supported by findings documenting, on the one side, a negative correlation between mouth-looking time and the amplitude of electrophysiological responses to incongruent stimuli (6-to-9-months-olds who look less at the mouth display more intense neurofunctional elaboration); on the other, a positive correlation connecting the phenomenon with later language outcomes (receptive and expressive vocabulary size, Kushnerenko et al., 2013a and 2013b).

The instructive role of the talking mouth is also demonstrated by studies documenting the existence of phonological statistical learning in the visual/sensorimotor modality.

In this connection, Teinonen and colleagues (2008) exposed 6-months-old participants to /ba/-to-/ga/ auditory continua as differentially coupled with visual information: bimodal in one case (auditory syllables were coupled with the coherent articulatory gesture according to their position, relatively to the midpoint of the continuum); unimodal in the other case (tokens were paired with either visual /ba/ or visual /ga/). As a result, subjects in the first condition appeared to discriminate the two percepts, while the others did not: the influence of visual information had led them to the formation of either a bimodal or unimodal category.

Comparable results were later produced by Ter Schure, Junge & Boersma (2016), with the important addition that infants in the bimodal distribution looked more at the mouth than infants in the unimodal one.

---

[22] As an extreme example, it has long been known that children with disorders of the autism spectrum display a deficit in speech multimodal integration, and the atypically low engagement towards the other's gaze is a fundamental feature of those pathologies.

The detection of visual effects over statistical category learning at 6 months testifies to an initial exploitation of articulation analysis before the first birthday (at least in the observed conditions).

Overall, while not playing a decisive role soon after birth, articulation seemingly becomes gradually more exploited throughout the first year, when the observation of the other's mouth yields significant contributions to specialization.

However, this is not to maintain that infants learn to perceive articulatory movements through simple observation: a more articulated and complex process underlies this aspect, crucially involving the joint development of one's own production abilities. After an excursus detailing historical and current hypothesis regarding the role played by sensorimotor information in adult speech perception, Section 3.3 will address this subject from a developmental perspective and mark the passage to the core-topics of this thesis.

3. Neuropsychological mechanisms underlying multisensory speech perception

3.1 Excursus: The mechanics of multisensoriality in speech perception

«The limits of language (of that language which alone I understand) mean the limits of my world»

Ludwig Wittgenstein, Tractatus Logico-Philosophicus

Due to its psychological and sociocultural complexity, language has long been regarded as a primary, excellent example of human cognition's possibilities. On the one side, the attribution of this major anthropological value has guaranteed the topic a central place in the history of human thought. On the other, it has contributed to the development of a vision of language as an entity of its own nature, with its most accomplished expression in the Chomskyian theoretical apparatus.

Through all its numerous revisions and progresses, the generativist approach has nurtured the idea of a differentiation between a faculty of language 'in the narrow sense' as opposed to 'a faculty of language in the broad sense': the former concerning the organization of the linguistic components into a syntax, constituted by complex hierarchical and abstract computations that only apply to human communication; the latter entailing sensorimotor and conceptual-intentional abilities, also shared by other animals (cf. Hauser, Chomsky & Fitch, 2002). An important by-product of this vision is that 'narrow' linguistic skills cover a leading role: they represent the real, uniquely human essence of language, while the 'broad' ones (while not denying their necessity) appear relegated to a supporting function. In this sense, this system of thought has contributed to the underevaluation of the role played by sensory processing in the linguistic dimension and, crucially for this thesis, in language acquisition.

This theoretical standpoint can be contextualized in the already remembered cultural climate committed to the general disembodiment of the cognitive sphere (of whom it represents an outstanding expression). Paradoxically enough, one of the first criticisms to this general ‘disembodied’ view came from the study of language itself, through the proposition of a Motor Theory of Speech Perception (Lieberman et al., 1967; Lieberman & Mattingly, 1985).

The Motor Theory came as a first answer to the factual evidence that the auditory signal *per se* is not invariant. As is the case of any first attempt, its answers to the problem proved to be simplistic: «the motor gestures are not the means to sounds (…) rather, they are, in themselves, the essential phonetic units. On the other side, sounds are not the true objects of perception (…) rather, they only supply the information for immediate perception of the gestures» (Liberman & Mattingly, 1985, page 31). Its proposition raised diffuse criticism and easy objections, among which studies highlighting the fact that brain damages affecting areas traditionally associated with speech production do not lead to perceptual inability.

This notwithstanding, subsequent scientific progress reopened the issue. Most notably, the discovery of populations of neurons in the primate brain (area F5, thought to correspond to the human ventral premotor cortex) that display a ‘mirroring action’, i.e. that fire both when an action is produced and when it is observed (cf. Di Pellegrino and colleagues, 1992).

Because speech production can be seen as a particular form of action, one of the hypothesis following this discovery was that the mirror system could be involved in its perception.

Invasive experimental procedures allowed to demonstrate this phenomenon *via* the examination of the primates’ production and perception of conspecifics’ vocal calls and mouthings (Ferrari et al., 2003; cf. Miller et al., 2010 and 2015 for more recent results in an analogous direction)[23]. Substantive support also came from experiments with human subjects based on the Transcranial Magnetic Stimulation Technique. With this methodology, some studies showed that the perception of speech stimuli crucially involving tongue movements evokes an increase in the electrophysiological potentials coming from the tongue muscles (Fadiga et al., 2002) and, more decisively, that a stimulation of the motor areas controlling certain articulators can temporarily disrupt the processing of sounds that are produced with the same[24] (while sparing the elaboration of

---

[23] Not limited to primates, findings about the developmental interconnection between perception and production have also been reported about birds, e.g. Nottebohm et al. (1990), Prather, Okanoya & Bolhuis (2017).

[24] Transcranial Magnetic Stimulation (TMS) is a methodology allowing the induction of electric activity in targeted micro-areas of the brain. By this means, it is possible to interfere with the spontaneous electrophysiologic response to a stimulus and observe if variations are detected, under such conditions, in its elaboration.

stimuli that are not, cf. D'Ausilio et al., 2009 and 2012; Möttönen & Watkins, 2009; Rogers et al., 2014).

Similar results initially seemed again to nurture a strong sensorimotor hypothesis about speech perception; however, important counterevidence came once more from the analysis of brain damages affecting areas associated with speech motor knowledge, which demonstrate the possibility of a successful functional dissociation between production and perception – at least in adult subjects and in optimal listening conditions (e.g. Lotto, Hickok & Holt, 2009; Hickok, 2010; Rogalsky et al., 2011).

Meanwhile, and partly independently from this theoretical debate, a wide body of evidence has documented the participation of sensorimotor information to speech perception, by evaluating the issue in human adults, children and infants through behavioral and neurofunctional assessments, post-damage investigations, developmental disorders observations and computational simulations (cf. the comprehensive review on the subject proposed by Skipper, Devlin & Lametti, 2017).

The main conclusion yield by this line of research is that available data are not consistent with auditory-only, nor with motor-only theories of perception. Two main consequences for current investigations can be drawn: firstly, the issue of a putatively prevailing role of one system over the other is not a worthwhile direction of inquiry and has to be discarded; rather, it is the fact that speech perception works on two connected binaries that has to be explicated. Secondly and consequently, when working in the multimodal perception framework, the aim is no more to demonstrate the sensorimotor participation, but rather to characterize its functions, in adults and developmentally.

In this connection, the motor system appears to play «ubiquitous, specific and context dependent [perceptual] roles»[25]. More precisely, its recruitment is a regularly occurring part of the process but, most probably, its role is differently important depending on the conditions of the communicative setting (conversely, the same can be said of the auditory one). In particular, as to current knowledge, it may play a preponderant role in under-optimal listening and attentional conditions, i.e. when the environment is somewhat noisy or distracting (worth to be noted, such examples constitute ordinary and frequent situations – cf. D'Ausilio et al., 2012; van Wassenhove, Grant & Poeppel, 2005).

As Skipper and colleagues point out, this perspective has been efficaciously formalized in the Analysis-by-Synthesis framework (Halle & Stevens, 1962; Bever & Poeppel, 2010), regarding speech perception as an inferential process acting on limited invariant information coming from different sources.

---

[25] Skipper, Devlin & Lametti, 2017, p. 79.

Not less importantly, the authors also underline that, as traditionally formulated, the research question regarding the perceptual involvement of the speech motor system is misconceived and misleading. In fact, *a* language-related motor system does not exist: rather (and as it will be better described in the next chapter, Foreword), the dynamics of speech production consists of multiple and interwoven fine-grained movements, realized by several articulators that, in turn, are directed by an equally distributed neurofunctional network[26]. This observation also entails that studies providing neural evidence against the relevance of sensorimotor implication could importantly be re-evaluated; for example, patients suffering from a brain damage to circuits traditionally associated to the control of motor speech patterns may not have lost *all* the linguistic sensorimotor experience that they have accumulated in the years[27]. Indeed, a major limitation of the studies targeting the link between perception and production, during development or adult age, lies in the assessment of productive abilities, often too coarse-grained (cf. Chap. 4, Section 5 for a further description of this aspect).

In a globally consistent direction with what observed by Skipper, Devlin & Lametti, a recent computational model (Barnaud et al., 2017 – COSMO model) implements the combination of auditory and sensorimotor information in the perceptual process as that of two coexisting but different decoders that, according with the principles of Analysis-by-Synthesis, work in the frame of a Bayesian fusion.

The authors report that COSMO allows a coherent explanation for different (and sometimes contradictory) sets of results, i.e. (a) findings observing that the motor system appears to be more importantly involved in adverse auditory or cognitive conditions (e.g. Wilson & Iacoboni, 2006; Zekveld et al., 2006); (b) results showing that Transcranial Magnetic Stimulation of motor areas temporally disrupts the elaboration of speech stimuli, but that a permanent brain damage to speech motor areas does not prevent perception; (c) studies highlighting a difference in the neurofunctional correlates characterizing the production and the perception of the same linguistic sounds (and thus arguing against the processing of articulatory actions during perception, cf. Cheung et al., 2016).

As to Barnaud and colleagues, those data would be interpretable by taking into account three functional attributes possibly characterizing the auditory/sensorimotor perceptual interaction, and namely: (i) redundancy, describing the fact that two different inferential processes accumulate evidence which is later fused in a categorical decision (thus, if evidence is controversial, as in

---

[26] Important to remember in this direction, the research program of Articulatory Phonology that, through the years, have furnished a systematic and detailed instrumental analyses of the articulatory gestures.

[27] «If we take seriously the complexity of speech production, any one production sub-process will involve its own distributed set of brain regions that might each play a role in speech perception and do so at different times, in different contexts». Skipper, Devlin & Lametti, 2017, p. 80.

audiovisual illusions, or if the fusion is not efficient, as in Transcranial Stimulation, the process leads to a wrong conclusion); (ii) complementarity: positing that (by physical nature) auditory processing happens in the range of a narrow band and motor processing in that of a larger one, one can explain the fact that the 'auditory decoder' outperforms the 'motor decoder' in optimal conditions and *vice versa* (and one can justify the fact that brain-damaged patients tested in good auditory settings succeed in perceptual tasks); (iii) specificity: when COSMO is applied to the processing of CV syllables made up of a plosive C, the model is better able characterize vowels via the auditory processor and the place of articulation of consonant via the sensorimotor processor.

Concluding this excursus, this latter point captures an important difference distinguishing vowels from consonants. Briefly mentioned in Chap. 1 (pages 8-9), this aspect can be adequately addressed at this point of the dissertation.

3.2 Developmental sensorimotor contributions to vowels and consonants: a hypothesis for future research

Though often collapsed under the same 'segmental' etiquette, vowels and consonants represent two sharply different entities, from both the production and the perception point of view. As it has been recently underlined, such physical differences also affect the linguistic experience that young learners can make with the two percepts. For example, vowels are (prototypically) more audible than consonants and from earlier in life (cf. Nishibayashi & Nazzi, 2016).

With the aim of contributing to this research direction, it can be observed that, from a complementary perspective, the production of consonants also requires a significantly greater motor effort when compared with the production of vowels. On such bases, it can be speculated that the perceptual elaboration of the two phonological classes could entail a different degree of participation from sensorimotor information: greater for consonants, smaller for vowels.

Following this line, intra-categorical differences can be predicted as well. For example, it seems conceivable that the degree of sensorimotor recruitment could (inversely) vary as a function of the audibility of the consonant, or that it could depend on the features which are processed (e.g. [sonority] being mostly assessable auditorily, place of articulation through sensorimotor recognition).

In a developmental perspective, this line of reasoning entails that, during phonological acquisition, a higher degree of sensorimotor knowledge – as acquired via experience with language-specific speech production – should be needed in order to gain full perceptual efficiency with consonant categories, in general and/or with some of them in particular (i.e. those requiring a particularly fine-grained sensorimotor effort of production, such as affricates).

In general terms, such conjectures appear indirectly supported by the experimental evidence displaying age-related differences in the timeline of typical perceptual specialization, where (as mentioned in Chap. 1, page 8-9) the process appears to be completed first for vowels and later for consonants (especially for motorily demanding ones). However, specifically focused research would be necessary to evaluate the topic in detail.

At first, this point of view may seem in contrast with a well-proved phenomenon, i.e. the consonantal bias displayed (for some languages, cf. Chap. 1, Section 3.2) in early lexical processing. However, this is not so.

Firstly, recalling the balanced division of labor that current research attributes to the auditory and motor system in perception (cf. Section 3.1), it seems possible to hypothesize that infants could indeed process consonant contrasts above chance levels even if not fully helped by sensorimotor information (especially in optimal listening conditions, e.g. accented syllable with no environmental noise). One example, in this respect, is the already cited case of [sonority]. This possibility would agree with current visions of multisensoriality in speech perception, which do not argue for the superiority of one source of information over the other, but for specialized contributions.

Secondly, the C-bias has recently been found to emerge from an earlier V-bias, declining between the 6th and the 8th month in French-learning infants (Nishibayashi & Nazzi, 2016). Remarkably, this period corresponds to the moment in which consonants typically emerge in babbling, becoming progressively attuned to the native language targets in the following months (e.g. de Boysson-Bardies & Vihman, 1991; de Boysson-Bardies, 1993).

Even if not directly addressed by the main experimental procedure reported in this thesis, the described hypothesis is an important aspect in order to ground phonological development into sensorimotor learning.

A preliminary attempt to deal with this question is described in Appendix I.

3.3 The perception/production link: a developmental perspective

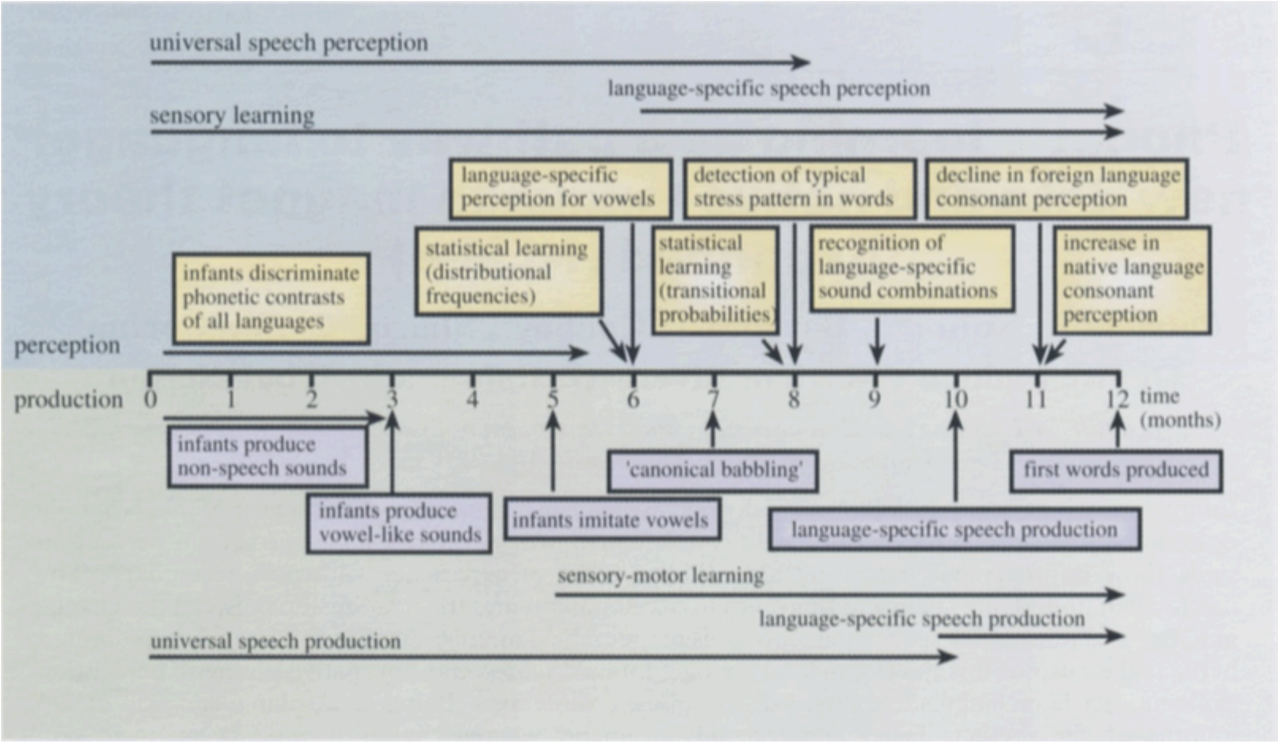
«Not having heard something is not as good as having heard it;
having heard it is not as good as having seen it;
having seen it is not as good as knowing it;
knowing it is not as good as putting it into practice»
Xun Zi, Book 8. Eric Hutton, ed., 2014

If the contribution of sensorimotor information to speech perception in healthy adults has been proved, from a developmental perspective, this raises the questions of when, how and why this mechanism is established.

Before proceeding to review the studies addressing this topic, it appears worthwhile to operate some terminological distinctions. In particular, language-related 'sensorymotor knowledge' can be defined as stored information about the motor and proprioceptive processes related to speech movements. Thus, 'sensorimotor processing' corresponds to a perceptual elaboration that recruits this knowledge and 'sensorimotor learning' defines the activities through which it is achieved, namely 'sensorimotor experience': the direct experience that the subject makes with the production and proprioception of speech movements.

Linguistic sensorimotor experience begins early in life. Importantly, though, not as early as experience with auditory and visual percepts.

Figure 4 – Timeline of infants' perception and production in the first year of life (Kuhl et al., 2008)



The salience of the experience with production during the first year is by no means comparable to that with perception[28] and this is due (among other things) to physiological constraints of anatomical and functional nature.

As compared with the auditory system, the articulatory system is far less mature at birth. In fact, it reaches the adult state through a long morphological evolution ending with puberty (cf. Fitch & Giedd, 1999; Mugitani & Hiroa, 2012). Besides, as this apparatus develops, one needs to learn to control it, i.e. to produce voluntary movements mirroring desired motor targets. While the discovery and the exploration of this capacity begin with babbling, the development of full speech motor control is characterized by a protracted acquisition of fine-grained sensorimotor skills, which are progressively automatized until adolescence (e.g. Walsh & Smith, 2002; Smith & Zelaznik, 2004).

Given the chronological misalignment characterizing perception and production, the timeline of their progressive connection in ontogenesis is complex to characterize.

Starting from the origins, as is well-known, newborns display mimicry of the other’s face (e.g. Meltzoff & Moore, 1977 and 1989). This also applies to speech movements (e.g. Chen et al., 2004 and Coulon et al., 2013 where, soon after birth, mouth-opening is preferentially operated when listening or looking at /a/ as compared to different sounds[29]). As a consequence, imitation has historically been proposed as a mechanism allowing the ‘cross-modal mapping’.

> «Human infants listen to ambient language spontaneously and attempt to produce sound patterns that match what they hear. In other words, infants acquire the specific inventory of phonetic units, words, and prosodic features employed by a particular language in part through imitation» (Kuhl & Meltzoff, 1996, p. 2427).

A crucial point in this view is the implication that, to a certain extent, the learner would need to judge the similarity between her/his own productions and the sounds perceived in the input, which would require a considerable attentional and mnemonic effort. In a slightly but significantly different direction, Messum and colleagues have argued that, rather than in imitation *per se*, the foundational unit linking perception and production has to be looked for in ‘mirrored vocal interactions’, i.e. the bidirectional interactions characterizing the communicative routine connecting parents and infants. Imitation is part of those settings, but it is also importantly performed by the

---

[28] As it can be noted, the proposed comparison does not take into account visual development. Though representing an important simplification (and not entailing any denying of the possible relevance of discussing this aspect in more detail), this choice is taken both for the sake of simplicity and because, in the scientific perspective assumed, the visual system acts as the receptor of information that is sensorimotor in nature.

[29] Cf. also Guellaï, Streri & Yeung (2014) for a further description of such aspects, comprising the criticism that have been addressed to such findings.

caregiver, who reformulates the infant's production. By this means, the expert social partner provides (positive or negative) evidence of the correspondence between the child's attempt and the correct model, acting as a «metaphorical mirror»[30].

After the foundations have been established, the connection between production and perception is obtained through experience in producing speech-specific movements. Capitalizing on this principle, effects of the emergence of production abilities over perception have been shown during the first year of life, with reference to both the auditory-only and the audiovisual modes of elaboration.

As a first case-study (De Paolis, Vihman & Keren-Portnoy, 2010), 10-to-16-months-olds were tested for their listening preference over speech passages containing non-words composed by consonants mirroring their individual production patterns. The latter were evaluated by means of a home-recording procedure, allowing to differentiate the acoustic stimuli in three categories: (a) consonants regularly produced by the participant; (b) consonants rarely produced, that can be realized at the age of testing (as in other peers' production patterns); (c) consonants not generally produced at this stage. As the results indicated, the infants displaying the richest production patterns preferred listening to consonants they did not yet produce on a regular basis.

This tendency has been interpreted as the marker of a more advanced developmental stage. The sensorimotor mastering of a sound would facilitate its processing by providing additional information. This, in turn, would allow more freedom to the subject's attention to address itself to new learnable objects. Agreeing results have been obtained with Italian children through a similar paradigm, with the difference that the groups were tested longitudinally and the stimuli consisted in isolated non-words (Majorano, Vihman & De Paolis, 2013).

Further evidence in this connection has been provided by De Paolis, Vihman & Nakai (2013), comparing 12-months-old English and Welsh natives on their processing of consonants having equal frequency in the auditory input, but unequal frequency in the first productions of the two languages. English participants, who are known to produce more often some of the proposed stimuli and more rarely some others, mirrored the patterns in De Paolis, Vihman and Keren-Portnoy (2010), i.e. they were more interested in listening to sounds that they did not regularly produce. The Welsh ones, who tend to produce all the experimental stimuli with equal frequency, did not show any differential preference.

---

[30] Messum & Howard (2015), p. 128.

These data were mirrored by findings in the audiovisual domain since, as reported by Altvater-Mackensen, Mani & Grossmann (2016), 6-months-old subjects belonging to the 'high-producers' category turn out to be better matchers of auditory and visual information (vowels).

Among other things, those studies also underline the fact that the detected sensorimotor effect is not necessarily connected with the individual profile in terms of specific correspondences (e.g., since a certain infant does not produce /n/, one would expect her/him not to be skilled at perceiving it); rather, the highlighted differences also pertain to high-producers *vs* low-producers: infants who have access to a more or less large knowledge of the motor and sensory patterns of the native language.

Incidentally, this is not to discard that the absence of *any* experience with a certain sensorimotor pattern has an effect over its elaboration.

In this sense, Streri & Yeung (2016) reported on an audiovisual matching task conducted with 3, 6 and 9-months-old French participants over three vowels: /a/, /u/ and /i/. The emerging looking patterns were compatible with a developmental increasing efficacy: while /a/ was easily matched by the younger participants, /u/ and /i/ became respectively better processed over the considered age range. This tendency appeared possibly connected with the French learners' production patterns, generally beginning to produce those stimuli in a consecutive order.

If the evidence hitherto considered is based on the effect of accumulated sensorimotor knowledge over speech perception, some results obtained via on-line tasks are also available[31]. Yeung & Werker (2013) observed that explicit manipulation of the oral articulators can alter the perceptual performance of 4.5-months-olds. Engaged in an audiovisual matching procedure, the subjects listened to /i/ or /u/ while seeing two articulating faces (one matching, the other mismatching). A group of infants completed the task while sucking on a pacifier, producing a [u] shaped configuration on the mouth; the other while playing with a teething toy, inducing a mouth shape more compatible with [i]. As the results highlighted, participants preferred watching at the mismatching audiovisual pair when making the same movements as those characterizing the sound they heard. In other words, as holds with reference to one's own sensorimotor *knowledge*, on-line sensorimotor *information* supports the elaboration of the stimulus, allowing the subject to devote attention to the 'novel, 'more relevant' one.

As displayed by Bruderer et al. (2015), this also holds with reference to auditory stimuli. The authors administered a group of 6-month-olds a discrimination task entailing the (non-native)

---

[31] As used in this context, the difference between 'on-line' and 'off-line' tasks consists in the fact that, while the first involve information which is made available to the participant during the procedure (e.g. the sensorial information carried by the use of the pacifier or the teething toy in Yeung & Werker, 2013), the second capitalize on information which is already stored in memory (e.g. the procedure used by De Paolis and colleagues).

Hindi contrast dental /d̪/ *vs* retroflex /ɖ/, which crucially involves a tongue-tip movement; the task was completed with or without a teething toy hindering the motricity of this articulator. The discrimination of the two items appeared temporally disrupted in infants with inhibited motricity.

Interestingly, while consistent with accompanying evidence, this result paves the way for a reflection upon the possible consequences on perception of developmental oro-motor disorders, which constitutes the topic of Chapter 3: as shown by this study or, similarly, by the described experiments entailing transcranial magnetic stimulation (pages 29-30), sensorimotor information can yield facilitatory or inhibitory effects, depending on its context of action.

3.4 Neural underpinnings: Creating the connection

In neurofunctional terms, the ontogenetic connection between production and perception has to be connected with findings concerning perceptual-related activity in the left Inferior Frontal regions, traditionally associated with the integration of multisensory information.

This pattern has been repeatedly identified. In 2006, Imada et al. reported the detection of a developmental continuum in the relative involvement of those areas which, irrelevant at birth, progressively intensifies among 6 and 12 months. Those results were interpreted as a reflex of production development across the first year of life, and particularly with reference to the onset of babbling towards the 6th month.

From a morphostructural point of view, Dubois and colleagues (2016) recently provided evidence that the circuitry associated with semantic elaboration matures earlier than the one associated with phonological processing and, in particular, with the mapping between sound and articulation[32]. Nevertheless, the latter displays a fast maturational realignment during the first five months of life, that the authors hypothetically attribute to the learning of the first speech-related cross-modal representations and, consequently, to the first combinatorial analyses of the input.

Finally, with reference to linguistic tasks other than simple passive listening, Kuhl et al. (2014) applied Magnetoencephalography to the auditory discrimination of native as compared to foreign syllables in participants of 7 and 11-12 months of age, i.e. straddling perceptual specialization. In the two cases, the recorded responses involved both circuitries supporting auditory processing (superior temporal areas) and areas connected with sensorimotor elaboration (Broca's Area, cerebellum). Crucially, while in younger infants those activations had comparable magnitude irrespectively of the proposed phonemes, at twelve months the discrimination of foreign stimuli

---

[32] The reference is clearly made to the 'dorsal' and 'ventral' pathways; first proposed as the two main neuromorphological substrates supporting language processing by Hickok & Poeppel (2004, 2007) and others (e.g. Warren & Warren, 2005; Wise, 2003).

elicited a greater sensorimotor contribution. These results fit with adult data (Section 3.1), arguing for early facilitatory effects of sensorimotor information in demanding tasks.

3.5 Towards the statement of the problem

To sum up, the contribution of sensorimotor elaboration is part of speech perception and the current problem in research is to better characterize its functions, both in adult's processing and in language acquisition.

Starting from late infancy, the multimodal elaboration of speech progressively becomes language-committed (i.e. not exclusively based upon the observation of domain-general features such as synchrony or physical similarity, but also supported by phonological information extracted from speech movements). In other words, toddlers make their entry in the out-and-out multisensorial linguistic dimension, a redundant condition which boosters perceptual efficacy. A fundamental role seems to be played in this transition by both the observation of the other's talking mouth and the acquisition of sensorimotor knowledge through experience with production.

Sensorimotor contributions to perception have indeed been observed starting from the first year of life, yielding enhancing effects on perceptual efficacy. Importantly, this enhancement is detectable independently from the processing modality (i.e. be the stimuli conveyed auditorily or audiovisually). Developmental differences are detectable in this context and appear connected with the individual achievement of sensorimotor knowledge as derived from speech-related experience.

However, not surprisingly, the development of such aspects does not end up with infancy. Rather, as it will be described in the next sections, once the perception/production relationship is set, the refinement of this cognitive interaction is completed by the end of childhood, when both multimodal processing and sensorimotor learning continue their ontogenetic trajectory.

4. Children: Multimodal perception and ongoing sensorimotor specialization

4.1 Multimodal speech perception during the childhood years

As a matter of fact, children appear less sensitive than adults to the visual component of speech and this has been remarked since the discovery of the McGurk effect.

In the original study, 3-to-4 and 7-to-8-years-old subjects were tested alongside with adults and appeared significantly less influenced by incongruent visual information (McGurk & MacDonald, 1976). Converging results were then collected by Massaro (1984), who also extended the investigation to the processing of visual silent articulation and documented a correlation between this and the audiovisual performance: children are not as skilled as adults in extracting linguistic information from the visual signal (cf. also Massaro & Cohen, 1986). On the same grounds, Hockley

& Polka described gradual emergence of responsiveness to visual features in the McGurk illusion's perception that, significantly, at 12 years is not yet fully adult-like (Hockley & Polka, 1994).

Several studies have replicated and investigated the reasons behind such results.

Among these, Erdener & Burnham (2013) displayed that, in children from the preschool to the school age, the degree of visual influence in the McGurk task is predicted by both their lip-reading abilities and their degree of perceptual specialization (as assessed *via* a same/different discrimination task entailing phonological and allophonic contrasts).

In a comparable direction, Lalonde & Holt (2014) assessed the multimodal processing of monosyllables in 6-to-8-years-olds, as compared to adults in three conditions: mismatching detection, discrimination and identification. The ratio behind the experiment lied in the fact that the three tasks required a growing ability to rely upon language-specific (as opposed to general) auditory and visual cues. Thus, based on the computation of the visual enhancement demonstrated by the groups[33], the authors showed that, while adults displayed greater visual improvement for tasks requiring higher levels of phonological elaboration, in younger participants the degree of visual support did not vary with the experimental condition.

Taken together, the tendencies detected in Erdener & Burnham (2013) and Lalonde & Holt (2014) suggest that, while adults can rely on both general and language-specific visual information, children may not have full access to the latter (cf. Baart et al., 2015 in the same direction).

As an alternative (or, rather, complementary) hypothesis, children could also be less able to rely on phonological visual cues due to excessive cognitive demands. Support in this direction comes from a study highlighting that even the general mechanisms underlying multimodal binding (such as the detection of temporal synchrony) may still be immature during childhood (Lewkowicz & Flomm, 2014). In this respect, Nardini, Bedford & Mareschal (2010) advanced a tempting hypothesis. In the authors' perspective, avoiding to integrate could be ontogenetically advantageous, as the prevalence of unimodal processing could grant gains in processing speed by allowing children to follow the fastest-available 'single' cue.

Turning from the origins to the consequences of the low visual enhancement characterizing childhood, in-noise perception represents a fundamental issue.

As it is known, the information derived from the visual signal is particularly supportive, for adults, in noisy conditions (e.g. MacLeod & Summerfield, 1987). Ross et al. (2011) evaluated this ability in five age-groups (5-to-7; 8-to-9; 10-to-11; 12-to-14 and 16-to-56-years). A speech-in-noise identification task was administered at various noise levels, chosen to cover a range of recognized

---

[33] I.e. the improvement in performance observed between the audiovisual task and an auditory-only baseline condition.

stimuli going from approximately 0% to almost 100% (isolated monosyllabic words covered with pink noise). As the results showed, children displayed significantly less visual enhancement than adults until adolescence and a significant age-related increase. In the authors' view, this suggested that «a considerable amount of multisensory learning remains to be achieved during the later schooling years»[34], i.e. that children do not successfully cope with visual speech cues because they lack (at least in part) the sensorimotor knowledge signaled by such cues.

Such pattern appears confirmed by Taitelbaum-Swead et al. (2016), who evaluated subjects among 5 and 80 years of age for the in-noise identification of monosyllabic (meaningful and non-sense) words (white noise, 0dB SNR). In the authors' results, this ability follows an inverse U-shaped trajectory across ages, with 4-to-5 and to 65-to-80-years-olds' performances as the lowest points and that of the 20-to-30 group representing the apex.

The difficulties that children display in processing auditory stimuli covered with noise has been described in Chap. 1 (Section 4.2). The additional finding that the availability of visual cues does not fully sustain them in those contexts has delicate consequences for pedagogical theories and practices (e.g. the organization of communicative interactions in the classroom).

At any rate, the weaker phonological advantage of multisensoriality in children as compared with adults does not imply that they do absolutely not benefit from visual cues.

In this connection, Jerger and colleagues (2014; 2017 and 2018) have recently highlighted that, given a low-fidelity stimulus (i.e. a stimulus where the sound of the onset consonant has been cut), the administration of a visual prime constituted by the silent articulation of the cut consonant can help young children (preschoolers and school-aged), provided the visual cues are sharply perceptible ([b]). Relevantly, Lalonde & Holt (2015) provided evidence in the same direction and with similar stimuli: in matching, discrimination and identification tasks, preschoolers appeared able to rely upon sharply perceptible visual cues ([ba] *vs* [ga]) but not upon poorly perceptible visual cues ([ba] *vs* [ma]). This latter aspect will be further examined in connection with the experimental procedure described in Chapter 4.

---

[34] Ross et al. (2011), p. 2329.

4.2 Neural underpinnings: The long way towards ‘visual’ language

Coherently with behavioral results, studies addressing the issue from the neurofunctional point of view provide evidence of multiple developmental changes happening through childhood.

From an electrophysiological perspective, the enhancing effect of audiovisual information is known to modulate adults’ responses, by producing waves at shortened latencies and with attenuated amplitudes as compared to auditory-only stimuli. As highlighted by Knowland and colleagues (2014), while the latency effect is detectable since the age of 6 years, the amplitude effect improves during development and is not fully mature at 12 years. In the authors’ interpretation, these results may reflect, on the one side, an early emergence of the temporal advantage provided by multimodal information. On the other, being amplitude attenuation putatively connected with the competition among information of different nature, they may hint at the developmental emergence of such redundancy (cf. also Kaganovich & Schumaker, 2014 on the same topic).

Pointing to similar conclusions, Dick et al. (2010) used fMRI to test functional reactions to the passive listening of short stories, presented in either auditory-only or audiovisual modality.

The audiovisual signal was found to activate an analogous network across 8-to-11-years-olds participants and adults, including areas associated with the integration of auditory and sensorimotor information (inferior frontal and premotor cortex, supramarginal gyrus, posterior superior temporal gyrus, planum temporale and posterior superior temporal sulcus). Crucially, though, an analysis of the connectivity among these areas revealed age-related differences. Namely, a less decisive involvement of the posterior inferior frontal gyrus and of the premotor cortex was detected in younger participants, suggesting that «the development of audiovisual speech comprehension proceeds through changes in the functional interactions among brain regions involved in language production and perception»[35].

In an interesting direction, Nath, Fava & Beauchamp (2011) investigated the neurofunctional correlates (fMRI) of individual differences in the perception of the McGurk effect among 6 and 12 years. In this connection, the recruitment of the superior temporal sulcus (a critical site for multisensory phonological integration) appeared more significant in children who more often reported to perceive the effect.

In conclusion, the emerging picture suggests that the neurofunctional and neuromorphological substrates of multisensory integration mature throughout childhood, supported

---

[35] Dick et al. (2010), p. 112.

by a growing association among phonological sounds, their motor and proprioceptive correlates and their visual reflexes.

4.3 Sensorimotor specialization during childhood years and its contribution to speech perception

Similarly to what described with reference to infants (Section 3.3), Turner et al. (2015) reported an on-line demonstration of the automatic contribution provided by sensorimotor information to perception during childhood. 7-years-old participants were administered videos of silently spoken words under two conditions: holding a tongue depressor (which substantially blocks facial motricity) or squeezing a ball with one hand. Coherently with former evidence, the participants with inhibited oral motricity scored significantly worst in identification.

These results are in agreement with those described by Yeung & Werker (2013) and Bruderer et al. (2015, cf. page 37-38). Taken together with adult research based on Magnetic Transcranial Stimulation, this line of research homogeneously demonstrates the disruptive effects exerted by strong speech-motor inhibitions on perceptual abilities.

As argued in Section 3.1 (pages 29-30), this does not support the standpoint that movements (and not sounds) represents the perceptual primitives. Rather, such data point to the fact that efficient speech perception is redundant by nature: a process where multimodal information is fused into a categorical decision, so that, when one of the channels is seriously compromised, a drop in performance takes place. In this perspective, the inhibition of the motor apparatus can be seen as the natural counterpart of communicative settings where the auditory channel is so noisy to be completely unexploitable, i.e. silent articulation.

Silent articulation, or ‘visual-only’ perception is often reported as being the most difficult receptive condition (e.g. Taitelbaum-Swead & Fostick, 2016). Tough, a remark about this point can be made. Experiments arguing in this sense usually take three terms of comparison: auditory-only, audiovisual and visual-only speech elaboration. Yet, importantly, carefulness is required when using such terminology. Specifically, what is ‘auditory-only’ in such procedures is the *modality* in which the stimuli are administered, not the *elaboration*. On the contrary, speech elaboration is supported by multisensory knowledge both when the stimuli are received audiovisually and auditorily (unless one of the two neurofunctional paths is somehow disrupted, as in the experiments mentioned above). In other words, the actual ‘auditory-only’ *processing* modality is the one described in studies hindering the recruitment of motor information: an inefficient condition.

An implication of this *caveats* is that, when stimuli are conveyed through the auditory-only modality, their normal processing provides access to both auditory and sensorimotor information while, when they are delivered in the visual-only modality, only the sensorimotor modality is available. Thus, if a primacy of the auditory channel exists, it also consists in its inherent access to redundancy and not (not only) in its robust physical characteristics.

Paralleling infant research, the investigations on sensorimotor perceptual contributions during childhood have also evaluated the off-line effects of accumulated sensorimotor knowledge. In this direction, research have capitalized on the fact that certain phonological contrasts require the control of particularly fine-grained motor schemes and, therefore, are only completely mastered late in development.

In this connection, McAllister Byun & Tiede (2017) assessed discrimination in late childhood (English participants ranging from 9 to 14 years of age) by means of a forced-choice task containing real words varying on a phonetic continuum going from /ɹ/ (a late emerging sound) to /w/ (similar, but earlier-emerging). Production skills were also assessed, by eliciting the articulation of the experimental sounds and judging the recorded realizations on the basis of their waveform. The results highlighted a significant association between perception and production proficiency.

Evidence in this direction has also been provided by Altvater-Mackensen & Fikkert (2010) with younger infants (Dutch 14-months-olds), as engaged in a word-learning experiment entailing both ‘easy’ and ‘difficult’ consonants (i.e. stops and fricatives). Based on the Switch Task (cf. Chap. 1, page 12), the procedure required the learning of a new word and its discrimination when a change in one of the consonants is produced (either a change from stop to fricative or from fricative to stop, e.g. from /paap/ to /faap/). The results highlighted asymmetrical patterns: firstly, while the infants were able to detect the manner-of-articulation change in the initial (and stressed) position, in final position they were not. Furthermore, the significance of the group’s discrimination scores was statistically explained by a consistent ability to detect the change from fricatives to stops, but not *vice versa*. This pattern mirrored the participants’ production skills (toddlers are generally more prone to substitute fricative with stops than the contrary and, moreover, master the contrast earlier for the onset than for the final position).

Such results are relevant to what argued in Section 3.2 (‘Developmental sensorimotor contributions to vowels and consonants: A hypothesis for future research’) and represent a first explicit connection between research on the phonological specification of early word forms and research targeting the sensorimotor contribution to the formation of phonological categories.

Yet, significantly, the task used proved to constitute the most difficult procedure in this context (‘Switch Task’, cf. Chap. 1, page 12). Hence, further research applying a wide range of methodologies is needed in order to evaluate these results.

To close this chapter by enlarging the perspective assumed, effects of experience with productions over sensorimotor perceptual influences are also detectable on cross-linguistic bases.

In this connection, a weaker influence of visual articulatory information has repeatedly been reported in Japanese as opposed to English speaking adults (e.g. Sekiyama & Tohkura, 1991 and 1993). From the ontogenetic point of view, Sekiyama & Burnham (2008) highlighted that, at 6 years, the effect of articulatory information on a perceptual decision (McGurk effect) is equally low in English and Japanese 6-years-old children; crucially, it increases in the English subjects around the 8th year of life but remains stably low in the Japanese ones (more precisely, the two groups of speakers differ in visual information processing not only for accuracy, but also for rapidity).

Interestingly, Japanese speakers may be led to develop this different elaboration of visual information by reason of differences inherent to their native phonological system. Specifically: «Japanese syllable identification may be less difficult with auditory information alone due to the smaller number of vowels (…) and lack of some consonant contrasts that occur in English»[36]. Moreover, and as the authors underline, the syllabic structure of English is more complex and varied than the Japanese one, notably entailing clusters and phonemic contrasts characterized by small acoustic differences and concomitant visual differentiations (e.g. /θ/ *vs* /s/ or *vs* /v/).

Tentatively, such cross-linguistic differences between English and Japanese could also be seen as elements promoting a relatively dissimilar (i.e. higher and lower) sensorimotor experience of the language patterns. This could produce a differential accumulation of sensorimotor knowledge and, in turn, a differential involvement of such information during speech processing. This direction of research remains to be further investigated.

---

[36] Sekiyama & Burnham (2008), p. 318. Cf. also Maddieson (2009) for an approach towards the calculation of complexity in phonological systems.

5. Statement of the problem - Part One

This chapter has explored the different sensory dimensions pertaining to speech perception and, particularly, the role played by sensorimotor information in this context.

The issue has been addressed from a developmental perspective, describing the ontogenetic emergence of sensorimotor contributions. Set during the babbling stage, those become slowly finer-grained through experience with production and, once acquired, produce more efficient perceptual capacities.

The existence of correspondences between perceptual and productive abilities has been documented in adults[37]. Conversely, the phenomenon still needs to be characterized from an ontogenetic perspective, investigating both its emergence and its significance. In this connection, further investigations on perceptual skills at different levels of sensorimotor proficiency appear legitimate and worthwhile.

This dissertation addressed the topic with reference to different categories of individuals. Namely, young adults (seen as the 'final state' in the ontogenetic picture) as compared with typically developing children in two age ranges: preschoolers (known to display a very low level of sensorimotor enhancement) and school-aged subjects, characterized – as to the reviewed literature – by initial advancements in this connection. Those subjects were taken together with a fourth category, i.e. children affected by Childhood Apraxia of speech: a developmental disorder hindering typical speech sensorimotor development.

This pathology and some connected atypical conditions are described in Chap. 3.

---

[37] Worthwhile to remark, even if they were not addressed in this chapter, experimental phonetics has produced very specific studies in this respect, e.g. Newman, 2003; Perkell et al., 2004a and b; Villacorta, Perkell & Guenther, 2007.

3. DEVELOPMENTAL PRODUCTION DISORDERS AND THEIR PERCEPTUAL CORRELATES

To recapitulate Chap. 2, a fundamental difference exists between perception and production in development: while the auditory apparatus supporting the former is able to operate as early as the prenatal period, the articulatory endowment reaches maturation well later in ontogenesis (cf. Foreword below). This state of affairs implies a fundamental consequence, i.e. although the signal is multimodal since the beginning of life, visual-sensorimotor information is fully exploitable only later, when the subject gradually accumulates sensorimotor knowledge through experience with production.

Given these premises, an excursus concerning the main characteristics of speech-related motor development appears worthwhile. This description will yield complementary information about the ontogenetic association of perception and production and will open the perspective on individuals with production disorders.

Foreword: on speech production and its development

Across the most different languages, speech articulation is a remarkably fine-grained motor skill, involving the coordinated action of approximately one hundred different muscles and the production of many distinctive sounds per second (cf. Duffy, 2000; Ladefoged & Maddieson, 1996).

To develop such ability, young learners pass through two interconnected and long-lasting ontogenetic paths: anatomical and neurofunctional development, i.e. the physical maturation of phono-articulatory organs and the achievement of the neurocognitive ability to control their movements.

Anatomical-physiological development is dramatic in the first year, when an outright «restructuring»[38] takes place. In this period, the vocal tract bends to form a right angle, thus provoking the descent of the anterior part of the tongue, the larynx, the hyoid and the epiglottis (which is distanced from the velum). Completed by a considerable global growth in length, width and volume, this postnatal reshaping ultimately differentiates the human vocal structure from that of the primates (e.g. Ficth & Giedd, 1999; Vorperian, 2005 and 2009; Mugitani & Hiroya, 2012).

Speech-related anatomy also evolves in more 'accidental' ways: for instance, the teeth appear during the first year and, although articulation is not their primary function, they contribute to the production of many sounds of the world's languages.

---

[38] Vorperian (2005), p. 338.

Importantly, the anatomical development of the articulatory system is not homogeneous, as different articulators develop at different rates. As Moore (2004) underlined, this also entails that children learn speech movements in the context of an evolving homeostasis: «the mass of the mandibule, for example, increases very rapidly during the period of initial speech acquisition, requiring the child to generate new speech behaviors without the benefit of a known, constant sensorimotor system»[39]. In other words, speech articulation has to be practiced and learned while relying on a physical endowment that is not fully adult-like and is also under physical transformation.

Clearly, these circumstances do not prevent phonological acquisition (also because the rate of structural change becomes significantly slower during childhood, providing sufficient stability). Nevertheless, they point to the complex dynamics characterizing speech motor development: such is the anatomical-physiological context in which the neurofunctional control of the speech motor system is acquired.

In this connection, speech production is controlled through efferent neural commands directed to the realization of voluntary movements. Those involve the fast, both parallel and serial displacement of several articulators that, though intentional, is highly automatized (i.e. one does not need to explicitly think about how to realize a sound or sound sequence – at least while speaking the native or a well-practiced language).

The link connecting cognitive triggering and speech action is commonly described as composed by two cognitive phases, followed by their motor implementation.

Table 1 – Speech Motor Control

(i) planning phase	Retrieval of invariant motor plans (e.g. to produce /b/, an opening/closure of the lips is required)
(ii) programming phase	Adaptation of the motor schema to the contingent phonetic context (e.g. the [b] at the onset of [bɔːl] will be slightly different from the [b] at the onset of [ˈbeɪbɪ], due to coarticulatory requirements)
(iii) execution	Execution of the motor program (e.g. the lips open and close while beginning the protrusion necessary for the following vowel in [bɔːl])

Observing these capacities during development, instrumental studies conducted in the last decades have shown that children's linguistic movements are less rapid, less accurate and less consistent than the adult ones and complete speech motor control is only reached at the onset of adolescence (e.g. Smith & Goffmann, 1998; Goffmann & Smith, 1999; Walsh & Smith, 2002; Smith

---

[39] Moore (2004), p. 193.

& Zelaznik, 2004; Cheng et al., 2007). In this protracted evolution, the babbling and the first multimodal interactions set the action-perception cycle; but the creation of fine-grained connections among motor, somatosensory and auditory information is only attained after a long journey through linguistic experience.

The attainment of articulatory cognitive control proceeds through non-simultaneous developmental stages. It begins with an early mastering of the jaw's vertical movements for opening/closure (allowing the production of some universally early emerging phonemes, i.e. /a/, /p/, /b/, /m/), which is followed by the progressive attainment of the independence of the lower lip (permitting facial retraction and rounding, as necessary to realize mid, closed and rounded vowels). The last achievement in this progression is represented by the control of the tongue (and especially of its tip) which gradually develops until the end of childhood and is involved in the realization of the motorily most challenging distinctive features (e.g. Hayden & Square, 1986; Green et al., 2000; Green, Moore & Reilly, 2002; Smith & Zelaznik, 2004; Green & Nip, 2010; Murdoch, Cheng & Goozée, 2011). Alongside this progression in the control of places of articulation, phonatory abilities are refined and muscular tone and strength increases, allowing progressive experience with the different manners of articulation included in the target phonology (e.g. Boliek et al., 1997).

Importantly, while in traditional accounts the completion of the phonological inventory is claimed to occur during the preschool period, such evidence shows that this is a substantial, though acceptable, simplification: preschoolers' productions are close enough to the target to be perceived (i.e. cognitively reconstructed) by the listener as adequate but, in sensorimotor terms, they are not. This situation probably yields perceptual consequences, contributing to the scarcely efficient exploitation of visuo-motor information observed during childhood (Chap. 2, Section 4). These difficulties are overcome as age grows in the context of typical development, but carry critical implications for individuals with developmental production disorders.

1. When the speaking mouth does not follow the hearing ear: Evidence from developmental production disorders

By definition, clinical populations affected by developmental speech disorders do not follow the typical progression in phonological production: rather, they suffer from delayed emergence and/or atypical realization of the speech sounds.

With the aim of examining the relationship between perception and production in this context, evidence from three domains will be considered, namely: developmental phonological disorders, childhood apraxia of speech and childhood-onset stuttering.

Table 2 – production disorders yielding perceptual consequences

	Speech sound disorders	**Speech fluency disorder**
Absence of motor deficit	Developmental phonological disorder	
Presence of motor deficit	Childhood apraxia of speech	Childhood-onset stuttering

As summarized in Table 2, the three disorders affect, on the one side (left column), the capacity to articulate the speech sounds (speech sound disorders, i.e. early-onset «persistent difficulties with speech sound production that interferes with intelligibility or prevent verbal communication»[40]); on the other (right column), the ability to control the time patterns underlying correct articulation (speech fluency disorder, i.e. stuttering). In stuttering and apraxia of speech the deficit is underlined by inadequate motricity of the articulators, while this is not the case in phonological disorders (generally attributed to language-related attentional and mnemonic processes).

Relevant to this thesis' purposes, these conditions share a fundamental characteristic: in their idiopathic form (i.e. congenital, the only considered in this review), they occur as primary disorders in otherwise healthy individuals, thus providing an opportunity to assess the effects of lacking articulatory experience over speech perception development.

Indeed, such conditions have long been regarded as 'output disorders', disturbing the delivery of verbal messages in individuals otherwise fully able to both conceive and receive them. Not undisputed in the context of adult-acquired speech pathologies[41], this standpoint is openly wrong in connection with developmental disorders. During ontogenesis, the paths leading to global linguistic proficiency are interconnected and interdependent; thus, a deficit in one of these domains can yield cascading effects on the others. As underlined by D'Souza & Karmiloff-Smith (2011): «the infant brain starts out highly interconnected, and it is only over developmental time that neural networks become increasingly specialized – that is, relatively modularized». On account of this, children presenting production disorders cannot be described as following a normal development *except for* speech production. Rather, due to their primary deficit, they undertake a globally deviating linguistic development (ending up with producing more or less satisfactory outcomes, due to a constellation of factors). Consistently with this point, they also show reduced perceptual acuity.

---

[40] DSM-5, Speech Sound Disorders.

[41] Cf. Randazzo M. (2016) for recent and pertinent evidence about audiovisual speech perception in adults with acquired, secondary apraxia of speech.

Evaluating adults with persistent developmental stuttering, Corbera et al. (2005) highlighted that, as compared with non-stutterers, they displayed atypical Mismatch Negativity Responses to auditory speech contrasts – while, relevantly, not to pure tones – and these were statistically connected with the participants' fluency profile, as measured through a self-evaluation questionnaire. In the same direction, Chunming et al. (2016) reported on a group of adult stutterers performing significantly lower than controls in auditory identification tasks. As observed through fMRI, their performance was correlated to a difference in the activation of the left anterior insula (part of the system controlling speech motor abilities) and this differentiation was also detectable when the same participants were engaged in a speech production task. In addition, the connectivity linking motor and auditory areas differed significantly between stutterers and non-stutterers and the strength of this connection correlated with the performance obtained in the perception task.

Analyzing the issue in the context of phonological disorder, Nijland et al. (2009) documented perceptual difficulties in this clinical population. In their study, participants ranging from young to mid-childhood (5.5-to-7.11 years of age) scored under the age norm in auditory discrimination and identification, and such results correlated with the participants' degree of accuracy in articulation (calculated as the percentage of correctly produced consonants in elicited speech passages).

Children with phonological disorder(s)[42] have also been evaluated from a multimodal perspective. In this connection, Desjardins, Rogers & Werker (1997) tested the explicit hypothesis that «experience correctly producing consonants plays a role in developing the underlying representation which mediates the perception of visible speech». With this aim, the participants of this study were assessed for articulatory proficiency and then tested for the identification of syllables repeated in auditory-only, visual-only and (congruent or incongruent) audiovisual modality. In agreement with the authors' hypothesis, the participants displaying poorer articulation skills also showed lesser visual perceptual proficiency. In the same connection, Dodd et al. (2008) reported a strongly auditory-biased elaboration of the McGurk effect in preschoolers with a diagnosis of phonological disorder.

Overall, the study of developmental production disorders has testified of reduced perceptual efficacy in the subjects concerned and of correlations between the latter and production skills. Albeit obtained from a reduced amount of studies, such evidence points to the significance of an adequate sensorimotor experience for the development of speech perception.

---

[42] Sub-types of developmental phonological disorder indeed exist. For the sake of simplicity, this chapter makes reference to such conditions as to a unitary entity.

This thesis aims at contributing to this direction of inquiry by focusing on childhood apraxia of speech – a condition which is particularly relevant to this scientific context.

2. Childhood Apraxia of Speech, or the uncontrolled movement

2.1 On this condition

Recalling the three stages of speech motor control described at the beginning of this chapter (Table 1), Childhood Apraxia of Speech (CAS) can be defined as a deficit hindering the process at the root: despite the absence of structural muscular deficits and for unclear neurobiological reasons[43], apraxic subjects are not able to acquire well-formed and stable motor plans for speech. Thus, when trying to perform speech movements, they turn out to be impaired in both precision and consistency: not only are their productions inaccurate, but the same phonological type can be realized in several different ways that vary in an unpredictable fashion (cf. ASHA, 2007).

Such inconsistency in language-related motricity is a crucial phonetic expression of CAS, that can be illustrated through Bernstein's description of the 'degrees of freedom problem' (1967): given that the human effectors can be used in multiple ways to approximate the same motor goals, how can coordination be efficient and fast? Freedom of movement must somehow be cognitively constrained, otherwise it would represent a disturbing factor.

Ontogenesis provides an answer to such issue, through the progressive acquisition of motor schemas (i.e. fixed routines applied to the execution of recurrent motor goals, e.g. Sporns & Edelman, 1993). The young child progressively selects optimal configurations associated with linguistic motoric purposes and, by repeating them, these gradually become accurate, consistent and automatized[44]. This fundamental goal is not accomplished in subjects suffering from CAS, which prevents the correct multimodal connection, since the babbling phase, between motor cause and acoustic consequence. Hence, oral movements are scarcely controlled and yield scarcely predictable results. In a nutshell, CAS causes a radically unreliable experience with speech production that, more or less severe as the case may be (cf. Section 2.2), jeopardizes the acquisition of sensorimotor linguistic knowledge.

In typical development, the execution of well-controlled speech actions produces sensorimotor knowledge through the coupling of motor and proprioceptive information. More in

---

[43] Apraxia, as well as all other congenital speech and language disorders, is due to genetical reasons. To date, however, those have not been fully clarified (e.g. Chilosi et al., 2015).

[44] Worthwhile to remark, the issue of the units of automatization (be those phonemes, syllables, words, high-frequency sentences or all of them) is a highly debated one, with reference to both adults and children (e.g. Smith, 2006).

detail, the execution of speech motor schemas produces proprioceptive sensations that, once repeated and memorized, provide the agent with a *forward* cognitive model, allowing the prediction of the expected sensory consequences when performing a certain speech action (e.g. when the subject wants to utter /b/, the proprioceptive sensation of lips closing and opening is expected). On the other side, the actual sensory correlates following the action are compared to this internal model through a *feedback* path, enabling to verify one's efficiency.

This cognitive cycle captures the type of learning that the individual affected by CAS is missing (a loss causing proprioceptive deficits, cf. Newmeyer et al., 2009; Iuzzini-Seigel et al., 2015). Ultimately, CAS entails a significant reduction of sensorimotor experience with speech, a partial 'dis-embodiment' of this aspect of cognition. In the words of Dawson (2014): «embodied cognitive science emphasizes feedback between an agent and the world (...) This, in turn, suggests that agents with different kinds of bodies can be differentiated in terms of degrees of embodiment»[45].

In consideration of all of these premises, an investigation of the perceptual correlates of this pathology appears to be highly desirable. The hypothesis is that, while phonological disorders (entailing reduced production experience without sensorimotor deficit) and stuttering (characterized by difficult but not disrupted motricity) affect speech perception, CAS should carry even more significant consequences.

Before addressing the central topic of perceptual findings in CAS, the next section provides a closer description of its production correlates.

---

[45] Dawson M. (2014), page 62.

2.2 Speech production in idiopathic Childhood Apraxia of Speech

> «The impression that the listener receives of the speech of a person with moderate to severe CAS is of effort and struggle».
>
> Velleman (2011), p. 82

The phonetic and phonological correlates of idiopathic CAS are defined by a set of coexisting symptoms, constituting the perceivable manifestations of this central sensorimotor deficit. Available evidence can be summarized as follows.

Table 3 – Phonetic and phonological symptoms of Childhood Apraxia of Speech

Anomalous babbling patterns, i.e. delayed, scarce, non-duplicated, and/or non-variegated (as reconstructed through parental questionnaire)	Aziz et al. (2010); Davis & Velleman (2000); Highman et al. (2012); LeNormand (2000); Velleman (2011); ASHA (2007)
Incomplete or atypical phonological inventory (the order of acquisition may be altered and sounds that are not phonematic in the target language can be present)	Davis et al. (1998); Davis & Velleman (2000); Velleman (2011); ASHA (2007)
‘Groping’, i.e. production of oral movements directed to the research of an articulatory configuration	Davis & Velleman (2000); ASHA Technical Report on Childhood Apraxia of Speech (2007)
‘Phonological inconsistency’, i.e. high, atypical variability in the phonetic realization of the same phonological target; constituting the perceivable correlate of motor inconsistency and particularly hindering intelligibility[46]	Aziz et al. (2010); Davis et al. (1998); Davis & Velleman (2000); Highman et al. (2012); Lewis et al. (2004); Marquardt & Jacks (2004); ASHA (2007)
Deletion of phonological material (concerning segments or syllables)	Aziz et al. (2010); Davis et al. (1998); Davis & Velleman (2000); Highman et al. (2012); ASHA (2007)
Phonemes substitution	Aziz et al. (2010); Davis et al. (1998); Davis & Velleman (2000); Highman et al. (2012); ASHA (2007)

---

[46] It is to be noted that phonological inconsistency entails great difficulties for the listener when trying to ‘translate’ the disturbed message (e.g. in a 20-minutes session recorded by the author, a patient aged X years utters 17 times the Italian word /rana/ (i.e. frog), by realizing it in 6 different tokens (i.e. [ra]; [rʹʔa:na]; [rʹa:da]; [ˈɰa:na]; [ˈla:na]; [ˈɾa:na]); another patient produces 9 different tokens of the word /kane/ (dog) over 14 total productions.

Phonemes distortion (i.e. use of a non-typical sound, e.g. lateral /l/ or /s/)	Aziz et al. (2010); Davis et al. (1998); Davis & Velleman (2000); Highman et al. (2012); ASHA (2007)
Anomalous prosody (and with particular reference to the word-form level), stemming from the disfluent production of speech motor schemas	Aziz et al. (2010); Davis et al. (1998); Davis & Velleman (2000); Highman et al. (2012); Shriberg et al. (1997, 2017); ASHA (2007)
Positive correlation of speech errors with length and structural complexity of the phonological target (due to deficitarian motoric coordination which ultimately makes of coarticulation the most difficult task)	Aziz et al. (2010); Davis & Velleman (2000)

To sum up, speech production in apraxic children is scarcely intelligible due to widespread and substantial phonological errors, exacerbated by their inconsistent occurrence and by significant coarticulation difficulties.

It is worthwhile to point out that this disorder covers a wide range of severity levels. To illustrate this point, three of the subjects taking part in a former study conducted by the author can be taken as reference (Lorenzini, 2014[47]). To begin with, Subject 1 (female, 5.6 years) presented a very mild form of CAS, essentially limited to difficulties in articulating the finer-grained contrast of the Italian phonological system, that were realized through incorrect and inconsistent attempts. By contrast, Subject 2 (male, 5.6 years) represented a prototypical, scarcely intelligible apraxic child: challenging coarticulatory patterns (e.g. consonant clusters or words with more than two syllables) were largely omitted through deletion and/or substitution of the target phoneme sequences; atypical sounds[48] were frequent and phonological inconsistency largely exceeded 50% of total productions (as measured over high-frequency single words like /kane/ 'dog' or /orset:o/ 'little bear'). Finally, Subject 3 (male, 6.4 years) appeared affected by an extremely severe form of the disorder (worsen by a complex and unstable context of multilingualism characterizing his environmental input). When elicited through a naming task, his linguistic production was limited to a few sounds (mostly vowels or syllables) and a small number of highly automatized words (e.g. /mam:a/ 'mother' and /dito/ 'finger'). This severely affected behavior was underlined by spontaneous attempts to exploit the

---

[47] Lorenzini I. (2014). *Dis-embodied Language: language acquisition in the lack of oro-articulatory references. Speech characteristics of Italian children with Apraxia of Speech*. M.A. Thesis. Data presented at the XI national AISV conference (Italian Association of Voice Sciences), Università Alma Mater di Bologna, January 28th-30th 2015.

[48] E.g., in this subject, the use of the voiceless lateral fricative /ɬ/ as an approximation of consonant clusters such as /fr/ or /dʒ/.

gestural channel for communicative purposes, by means of which the child provided rich pantomimic descriptions of the target pictures[49].

Importantly, the described difficulties do not represent a mere articulatory phenomenon; rather, their early onset shapes language acquisition, yielding an at least partially altered phonological encoding. In this connection, Preston et al. (2014) provided evidence of atypical electrophysiological processes in participants with CAS producing isolated words. Specifically, as compared with an age-matched control group, apraxic participants displayed reduced responses during a time window which is normally associated with the phonological encoding of word forms. It is worth noting that those data were obtained during late childhood (participants ranging from 9 to 15 years of age).

From a complementary perspective, Fiori et al. (2016) reported evidence of reduced brain connectivity in this clinical population, concerning areas supporting both speech and language functions. Moreover, the strength of the connectivity patterns identified was correlated with both the speech motor ability of the subjects[50] and their lexical production skills.

2.3 Speech perception in Childhood Apraxia of Speech: Former findings

As for the other speech disorders described in Section 1, children with CAS score below typically developing peers in experiments targeting perception, and some evidence has been accumulated about a possible contribution of their production deficit to the phenomenon.

To begin with, this population has been found to display difficulties in discrimination tasks entailing both vowels and consonants (Bridgeman & Snowling, 1988; Maassen, Groenen & Crul, 2003). Importantly, those data have been collected in participants between mid and late childhood (from 7 to 11 and from 6.11 to 9.6 years, respectively) and it is to be noted that difficulties in the perception of vowels are particularly atypical in this period. Similar results have been highlighted in rhyme-judgment experiments, straddling prosodic/perceptual and meta-phonological abilities (Marion, Sussman & Marquardt, 1993).

In more detail, Froud & Khamis-Dakwar (2012) have assessed the presence of the Mismatch Negativity Response in English-speaking, 5-to-8-years-old apraxic children listening to phonemic and allophonic contrasts. As the authors reported, neurofunctional reactions appeared significantly altered in this clinical population who, differently from the control subjects, displayed weaker, ‘immature’ negativity signatures to the phoneme opposition alongside with an electrophysiological reaction to allophones which – as is well known – does not normally occur in subjects of this age (the

---

[49] Cf. in this direction Davis & Velleman (2000), who report a tendency towards the establishment of home-sign systems in extremely severe apraxic children.

[50] Diadochokinetic Rate, cf. Chap. 4, Section 5.1.

automatic identification of non-phonemic contrasts declining during infancy). Ultimately, such results provide a demonstration of low perceptual commitment to linguistic sounds in these participants.

The relationship between sensorimotor production and perceptual abilities has also been targeted in this context. In this respect, Rvachew & Jamieson (1989) underlined that, when 5-years-old apraxic subjects are engaged in the identification of words ranging on acoustic continua based on phonemes that are associated with their production difficulties (/s/, /ʃ/ and /f/), errors in the two modalities mirror one another. In the same direction, and with a substantially analogous methodology, Groenen et al. (1996) highlighted a positive correlation at 8.9 years (mean age) between the ability to properly realize the distinctive features characterizing the proposed stimuli (/d/ and /b/) and their discrimination (as embedded in acoustic continua). Analogously, the previously described study by Nijland (2009) also entailed apraxic participants (cf. page 51).

Finally, McAllister Byun (2012) reported a case study on a 4-years-old subject, arguing for diminished perception, in a non-word discrimination task, of the phonological contrasts that the child more often neutralized in production (e.g. /dɪz/ - /gɪz/). In the author's opinion, these data «constitute evidence that a primary production deficit can cause decreased perceptual ability».

The studies addressing this last issue have been carried out in the framework of a very specific perspective, aimed at assessing whether a scarce ability in producing certain sounds entails difficulties with its processing. This direction of research is partially divergent from the one investigating the topic in typical development which, as described in Chap. 2 (Sections 3.3 and 4.3), also highlights a more widespread pattern in which globally less proficient sensorimotor profiles appear associated with generally less efficient perception skills. This difference points to the disadvantageous disconnection sometimes characterizing research on typical and atypical development, preventing adequate comparisons between data coming from the two fields and hints at the many perceptual aspects that still have to be investigated in the context of apraxia of speech.

Among these (as the reader may have noticed), to the author's knowledge, no research has to date directly addressed speech-in-noise and multimodal linguistic elaboration in this population. Considering the strong sensorimotor involvement that lies at the heart of this condition, studies in this direction appears to be fully justified. The experimental procedure described in Chapter 4 entails a first exploration of such aspects.

Statement of the problem – Part Two

With the aim of contributing to the study of the developmental relationship between perceptual and sensorimotor skills, the experimental procedure described in Chap. 4 was conducted with apraxic children as contrasted with typical ones. Precisely, it explored their discrimination skills (i) within an experimental setting explicitly eliciting reliance upon sensorimotor processing; (ii) with and without a strong audiovisual support; (iii) in connection with simple and complex consonant sounds.

As described (Chap. 2, pages 29-30), a major criticism that can be addressed against radical motor theories of perception consists in the fact that, if speech motor schemas really are the primitives, then subjects suffering from congenital or acquired motor limitations should be unable to process speech. Clearly, this is not the case. Conversely, the question becomes legitimate and relevant if reframed in perspectives that posit the complementary contribution of multimodal sources of information (such as the Analysis by Synthesis model).

PERCEPTUAL PROFICIENCY AT DIFFERENT LEVELS OF SENSORIMOTOR KNOWLEDGE

1. General rationale

Sensorimotor knowledge can take part in speech perception as long as it is accumulated through linguistic experience, i.e. inasmuch as speech motor schemas are repeatedly and accurately practiced.

Capitalizing on such premise, this research aimed at contributing to the study of the ontogenetic emergence of the sensorimotor contribution, with reference to the childhood period.

To this goal, perceptual proficiency was tested in an experimental procedure eliciting reliance upon sensorimotor processing, completed by individuals endowed with different degrees of sensorimotor development. In this respect, an 'Ontogenetic Amount Method' was followed, consisting in «comparing the abilities of individuals of different ages, brought up in functionally identical environments, on a common task»[51] and, by such means, evaluating «the effect that *amount* of experience and *amount* of maturation may have on development»[52].

The participants were: (i) healthy young adults, representing the ontogenetic 'final state' (i.e. fully-attained speech sensorimotor abilities); (ii) two groups of typically developing children (younger and older), endowed with immature sensorimotor knowledge and (iii) children with apraxia of speech, a congenital deficit hindering speech sensorimotor learning.

The experimental plan entailed two tasks, respectively targeting perception and production. Statistical analyses were performed on each task separately and on the correlational patterns emerging between the two.

---

[51] Burnham & Sekiyama (2012), page 64.

[52] Ibidem.

2. Participants

111 participants took part in this study. Among them, one adult, 9 (out of 57) typical children and 3 (out of 21) apraxic participants were excluded from the analyses due to lack of collaboration[53]. Hence, the rejection rate was 16% in typical children, 14% in the apraxia group and irrelevant in adults.

The collaborating participants were distributed as follows.

Mean Age: M.A. (yrs.mos); *Standard Deviation*: S.D.

32 young adults: M.A. 23.8 yrs; RANGE 19-31; S.D. 3yrs; 13 F.

25 typically developing pre-school children: M.A. 4.9yrs; RANGE 4-5.7; S.D. 6mos.; 13 F.

23 typically developing school-aged children: M.A. 7.5yrs; RANGE 6.9-8.9 S.D. 7.8mos.; 10 F.

18 children with apraxia of speech: M.A. 6.5yrs, RANGE 4-9.8; S.D. 15mos.; Linguistic Age 5.3yrs; RANGE 3.5-7; S.D. 1yr; 4 F.

Linguistic Age (i.e. the level of linguistic development) was assessed for the apraxic participants by the clinicians of the Institute of Research and Care Fondazione Stella Maris on the basis of a standardized procedure[54].

Typical children were control and experimental participants at the same time. For these purposes, two different age-groups were recruited (preschool age group and school-age group). This allowed, on the one side, to compare the apraxic participants for both linguistic and chronological age (preschoolers and school-aged children, respectively); on the other, to assess whether the healthy children's capacity to cope with the proposed stimuli changed significantly in correspondence with the onset of schooling.

Gender, age and all other relevant characteristics of the groups are detailed at the end of this section (in Tables 4 to 7).

Adults were undergraduate and PhD students from the Scuola Normale Superiore of Pisa and the University of Pisa (candidates with an expertise in language sciences, speech and language therapy or psychology were not recruited). Typical children were enrolled thanks to an agreement with their schools, the kindergarten and the elementary school of the Istituto Comprensivo Strenta Tongiorgi of Pisa. Participants with apraxia were recruited by virtue of the collaboration

---

[53] The rejected adult was visibly uncomfortable during the procedure and obtained an extremely deviant score; the children either did not consistently look at the screen or explicitly asked to stop the experiment.

[54] *Test di Comprensione Grammaticale per Bambini* – TCGB. Chilosi, Cipriani & Ruschi (2005), 2nd edition.

between our institution and the scientific hospital for child neurology and psychiatry Istituto di Ricovero e Cura a Carattere Scientifico Fondazione Stella Maris[55].

All participants were native Italian speakers. Children were monolingual; adults who were able to speak other languages were accepted, provided the latter were late-acquired.

For what pertains to the typical groups (adults and children), individuals who had suffered from any type of cognitively-relevant disorder or injury were not included, even if displaying full recovery. Participants who declared genetic familiarity for speech disorders were accepted, and this variable was considered in the statistical analyses. This aspect was only directly assessable in adults, as the majority of typical children's parents refused to give personal information: to cope with this problem, the experimenter relied on the teachers' judgment for this aspect.

Adults were characterized by a very high educational profile (i.e. enrolled in university or graduated). Taken together with the age-range of the participants (20-to-30 years, corresponding to a peak in perceptual efficiency across the life-span – e.g. Taitelbaum-Swead & Fostick, 2016), this aspect most probably introduced a positive bias in the group's profile. On the other hand, it granted a high homogeneity in the results and, importantly, set a benchmark of the best performance obtainable by naïfs participants in this study.

Candidate members for the Apraxia of Speech group were accepted if affected by a congenital form of the pathology, i.e. genetically-based, not arising from a brain damage or a wider syndromic condition (which could carry cognitive difficulties unrelated to those characterizing the genetic form of the disorder, cf. Chap. 3, Section 2.2 and 2.3).

Extremely severe, non-verbal children were not included.

The duration of the speech therapy was considered as a variable in the statistical analyses (when not already undertaken, a 0 score was attributed). Furthermore, it was asked if and for how long the individual had undertaken the P.R.O.M.P.T. therapy ('Prompts for Restructuring Oral Muscular Phonetic Targets' – cf. Dale & Hayden, 2013): this approach is based on the re-education of the cognitive connection among speech motor actions, their proprioception and their acoustic correlates, hence, significant relationships among this factor, production and perception skills were hypothesized.

---

[55] The Institute is one of the largest scientific hospitals for child neurology and psychiatry in Italy and it is a reference point for the diagnosis and the rehabilitation of Italian children with apraxia of speech. Dr. Anna Chilosi, who leads the unit of Speech and Language Disorders, and Dr. Beatrice Franchi (speech therapist) directed the recruitment of the participants and gave very valuable insights during the fulfillment of this study.

Table 4 – Adult participants

Participant	**Age (Years)**	**Gender**	**Familial history**	**Socio-Economic Status**
1.	27	M	Positive (stuttering)	PhD student
2.	25	M	Negative	PhD student
3.	21	M	Negative	Undergraduate student
4.	27	F	Negative	PhD student
5.	21	M	Negative	Undergraduate student
6.	22	M	Negative	Undergraduate student
7.	24	M	Negative	PhD student
8.	27	F	Negative	PhD student
9.	20	M	Negative	Undergraduate student
10.	27	M	Negative	PhD student
11.	24	F	Negative	PhD student
12.	25	M	Negative	PhD student
13.	19	F	Negative	Undergraduate student
14.	22	F	Negative	Undergraduate student
15.	27	F	Negative	Undergraduate student
16.	26	M	Negative	Undergraduate student
17.	24	M	Negative	Undergraduate student
18.	28	M	Negative	Undergraduate student
19.	24	M	Negative	Undergraduate student
20.	26	M	Negative	Undergraduate student
21.	21	M	Negative	Undergraduate student
22.	28	F	Negative	Undergraduate student
23.	30	F	Negative	Undergraduate student
24.	31	F	Negative	PhD student
25.	21	M	Positive (stuttering)	Undergraduate student
26.	20	F	Negative	Undergraduate student
27.	22	F	Positive (deafness)	Undergraduate student
28.	22	M	Positive (speech sound disorders)	Undergraduate student
29.	22	F	Negative	Undergraduate student
30.	21	M	Negative	Undergraduate student
31.	20	M	Negative	Undergraduate student
32.	20	F	Positive (dyslexia)	Undergraduate student

Table 5 – Typically developing preschoolers

Participant	**Age (yrs.mos)**	**Gender**
1.	4.3	M
2.	6	F
3.	5	M
4.	5	F
5.	5.4	M
6.	4.4	F
7.	4	F
8.	5	M
9.	5.3	M
10.	5.5	F
11.	5.2	F
12.	5.5	M
13.	5	F
14.	4.8	M
15.	5.7	F
16.	4.7	F
17.	4.9	M
18.	5.2	F
19.	5.6	M
20.	5.6	M
21.	4	F
22.	4.5	F
23.	4	F
24.	5	M
25.	5	M

Table 6 – Typically developing school-aged children

Participant	**Age (yrs.mos)**	**Gender**
1.	8.3	M
2.	7.8	M
3.	8.3	M
4.	8.7	M
5.	8.7	F
6.	8	F
7.	8.9	F
8.	8.4	M
9.	8.7	M
10.	8.6	M
11.	8.1	F
12.	8.1	M
13.	7.5	F
14.	7.10	M
15.	7.1	F
16.	7.4	F
17.	7.6	F
18.	6.9	F
19.	7.2	M
20.	7.5	M
21.	7.1	M
22.	7.1	M
23.	7.9	F

Table 7 – Childhood Apraxia of Speech group

Participant	Age (yrs.mos)	Linguistic Age (yrs.mos)	Gender	Speech therapy (yrs.mos)	PROMPT therapy (yrs.mos)	Not-own speech motor schemas[56]
1.	7.1	5	M	5.5	1	/r/; /ʎ/
2.	5.8	4	F	2.8	0.4	/ɲ/; /b/; /d/; /g/; /ts/; /dz/; /tʃ/; /dʒ/; /v/; /ʃ/; /r/; /ʎ/
3.	5	N.A.	M	N.A.	N.A.	N.A.
4.	8.1	6.6	M	3	1.8	/ɲ/; /b/; /d/; /g/; /ts/; /dz/; /tʃ/; /dʒ/; /f/; /v/; /s/; /z/; /ʃ/; /r/; /ʎ/
5.	6.4	5	M	2.10	N.A.	/b/; /v/; /d/; /z/; /r/; /tʃ/; /dʒ/; /ʃ/; /ɲ/; /ʎ/; /g/
6.	7.3	7	M	3	2.4	N.A.
7.	6.1	5	M	3.6	2.2	/ts/; /dz/; /r/; /tʃ/; /dʒ/; /ʃ/
8.	4	N.A.	M	N.A.	N.A.	N.A.
9.	4.5	3.6	F	0.3	0	/l/; /r/; /ʃ/
10.	6.6	4.6	F	2.6	0.5	/dz/; /r/; /ʃ/; /ɲ/; /ʎ/ Atypical production of /s/; /tʃ/; /dʒ/
11.	7.1	5.6	M	3.4	0	/z/; /r/; /dʒ/; /ʃ/
12.	6.5	5	M	N.A.	0.4	/r/; /z/; /ts/; /dz/; /tʃ/; /dʒ/; /ʃ/
13.	7.2	7	F	3.8	3	/r/; /ɲ/; /ʎ/; /dz/; /ts/
14.	6.10	6	M	3	2.1	/s/; /z/; /tʃ/; /dʒ/
15.	5.3	5	M	2.6	0	/ɲ/; /g/; /ts/; /dz/; /v/; /s/; /z/; /r/; /ʎ/
16.	7	N.A.	M	1	1	N.A.
17.	9.8	7	M	7.1	2	/dz/; /ʃ/; /dʒ/
18.	7.3	4	M	N.A.	0	/ɲ/; /r/

---

[56] I.e. consonants that that the participant was not able to produce. Those data were taken from the more recent evaluation conducted by the neuropsychologist or the speech therapists of the IRCCS Fondazione Stella Maris.

3. Methods - Perception task

3.1 Factors constraining the experimental design

This study was conceived with the view of comparing largely different populations. As described, a special focus was put on apraxic children, a severe speech disorder with perceptual consequences (cf. Chap.3, Section 2). Thus, the need to assure the feasibility of the task with this latter group (as well as with the younger typical participants) had a primary importance. To this goal, possible issues of heterogeneous nature were taken into account.

To begin with, some potential disadvantages stemmed from the settings where children were tested.

Adults freely scheduled their appointment; they were encouraged to come in a good psychophysical condition and were examined in an optimal research environment. By contrast, younger participants could only be assessed in non-experimental settings and within a limited time. Specifically, the experimenter obtained the parental consent to carry out with each participant a single 30-minutes session, conducted either at school or at the rehabilitation institute.

In the latter contexts, a quiet room was uniformly provided. However, the participants' state of fatigue as due to prior school or clinical activities was not fully controllable. In particular, a large number of the apraxic participants took part in the experiment on the occasion of a periodical follow-up (i.e. during or after this activity), which consists in an articulated diagnostic protocol. Thus – what represents a common shortcoming in the fields of clinical and school research – the children's attentional state may have been weakened at the moment of the testing session. Considered in the context of the inherently less robust attention capacities of children (and especially of disordered children) in comparison with adults, this represented a particularly delicate aspect.

Furthermore, speech-in-noise tasks are known to be challenging for young participants (Chap. 1, Section 4.2 and Chap. 2, pages 40-41) and were expected to be possibly even more so for the apraxic ones, whose abilities in this connection had never been previously assessed.

Consequently, some important decisions were taken, aimed at obtaining a procedure that could at the same time minimize the cognitive load and optimize the evaluation of the fundamental research question, i.e.: whether associations between (high or low) proficiency in production and perception would emerge in the groups, as evaluated in a listening condition eliciting reliance upon sensorimotor processing.

To these goals, a single speech-in-noise audiovisual procedure was used. In this connection, the experiment did not directly evaluate the audiovisual enhancement provided by the administration of visual cues in noisy conditions (differing from some previous studies which targeted similar abilities, e.g. Lalonde & Holt, 2014). Albeit limiting the evidence which could be collected in this respect, this aspect did not interfere with the results (cf. Section 3.2, pages 68-70; Chapter 5, Section 1).

Moreover, the experiment was built upon a reduced amount of trials, i.e. 8 'same' and 8 'different'. Specifically, as Table 11 (page 71) reports, the latter were obtained by repeating in each order of presentation four phonological contrasts. This admittedly restricted choice proved to be appropriate because, as described in the next Sections, attention-related effects emerged in the performance of the apraxic participants (cf. pages 83-84 in particular).

Finally, CV syllables were preferred to meaningful words in order to avoid holistic word-recognition and elicit a segment-per-segment analysis. Pseudowords were also ruled out, since significant difficulties in their encoding and retrieval have been documented in apraxia of speech (cf. Shriberg et al. 2012).

3.2 Rationale

The experimental procedure consisted in a forced-choice (same/different) discrimination task.

Given the fundamental aim of this research, i.e. to evaluate possible connections between sensorimotor development and speech perception, a necessary precondition was to support the activation of such knowledge.

To this purpose, a speech-in-noise audiovisual task was used (White noise, 0 dB SNR) since, in this case, sensorimotor processing should give a particularly vivid contribution to perception (cf. Chap. 2, Section 3.1). Reliance upon sensorimotor elaboration was further supported by: (i) showing (only) the mouth region of the face during the whole procedure (thus offering a very detailed vision of the articulatory movements, that was unlikely to be ignored); (ii) warning participants that the audio track would only be residually reliable and inviting them to devote attention to the mouth movements (cf. page 76).

The task explored the participants' ability to cope with in-noise perception in two conditions, varying along two parameters:

Table 8 – Phonological parameters implemented in Condition 1 and 2

	Parameter	**Description**
Condition 1	Visibility	Speech-in-noise perception of CV syllables as supported by rich *vs* poor visual cues
Condition 2	Motor complexity	Speech-in-noise perception of CV syllables containing consonants whose production requires a high *vs* low level of motor proficiency

As shown below, condition 1 had two values for visibility, while motor complexity was controlled (i.e. invariably low).

Table 9 – Perception task, condition 1: experimental stimuli

Condition 1

	Rich visibility	Poor visibility
Low motor complexity	[ba] *vs* [ga]	[da] *vs* [na]

In Condition 2, the factors were reversed: motor complexity had two values, while visibility was controlled (i.e. invariably low).

Table 10 – Perception task, condition 2: experimental stimuli

Condition 2

	Low motor complexity	High motor complexity
Poor visibility	[da] *vs* [na]	[dza] *vs* [dʒa]

The opposition [poor visibility; low motor complexity] consisted of the same pair of stimuli in both conditions ([da] *vs* [na]), thus allowing a comparison of the results.

Before turning to a detailed description concerning the stimuli's characteristics and the experimental procedure, some important considerations on the experiment are in order, with the aim of explaining the logic behind each condition and discussing the predictions that could be derived from the physical (i.e. acoustic and articulatory) nature of the consonants used.

To begin with, condition 1 (Table 9) evaluated speech-in-noise discrimination with the support of rich *vs* poor visual speech cues. Its aim was to assess whether children with apraxia would show sensitivity to visual speech cues, as compared against typical age- and language-matched control participants.

Rich visual cues are well-known to provide a great perceptual enhancement in adults (e.g. Ross et al., 2007; Lalonde & Holt, 2015; Files et al., 2015) and such effect has been recently confirmed in children, starting from the preschool age (Jerger et al., 2014, 2017 and 2018; Lalonde & Holt, 2015). On the other hand, the apraxic population has never been compared with typical participants in this respect.

The perceptibility of poor visual cues represents a more controversial subject.

Their detectability during adulthood was discarded by past literature (e.g. Woodward & Barber, 1960) which, essentially, regarded as perceivable only the movements overtly recruiting the lips (e.g. bilabial phonemes) or the anterior part of the teeth (e.g. interdental phonemes). Recent research, however, has revealed the existence of more nuanced differences. In this connection, Files et al. (2015) enrolled young adults in a visual discrimination task based on CV contrasts which were 'far' or 'near' in their visual perceptual distance (i.e. e.g. [fa] *vs* [ʒa], far; [ta] *vs* [ʒa], near). The results showed that, albeit better in the 'far perceptual distance' condition, discrimination was above chance levels for the 'near perceptual distance' and the authors concluded that listeners «are sensitive to and make use of the [visual] details in natural speech stimuli»[57] (cf. Bernstein, 2012 in the same direction).

Thus, even if the topic still needs further investigation, the idea of a visual sensitivity limited to rich speech cues does not fit anymore with the understanding of audiovisual speech perception, but rather seems to be connected with past theoretical standpoints underestimating the importance of the visual signal.

While this conclusion holds for adults, recent studies have mostly concluded that poor visual cues are not exploited developmentally (cf. Jerger et al., 2014 and 2018; Lalonde & Holt, 2015). In this respect, it is worthwhile to point out that the poor visual cues that have been tested for discrimination during childhood are not of the same sort as the contrasts used by Files and colleagues, which always granted a residual (albeit low) visibility (cf. examples above). Rather, they are characterized by so low visibility as to be considerable non-visible (e.g. [ba] *vs* [ma] in Lalonde & Holt)[58].

The same is true for the poor visibility opposition used in this study, i.e. [da] *vs* [na] where, as in [ba] *vs* [ma], the point of articulation of the two consonants is exactly the same.

Given the above premises, it was hypothesized that:

(i) both adults and children would benefit from the rich visual cues distinguishing [ba] *vs* [ga], but not from the poor visual cues distinguishing [da] *vs* [na];

(ii) in the poor visibility condition, adults would efficiently rely upon the residual auditory signal, while children would not (cf. Chap. 1, Section 4.2).

Therefore, no significant effect of the trial-types on discrimination accuracy was expected in adults while, conversely, a significant facilitation for the richly visible trials was expected in typical children.

As previously mentioned (page 66-67), this experiment did not entail a base-line acoustic task comparing the children's auditory and audiovisual performance. Moreover, the former literature did not produce relevant data pertaining to the auditory-only perceptibility of the stimuli used (i.e. as covered with analogous noise and with Italian participants of the age tested). Therefore, this study

---

[57] Files et al. (2015), page 888.

[58] A commentary on this aspect entailing a proposition for future research will be proposed in Chapter 5 (Section 3).

did not provide information about the amount of perceptual enhancement stemming from the visual cues offered (the same is true for condition 2, cf. the paragraphs below). However, this limitation did not affect its goal, i.e. to assess whether or not children with apraxia would in absolute terms benefit from rich visibility speech cues. In this respect, the degree of informativity of the contrasts used was so objectively distant ([ba] *vs* [ga] prototypically visible, [da] *vs* [na] most probably not visible) that an advantage for the rich visibility trials was reasonably conceivable for any participant endowed with sensitivity to visual speech cues[59].

The specific research question was: would the apraxic participants better perform with the rich visual difference between [ba] and [ga] as compared with the poor visual difference between [da] and [na]?

Essentially: are they sensitive to visual speech information at all?

Condition 2 (cf. Table 10) had the goal to test the elaboration of stimuli varying in articulatory difficulty. The hypothesis was that, consistently with former literature, differences in the groups' performances would emerge, connected with the participants' level of sensorimotor development (cf. Chap. 2, Section 4.3; Chap. 3, Sections 1 and 2.3).

In this context, only poor visual cues were administered, in order to avoid the risk that participants could base their discrimination decisions on easily detectable visual traits. Therefore, as for the poorly visible trials of Condition 1, it was taken for granted that the participants (especially the children) would not benefit from the visual cues presented and would rely on the residual auditory signal.

Albeit representing a major limitation, the lack of a direct evaluation of the (relevant/irrelevant) role played by the visual cues in the discrimination judgment did not interfere with the goal of this condition, i.e. assessing the differential contribution of sensorimotor knowledge to the groups' performance. In fact, the sensorimotor perceptual contribution is activated by both auditory and audiovisual stimuli (Cf. Chap. 2, Section 3.1; Section 4.3, pages 43-44), particularly so in noisy conditions.

Moreover, this research question was further evaluated through an analysis of the correlations between perception and production scores (cf. Section 7).

Concluding this general description, Table 11 recapitulates the general structure of the perception experiment and spells out the labels used to describe its different parts. The same labels will be used in the following sections.

---

[59] This hypothesis particularly holds with reference to children aged from 7 to 9-years, whose ability to perceive [da] *vs* [na] as compared with [ba] *vs* [ga] turns out to be equal, when tested at the same SNR (albeit with a noise of different nature, i.e. speech-shaped pink noise: cf. Nishi et al., 2010).

Table 11 – perception task: terminology

Condition 1

Parameter: visibility

Trial-types	Rich visibility (Low motor complexity)	Poor visibility (Low motor complexity)
Set 1	[ba]$_1$ *vs* [ga]$_1$	[da]$_1$ *vs* [na]$_1$
Same trials	[ba]$_1$ *vs* [ba]$_2$; [ga]$_1$ *vs* [ga]$_2$; [da]$_1$ *vs* [da]$_2$; [na]$_1$ *vs* [na]$_2$	
Set 2	[ga]$_1$ *vs* [ba]$_1$	[na]$_1$ *vs* [da]$_1$

Condition 2

Parameter: motor complexity

Trial-types	Low motor complexity (Poor visibility)	High motor complexity (Poor visibility)
Set 1	[da]$_1$ *vs* [na]$_1$	[dza]$_1$ *vs* [dʒa]$_1$
Same trials	[da]$_1$ *vs* [da]$_2$; [na]$_1$ *vs* [na]$_2$; [dza]$_1$ *vs* [dza]$_2$; [dʒa]$_1$ *vs* [dʒa]$_2$	
Set 2	[na]$_1$ *vs* [da]$_1$	[dʒa]$_1$ *vs* [dza]$_1$

3.3 Characteristics of the stimuli

The stimuli were filmed in a soundproof cabin. White noise was superimposed on the audio track after controlling for intensity, F0 and duration; the degraded auditory signal was then re-synchronized to the corresponding video in AVS Video Editor.

Gaussian white noise was randomly generated in Praat at a Signal to Noise Ratio (SNR) of 0dB. White noise was preferred to other types of degradation since it represents an ‘ecological’ option, close to real-life situations.

The 0 dB SNR was chosen after a calibration procedure. Ten adults having the same profile of the participants (but not taking part into the experiment) undertook the task at different SNRs, with a between-session interval of one week. SNRs of -20, -10 and 0 dB were tested, until reaching a mean total score of 85% correct discrimination in the latter condition (Standard Deviation 10%).

The creation of the stimuli was carried out at the Laboratory of Linguistics of Scuola Normale Superiore, with the collaboration of Dr. Irene Ricci and Dr. Chiara Bertini.

The auditory and visual characteristics of the stimuli are reported in Tables 12 to 15.

It is to be noted that trials of the ‘same’ type (e.g. [ba] *vs* [ba]) were obtained by two different video-frames (e.g. [ba]$_1$ *vs* [ba]$_2$ – cf. Table 11). In this way, identity between elements in a pair was granted on linguistic-categorical, rather than on physical-contingent bases.

Table 12 – Auditory characteristics of the stimuli, condition 1

‘Different’ trials

	Rich visibility		**Poor visibility**		**Mean (SD)**
	[ba]$_1$	**[ga]$_1$**	**[da]$_1$**	**[na]$_1$**	
Duration (Ms)	1800	1820	1800	1835	1813.75 (17)
F0 (Hz)	235.25	238	239.24	240.11	236.8 (3.5)
Intensity (dB)	73.77	73.17	72.78	70	72.4 (1.7)
Input type frequency (occurrences in *Phonitalia*)	14659	9721	25748	69085	29803 (27031)

‘Same’ trials

	[ba]$_1$	**[ba]$_2$**	**[ga]$_1$**	**[ga]$_2$**	**[da]$_1$**	**[da]$_2$**	**[na]$_1$**	**[na]$_2$**
Duration (Ms)	1800	1800	1820	1835	1800	1800	1835	1850
Mean (SD) Duration	1800 (0)		1827 (10)		1800 (0)		1842.5 (10)	
F0 (Hz)	235.25	232.37	238	231.25	239.24	237.85	240.11	240.46
Mean (SD) F0	233.81 (2)		234.6 (4.8)		238.5 (1)		240.3 (0.25)	
Intensity (dB)	73.77	73.77	73.17	72.66	72.78	72.68	70	68.47
Mean (SD) Intensity	73.8 (0)		73 (0.4)		72.7 (0.07)		68.7 (0.3)	

Table 13 – Auditory characteristics of the stimuli, condition 2 (‘different’ trials)

‘Different’ trials

	Low motor complexity		**High motor complexity**		**Mean and SD**
	[da]$_1$	**[na]$_1$**	**[dʒa]$_1$**	**[dza]$_1$**	
Duration (Ms)	1800	1835	1810	1800	1811.25 (16)
F0 (Hz)	239.24	240.11	237	235	238 (2.3)
Intensity (dB)	72.78	70	74	72	72.2 (1.7)
Input type frequency	25748	69085	10066	12177	29269 (27438)

‘Same’ trials

	[da]$_1$	**[da]$_2$**	**[na]$_1$**	**[na]$_2$**	**[dza]$_1$**	**[dza]$_2$**	**[dʒa]$_1$**	**[dʒa]$_2$**
Duration (Ms)	1800	1800	1835	1880	1800	1760	1810	1800
Mean (SD) Duration	1800 (0)		1857.5 (32)		1780 (28)		1805 (7)	
F0 (Hz)	239.24	237.85	240.11	240.5	235	234.3	237	239.24
Mean (SD) F0	238.5 (1)		240.3 (0.25)		234.65 (0.55)		238.12 (1.6)	
Intensity (dB)	72.78	72.7	70	68.5	72	72.46	74	72.8
Mean (SD) Intensity	72.7 (0.07)		69.23 (1.08)		72.23 (0.33)		73.4 (0.85)	

Table 14 – Visual characteristics of the stimuli, condition 1

‘Different’ trials

	Rich visibility		**Poor visibility**		**Mean and SD**
	[ba]$_1$	**[ga]$_1$**	**[da]$_1$**	**[na]$_1$**	
Duration (Ms)	670	685	730	710	699 (26.6)
Movement onset (Ms)	390	520	560	500	492.5 (72.74)
Movement end (Ms)	1060	1205	1290	1210	1301 (332)

‘Same’ trials

	[ba]$_1$	**[ba]$_2$**	**[ga]$_1$**	**[ga]$_2$**	**[da]$_1$**	**[da]$_2$**	**[na]$_1$**	**[na]$_2$**
Duration (Ms)	670	590	685	660	730	745	710	730
Mean (SD) Duration	630 (56.5)		673 (18)		737.5 (10.6)		720 (14.14)	
Movement onset (Ms)	390	430	520	520	560	480	500	520
Mean (SD) Movement onset	410 (28.3)		520 (0)		520 (56.7)		510 (14.14)	
Movement end (Ms)	1060	920	1205	1180	1290	1225	1210	1250
Mean (SD) Movement end	990 (99)		1193 (18)		1258 (46)		1230 (28.3)	

Table 15 – Visual characteristics of the stimuli, condition 2

‘Different’ trials

	Low motor complexity		**High motor complexity**		**Mean and SD**
	[da]$_1$	**[na]$_1$**	**[dʒa]$_1$**	**[dza]$_1$**	
Duration (Ms)	730	710	900	900	817.5 (97.4)
Movement onset (Ms)	560	500	450	550	515 (50.7)
Movement end (Ms)	1290	1210	1350	1450	1325 (101.2)

‘Same’ trials

	[da]$_1$	**[da]$_2$**	**[na]$_1$**	**[na]$_2$**	**[dʒa]$_1$**	**[dʒa]$_2$**
Duration (Ms)	730	745	710	730	900	800
Mean (SD) Duration	737.5 (10.6)		720 (14)		815 (70.7)	
Movement onset (Ms)	560	480	500	520	450	450
Mean (SD) Movement onset	520 (56.6)		510 (14)		450 (0)	
Movement end (Ms)	1290	1225	1210	1250	1350	1250
Mean (SD) Movement end	1275.5 (46)		1230 (28)		1300 (70)	

3.4 Procedure

A discrimination task was preferred to an identification task because, normally, apraxic children do not master the complete phonological inventory (and, in fact, this was confirmed by the examination of the recruited participants, cf. Table 7, page 65).

The stimuli were delivered through Presentation (version 183071816) with an intra-stimulus interval of one second and were administrated via the same portable computer in all experimental sites. Participants sat at a distance of approximately 60 cm from the screen and, when presenting with mild vision problems, they wore their glasses. The sound was administered through headphones at a mean intensity of 70 dB.

The answers were recorded by means of a button box, calibrated relatively to each participant's dominant hand (i.e. the green button, equivalent to the 'same' answer, was placed in correspondence with the dominant hand). Younger participants did not use the button box autonomously, but rather communicated their answers to the experimenter; hence, reaction times were only recorded for adults, while discrimination accuracy was measured in all groups.

For children, the perception and production tasks were incorporated into a ludic activity, ending with a reward. Moreover, playful attention-getters (smiling faces) were inserted after each trial and a brief training phase was provided, based on trials that were not used in the experiment (e.g. [ka] *vs* [ma]). During the calibration procedure described in Section 3.3 (page 71), adults demonstrated no need for a training phase; therefore, for them, this was not included.

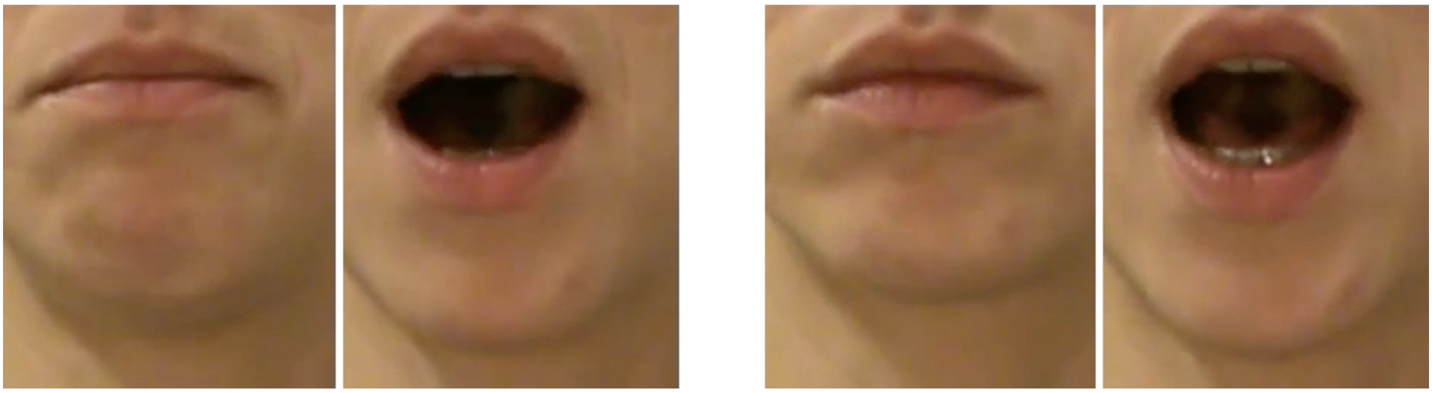
The stimuli were pseudo-randomly ordered and presented in all possible combinations (i.e. [ba]/[ga], [ga]/[ba] etc.), yielding a total of sixteen experimental trials (pooled across conditions, cf. Table 11, page 71). The order of administration of the two conditions was counterbalanced.

Given that two different video-clips were used in the 'same' trials, in the verbal instructions given before running the tasks, participants were asked to tell whether the mouth *said* the same syllable or two different syllables (rather than generically reporting whether what they perceived was identical or different). Furthermore, they were informed that the sound would not be fully reliable due to noise.

Considering the difficulty of the task, and in order to precisely focalize the attention of the children (in particular the apraxic ones), only the mouth portion of the speaking face was shown in the videos. As previously described (page 67), this decision also aimed at supporting perceptual reliance on sensorimotor processing.

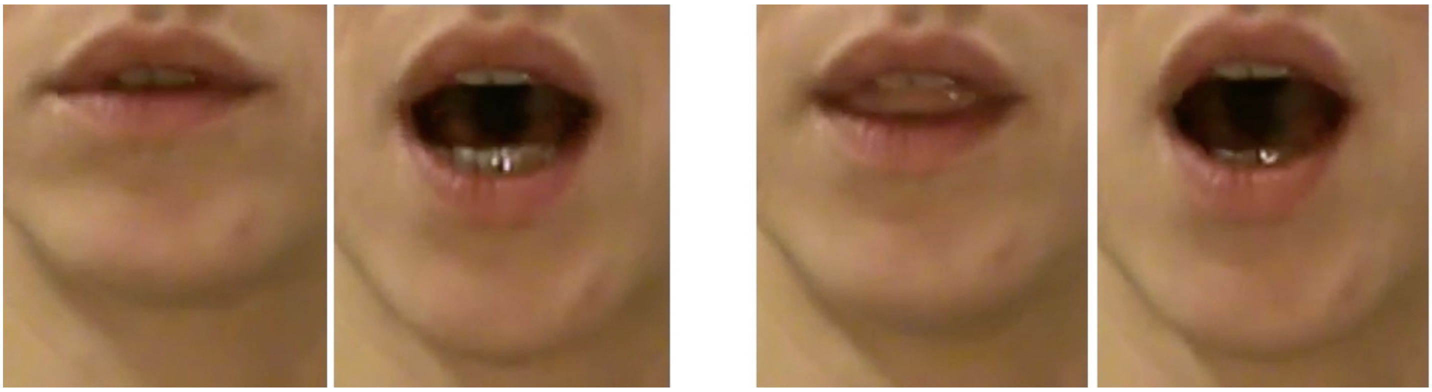
Figure 5 – Still frames from the perception tasks

Rich visibility; Low motor complexity



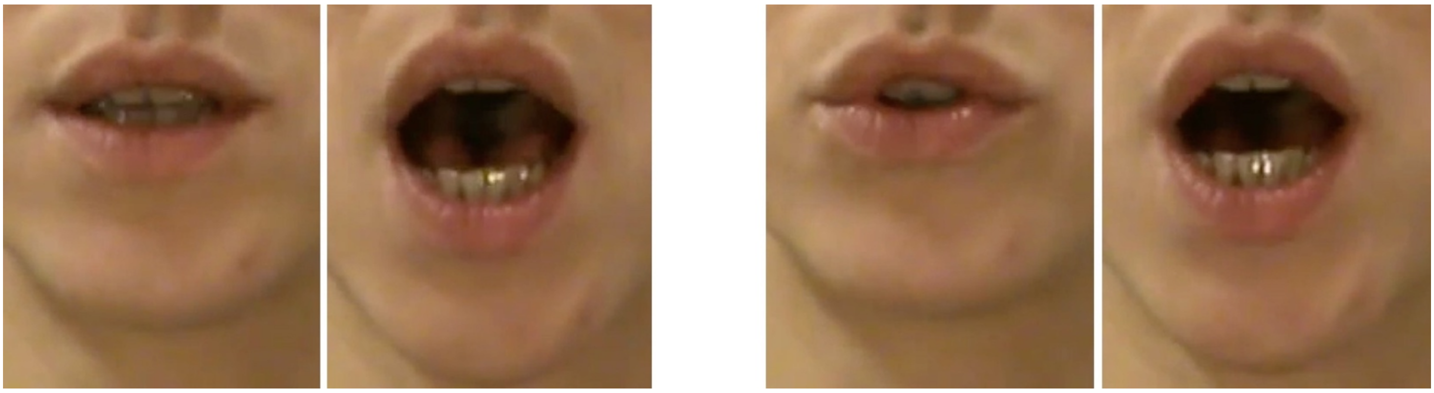
[b] [a] [g] [a]

Poor visibility; Low motor complexity



[d] [a] [n] [a]

Poor visibility; High motor complexity



[dz] [a] [dʒ] [a]

4. Results - Perception task

4.1 Overview – Mean Performances

To begin with a descriptive overview, the groups' total mean scores per condition were compared (two-sided, unpaired Wilcoxon Rank Sum test). As expected, adults abundantly outperformed both typical and apraxic children across conditions (cf. Tables 16 and 17).

Typically developing 4-to-5-years-olds and 7-to-9-years-olds obtained comparable scores across conditions ($p = .5$).

Conversely, they were more accurate than participants with apraxia in condition 1. This difference was significant when the comparison was made for the same chronological age (i.e. 7-to-9-years-olds *vs* apraxic participants: $p = 0.003$) and marginally significant when run for the same linguistic level (i.e. 4-to-5-years-old *vs* apraxic participants: $p = 0.05$).

An analogous trend characterized condition 2 (cf. Figure 7), but these differences were not significant ($p > 0.05$). However, as described in Section 4.5, such surface similarities in mean scores were determined by different patterns of association between the given variables.

Moreover, and significantly, this descriptive overview took into account both the 'same' and the 'different' trials, while only the latter (representing the actual experimental contrasts) were considered in the analyses described in Sections 4.4 and 4.5.

Significant values are reported in Tables 16 and 17 and Figures 6 and 7.

Figure 6 – Group comparison, total mean scores per condition, condition 1

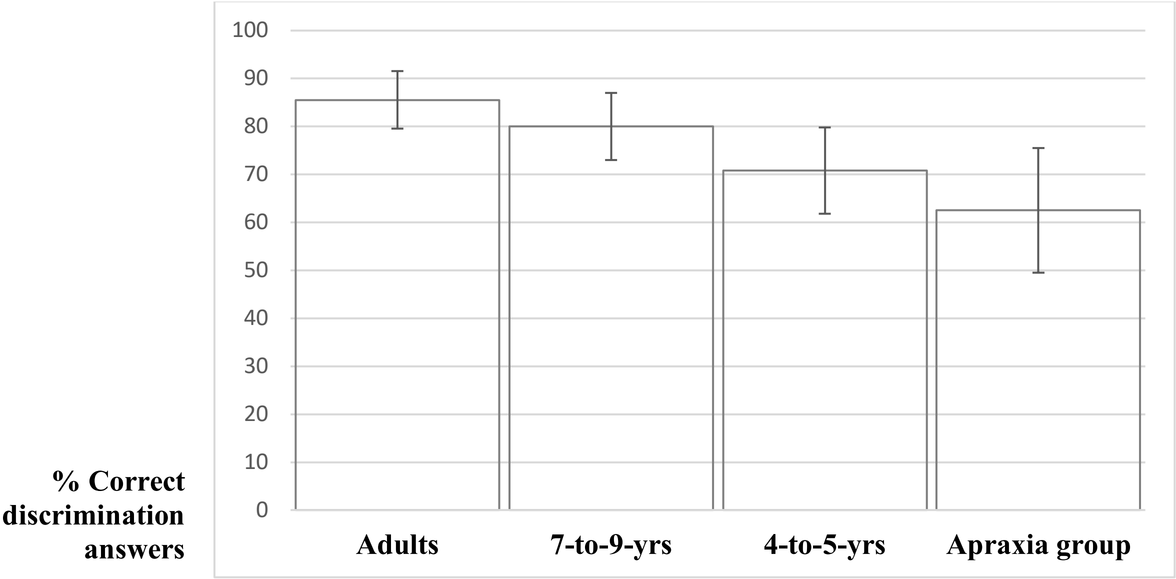


Table 16 – Group comparison, total mean scores per condition, condition 1

	Two-sided Unpaired Wilcoxon test
Adults – Typical children (4-to-5-years-olds)	W = 524.5; p-value = 0.006
Adults – Typical children (7-to-9-years-olds)	W = 507, p-value = 0.01
Adults – Apraxia group	W = 471, p-value = 1.6e-05
Typical children (4-to-5-years-olds) – Typical children (7-to-9-years-olds)	W = 233, p-value = .5
Typical children (4-to-5-years-olds) – Apraxia group	W = 267, p-value = 0.05
Typical children (7-to-9-years-olds) – Apraxia group	W = 301.5, p-value = 0.003

Figure 7 - Group comparison, total mean scores per condition,condition 2

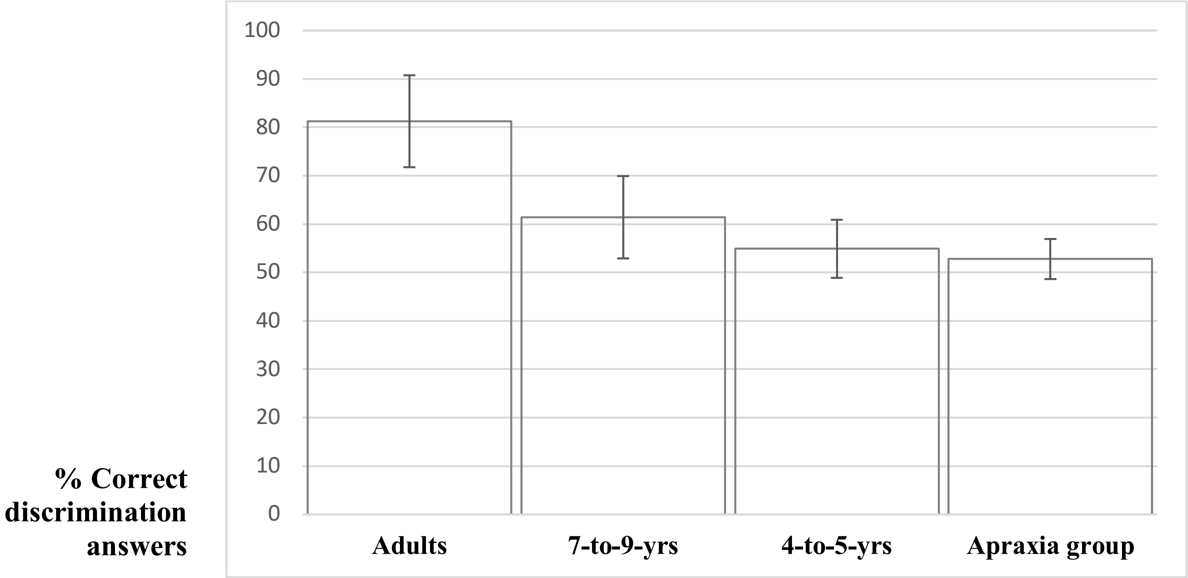


Table 17 – Group comparison, total mean scores per condition, condition 2

	Two-sided, unpaired Wilcoxon test
Adults – Typical children (4-to-5-years-olds)	W = 673.5, p-value = 7.4e-06
Adults – Typical children (7-to-9-years-ols)	W = 582, p-value = 0.0002
Adults – Apraxia group	W = 498.5, p-value = 1.6e-05
Typical children 4-to-5-years-olds – 7-to-9-years-ols	W = 242, p-value = 0.3
Typical children (4-to-5-years-olds) – Apraxia group	W = 254, p-value = 0.5
Typical children (7-to-9-years-ols) – Apraxia group	W = 262.5, p-value = 0.1

4.2 Associations between variables

Given that the distribution characterizing the groups was not normal and that the outputs of the perception tasks were binary in nature (i.e. 1 – right answer; 0 – wrong answer), the data were analyzed by means of Generalized Linear Mixed-effects Models, which are referred to in the next sections as Models 1 to 6 (cf. Quené & van der Bergh 2008 and Magezi 2015 on the appropriateness of such statistics in similar experiments).

The performance of each group was firstly examined by considering the scores obtained in the two conditions together (Section 4.3) and then by targeting them separately (Section 4.4 – condition 1 and Section 4.5 – condition 2). The outcome variable was DISCRIMINATION ACCURACY for all groups, while additional analyses were run on REACTION TIMES for adults. The predictor variables are specified case by case in the following sections. Participants were treated as the random factor.

4.3 Global perceptual profile: Model 1

With the preliminary aim of evaluating the groups' general abilities, scores pertaining to all experimental trials were examined together in Model 1.

The predictor variables were: AGE (z-scores); GENDER; ORDER OF ADMINISTRATION (i.e. condition 1-first or condition 2-first); CONDITION (1 or 2); PAIR-TYPE (i.e. 'same' *vs* 'different') and LINGUISTIC AGE (for the apraxia group, z-scores). It is worth noting that this first analysis did not evaluate the effect on accuracy and reaction times of the *trial-types* (cf. Table 11, page 71), but the effect on accuracy and reaction times of the *pair types* (i.e. 'same' as opposed to 'different').

Some general trends were observed (cf. Table 18). These differentiated, on the one side, apraxic from healthy participants of any age; on the other, adults from typical children.

Specifically, two main findings emerged:

(i) while, as it is normally expected, healthy participants – both children and adults – were more accurate in detecting the 'same' than the 'different' pairs ($p < 0.0001$ and $p < 0.05$ respectively), children with apraxia were not ($p > 0.05$). This can be explained by recalling that two different videoclips were used to create the 'same' pairings, so that identity was given through phonological categorization rather than strictly in terms of sound and vision (cf. Section 3.3, page 71). It can be concluded that, while the typical participants processed the stimuli through categorical-phonological judgements, the apraxia group was more sensitive to the contingent physical characteristics of each item. This first and fundamental divergence can be taken as evidence of a lower commitment to language-specific processing.

(ii) On the other hand, typically developing children, but not adults, displayed a very meaningful effect of condition on accuracy, performing better in the first than in the second ($p < 0.0001$ for preschooler; p < 0.001 for school-aged children). This was due to the strongly facilitating effect of the [rich visibility] trials, for which typical children obtained particularly high discrimination scores (i.e. 93% of mean correct answers in both groups). Notably, this trend was not detected in the apraxia group and this aspect appeared connected with attention-related factors (Section 4.4, pages 83-84).

The additional effect of gender emerged in the adult group where, across conditions, female participants obtained better accuracy scores (p = 0.04). Conversely, in this population, nor discrimination accuracy nor reaction times were influenced by the other predictors. In particular, no effect of the experimental condition was detected.

Finally, age did not affect the groups' performances and this was not surprising, due to the small internal variation characterizing the variable (cf. Section 2).

Table 18 – Model 1 (global scores)

Group	**Predicted variable**	**Significant predictors**	**Coefficient estimate**	**Standard error**	**z value**	**p value**
Adults	ACCURACY	TRIAL TYPE (Different *vs* Same)	0.50	0.25	2.05	0.04
		GENDER (Female *vs* Male)	-0.62	0.30	-2.02	0.04
TDs (4-to-5-years-old)		TRIAL TYPE (Different *vs* Same)	1.48	0.25	5.88	3.9e-09
		Condition (1 *vs* 2)	-1.00	0.25	-4.07	4.6e-05
TDs (7-to-9-years-old)		TRIAL TYPE (Different *vs* Same)	2.07	0.29	7.16	8.3e-13
		CONDITION (1 *vs* 2)	-0.94	0.26	-3.58	0.0003

4.4 Condition 1 – Rich *vs* poor visibility: Models 2 to 4

Before describing the analyses that were applied to each condition of the experiment, some preliminary considerations are in order.

As previously described, the phonological oppositions targeted in the two conditions were implemented by means of two CV pairings repeated in all possible combinations (cf. Table 11, page 71). While limiting the processing load, this schema introduced the risk of undesired effects due to repetition of the trials which, e.g., could lead the more skilled participants to lose their interest in the task or the less proficient ones to artificially increase their accuracy.

In order to control for this factor, the stimuli were pseudo-randomly ordered in lists containing the former occurrence of each 'different' trial-type (e.g. [ba] *vs* [ga]) in the first half and the latter (e.g. [ga] *vs* [ba]) in the second half. This allowed the experimenter to distinguish two sets of responses and to compare them (Paired Wilcoxon Rank Sum test).

These two sets of responses are respectively called set 1 and set 2 (cf. Table 11, page 71).

As reported in the following paragraphs, significant effects did arise, from this analysis, for some of the participants. As a consequence, responses in sets 1 and 2 were separately analyzed.

DISCRIMINATION ACCURACY in condition 1 was evaluated in Models 2 and 3, targeting SET 1 and SET 2 respectively. The predictor variables were constituted by the TRIAL TYPE (rich visibility *vs* poor visibility) and the ORDER OF ADMINISTRATION of the two conditions; the participants were treated as the random factor. The same analyses were repeated with REACTION TIMES for the adults group (Models 4 and 5). The results obtained are detailed in tables 19 to 21, where only Models highlighting significant values are reported.

To begin with, the considered variables did not affect accuracy in adults.

On the other hand, reaction times in both set 1 and 2 were significantly shorter for trials characterized by rich *vs* poor visibility (p = 0.01 for the set 1; p = 0.0001 for set 2; cf. Table 21). This seems attributable to the well-known strong facilitation yielded in noisy contexts by macroscopic visual cues.

Both groups of typical children displayed significant better discrimination for the rich visibility as opposed to the poor visibility trials (4-to-5-years-olds: p = 0.02 in set 1, p = 0.002 in set 2; 7-to-9-years-olds: p = 0.006 in set 1, p = 0.0006 in set 2; cf. Tables 19 and 20).

Conversely, participants with apraxia followed the trend displayed by typical children only once outliers were removed (in the number of 3) and limitedly to set 1 (p = 0.003; Table 19).

The lack of any effect of trial-type in set 2 was determined by a significant drop in accuracy concerning the rich visibility trials ([ga] *vs* [ba]: percentage change = -36%; p = 0.04, Paired Wilcoxon Rank Sum test). This behavior seems to highlight an atypical difficulty to maintain (sustained or selective) attention during this brief task which, in turn, suggests a difference in processing costs in this population as compared with typical participants[60].

To sum up, compared with the controls, children with apraxia displayed weaker effects of the rich visual cues: as measured by the rich visibility trials, speech-in-noise audiovisual discrimination seemed to be more effortful for them, even in the context of a brief task.

Concomitantly, accuracy for the poorly visible trials ([da] *vs* [na] – set 1 and [na] *vs* [da] – set 2) was uniformly below chance levels in this group (47% in both set 1 and 2, SD 51% and 49% respectively), while it was overall above chance levels in typical children (55% for 4-to-5-years-olds, SD 36% and 58% in 7-to-9-years, SD 45% – mean scores across sets).

This reveals that, in apraxic participants, speech-in-noise perception turned out to be inefficient when not supported by rich visual cues.

---

[60] Who, in the two sets of rich visibility trials, uniformly displayed high accuracy: 91% in set 1 and 96% in set 2 for 4-to-5-years-olds; 96% in set 1 and 91% in set 2 for 7-to-9-years-olds.

Table 19 – Model 2 (Discrimination accuracy, condition 1, set 1)

Group	Predicted variable	Significant predictors	Coefficient estimate	Standard error	z value	p value
TDs (4-to-5-years-olds)	ACCURACY	TRIAL TYPE (poor *vs* rich visibility)	2.14	0.95	2.27	0.02
TDs (7-to-9-years-olds)		TRIAL TYPE (poor *vs* rich visibility)	3.00	1.10	2.72	0.006
CAS		TRIAL TYPE (poor *vs* rich visibility)	22	7.60	3.00	0.003

Table 20 – Model 3 (Discrimination accuracy, condition 1, set 2)

Group	Predicted variable	Significant predictors	Coefficient estimate	Standard error	z value	p value
TDs (4-to-5-years-olds)	ACCURACY	TRIAL TYPE (poor *vs* rich visibility)	23	7.40	3.12	0.002
TDs (7-to-9-years-olds)		TRIAL TYPE (poor *vs* rich visibility)	3.25	0.95	3.42	0.0006

Table 21 – Model 4 and 5 (Reaction Times in adults, condition 1, set 1 and 2)

Group	Predicted variable	Significant predictors	Coefficient estimate	Standard error	z value	p value
Adults Set 1	ACCURACY	TRIAL TYPE (poor *vs* rich visibility)	-139	53	-2.63	0.01
Adults Set 2		TRIAL TYPE (poor *vs* rich visibility)	-356	76	-4.5	0.0001

4.5 Condition 2 – High *vs* low motor complexity: Model 5 and 6

Accuracy and reaction times in condition 2 were analyzed in Models 6 and 7 (targeting set 1 and 2 respectively), where predictors were analogous to those considered in condition 1 (cf. page 83). The results are detailed in Table 22, where the only Model highlighting significant values is reported (Model 6).

No significant tendencies emerged in the adults group; differently from condition 1, the given variables did not even affect the reaction times. Such result can be explained by the more homogeneous nature of condition 2 which, albeit not affecting discrimination precision for mature participants, did not include oppositions as easy to detect as those comprised in condition 1 (i.e. [ba] *vs* [ga]).

On the other hand, typically developing children (both preschoolers and school-aged children) displayed a strong reduction in accuracy for trials characterized by high motor complexity ($p = 0.0007$ and $p = 0.003$, respectively).

The additional effect of the order of presentation of the conditions was observed in preschoolers, who obtained better scores when the experiment began with the second task ($p < 0.0001$). This point seems to depend on attention-related effects: since condition 2 was more complex, young children dealt with it better at the beginning of the experiment, when they relied on a fresh attentional state.

The described pattern held for set 1, but not for set 2 as, in the latter case, significant differences were not detected. Comparing the two sets, an accuracy increase was displayed by both groups (percentage of growth = 33%, $p = 0.02$ in 4-to-5-years-olds; 46%, $p = 0.06$ in 7-to-9-years-olds).

In other words, typical children showed a significantly greater difficulty in dealing with complex (*vs* easy) speech motor schemas but, when the same trials were repeated, displayed a rapid recovery in discrimination precision, supposedly primed by the previous occurrence.

As previously mentioned (Sections 3.1, page 66-67; Section 3.2, pages 69-70), a comparison between the performances obtained in this task with an acoustic baseline (i.e. with auditory-only input) was not provided in this study. Therefore, one cannot directly determine whether the discrimination performances observed in the children derived from an auditory or from an audiovisual processing of the test trials. Thus, acoustic-related factors might have been at play (cf. Chap. 5, Section 2, for a discussion of such aspect). At any rate, the highly specific pattern of correlations between perception and production abilities described in Section 7 support the conclusion that the results observed are connected with the different sensorimotor profiles characterizing the participants. In fact, as it will be described, the children's production abilities only positively correlated with the perception of low complexity motor schemas, while a reverse pattern was observed in adults (whose production abilities only positively correlated with the perception of high complexity motor schemas).

Further studies will be necessary to evaluate the relative importance of the acoustic and the sensorimotor aspect in comparable conditions but, in itself, the significant effect of the latter is globally supported by the results obtained in this dissertation.

Very different tendencies characterized apraxic participants as, in this group, accuracy in condition 2 was uniformly below chance levels, irrespectively of the degree of motor complexity of the consonants (40.3% mean correct answers across sets *vs* 52% in 4-to-5-years-olds and 50% in 7-to-9-years-olds).

On the one side, this result confirms the trend observed in condition 1: when the support of rich visibility is not provided, apraxic participants cannot efficiently rely upon an auditory signal masked by a 0 dB white noise. On the other, it introduces a remarkable difference between typical and apraxic children, as no effect of motor complexity is observed in the latter. As it will be described in Section 7, taken together with the observed correlation patterns, this data point to lack of sensorimotor perceptual contribution in this population: the group's perceptual abilities did not correlate with any of the production scores while, crucially, they correlated with type and duration of the speech therapy.

Table 22 – Model 6 (condition 1, set 1)

Group	**Predicted variable**	**Significant predictors**	**Coefficient estimate**	**Standard error**	**z value**	**p value**
TDs (4-to-5-years-olds)	ACCURACY	TRIAL TYPE (High *vs* Low motor complexity)	15	4.5	3.4	0.0007
		ORDER (1-first to 2-first)	37	8.7	4.3	2.16e-05
TDs (4-to-5-years-olds)		TRIAL TYPE (High *vs* Low motor complexity)	15	5.1	3	0.003

5. Methods – Production tasks

5.1 Rationale

The production tasks aimed at evaluating speech motor control.

As briefly mentioned in Chap. 2 (Section 3.1, page 31), a problematic aspect in research targeting the developmental relationship between perception and production is that the assessment of production skills is often rather coarse-grained. Most commonly, it relies on eliciting the articulation of some target sounds or by calculating the Percentage of Consonants Correct over a speech sample. However, auditorily adequate productions are not necessarily realized by children by means of mature speech movements (cf. Chap. 3, Foreword, page 49) and this raises an issue, particularly with reference to participants affected by speech disorders: experimental studies addressing the relationship between perception and production can only be fully accurate when instrumental observations of articulation are gathered.

Thus, although without the support of articulatory data, particular attention was devoted in this research to obtain a precise behavioral assessment of production skills. To this purpose, a highly sensitive instrument is represented by the Maximum Performance Rate Test, whose reliability in distinguishing healthy from speech-disordered participants has been consistently assessed[61] and whose heuristic value applied to Italian apraxic children has been recently confirmed (Chilosi et al. 2015).

Since sensorimotor knowledge is strictly connected with proficiency in articulation (the former deriving from the latter), the Maximum Performance Rate Test allows for its quantification. Precisely, it measures the upper limits of the participant's articulatory dexterity, thereby providing behavioral indexes that are firmly rooted in the individual's cognitive-motoric profile.

In its 'alternating' version (as used in this experiment[62]), the task is performed by asking to repeat as fast as possible a phonological sequence characterized by a clear-cut differentiation of articulatory movements. A commonly used stimulus is the pseudo-word/pataˈka/: albeit composed by very simple motor schemas (i.e. voiceless stops at easy places of articulation), this motoric pattern is challenging when repeated rapidly.

---

[61] E.g. Thoonen (1996); Thoonen et al. (1999); Williams & Stackhouse (2000).

[62] The Maximum Performance Rate task is defined 'alternating' if the experimental item(-s) contain(-s) different consonants, 'sequential' if the same consonant is repeated.

5.2 Stimuli and procedure

In line with the methodology in Chilosi et al. (2015), the test was implemented into a count-by-time procedure, i.e. the sequence was repeated for a given time (20 seconds).

Younger participants repeated the pseudo-word/pataka/; adults were also assessed by means of an additional trial that, mirroring the perception task, consisted of motorily complex consonants in alternating pattern (/ɲaraˈdʒa/).

Participants were instructed to repeat the item as fast as possible from the ‘start’ to the ‘stop’ signals given by the experimenter, with the additional suggestion to try to find their individual trade-off between accuracy and rapidity (i.e. they were explicitly allowed to *slightly* slow down the rate if not managing to produce the sequence accurately). Once again, a ludic formulation of the task was designed for young participants.

Each session took place in a quiet room and was filmed for off-line coding. The data were cross-analyzed by the experimenter in two separate sessions based on perceptual and spectrogram analysis.

Three indexes were calculated: (i) diadochokinetic rate, i.e. the ability of producing rapid and alternating movements, computed as the total number of trisyllables produced in the given time; (ii) accuracy, the ratio between correct repetitions and total repetitions; (iii) phonological inconsistency, the ratio between the number of different productions and the number of total productions (i.e. repetition-type and repetition-tokens).

Crucially, this procedure was repeated in two conditions, proposed in a counterbalanced order: normal modality of production and auditory-masking. In the latter case, participants performed the test while wearing headphones administrating white noise at 70dB. By this mean, the sound of one’s voice (i.e. the auditory feedback, a primary resource to monitor one’s performance in challenging contexts) became substantially unexploitable, thus eliciting greater reliance on the proprioceptive control of speech movements.

6. Results – Production tasks

6.1 Overview

Overall, the three indexes did not demonstrate the same reliability.

The measure most robustly differentiating the groups was represented by the diadochokinetic rate. In this respect, an age-related pattern emerged in typical participants: adults greatly outperformed both groups of children ($p < 0.0001$), and the older children outperformed the younger ($p = 0.0004$ in normal production; $p < 0.0001$ in auditory masking). On the other hand, apraxic participants appeared significantly less skilled then adults ($p < 0.0001$) and when compared with typical children of the same chronological age ($p = 0.007$ in normal production; $p = 0.001$ in auditory masking).

Conversely, accuracy turned out to be less informative. It significantly differentiated adults from apraxic children (p = 0.01 in normal production; p < 0.0001 in auditory masking) and from typically developing 7-to-8-years-olds (p = 0.0004 in normal production; p < 0.0001 in auditory masking); however, no dissimilarity was detected between the adults and the younger typical participants. This finding can be explained from a developmental perspective: while 4-to-5-years-olds displayed a low diadochokinetic rate (so low as to be comparable with that of the apraxic group), older children were significantly more rapid: this increase in velocity caused a significant drop in accuracy, in contrast to the younger children. Thus, additional factors seem to influence this result, quite independently from the articulatory perspective (a further description of this aspect will be given in Section 7, page 97).

More unexpectedly, accuracy did not relevantly distinguish typical *vs* apraxic children. Given the severity of this disorder, this should not be interpreted in the sense that these two populations were comparably accurate. Rather (and in agreement with what described in the previous paragraph), the Maximum Performance Rate may not be an optimal instrument of comparison in this respect: since this procedure is conceived to push the participants to the limits of *their own* articulatory performance, it can easily induce a loss in accuracy in young participants.

Similar considerations hold with reference to inconsistency. This value coherently differentiated adults from all other participants, but not typical from apraxic children. In this connection, an important aspect has to be underlined: differently from adults, typical children displayed difficulties to retain the pseudo-word (/pataka/) that, in the majority of the cases, they confounded with/pakata/; on the other hand, the apraxic participants' substitutions were much more atypical (e.g. [takapkta], [kapalal]). Thus, it seems conceivable that phonological memory skills and general ease of articulation could have determined the low performance of typical children, while more severe motor factors were at play in the apraxia group.

It should be remarked, with respect to this last index, that significant results only emerged in a coherent pattern when outliers were removed from the groups[63] (while their presence did not influence the other analyses): an aspect pointing to the large interindividual variability existing at this level (i.e. the degrees of freedom underlying one's patterns of productions, cf. Chap. 3, Section 2.1, page 52).

A last point to consider is that the apraxic participants are periodically evaluated by means of the Maximum Performance Rate during clinical follow-up; typical participants, conversely, are not used to it. An effect of familiarity with the procedure (albeit minimal) cannot be discarded.

---

[63] 6 adults; 1 participant from the apraxia group; 10 preschoolers and 7 school-aged children.

Figure 8 - Group comparison, mean scores, diadochokinetic rate (normal mode of production)

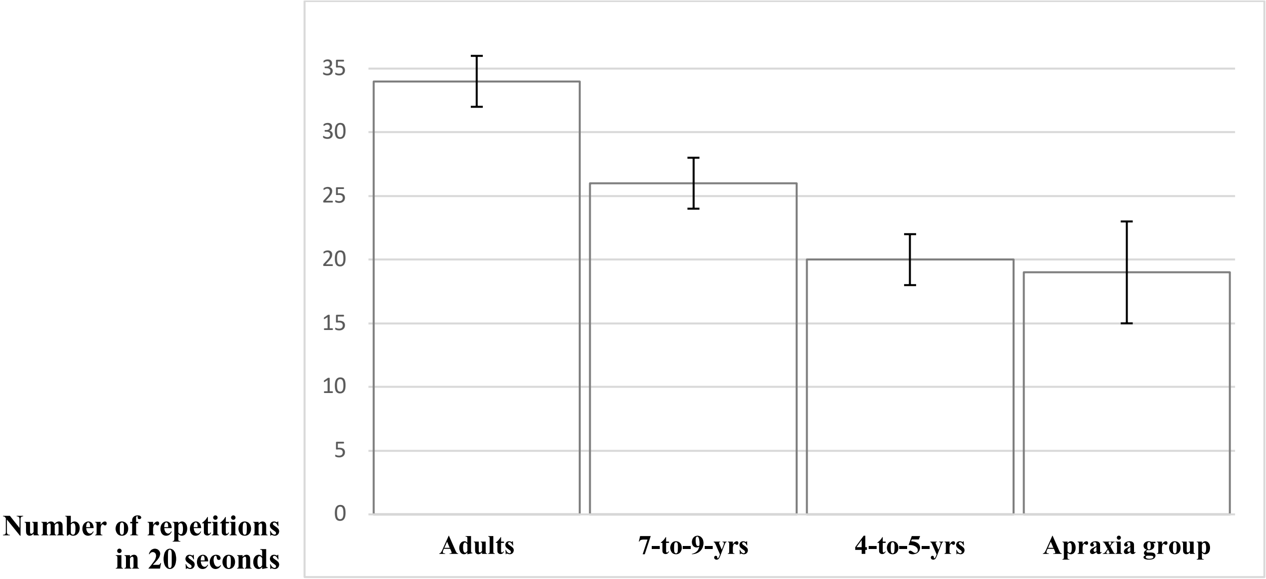


Figure 9 - Group comparison, mean scores, accuracy (normal mode of production)

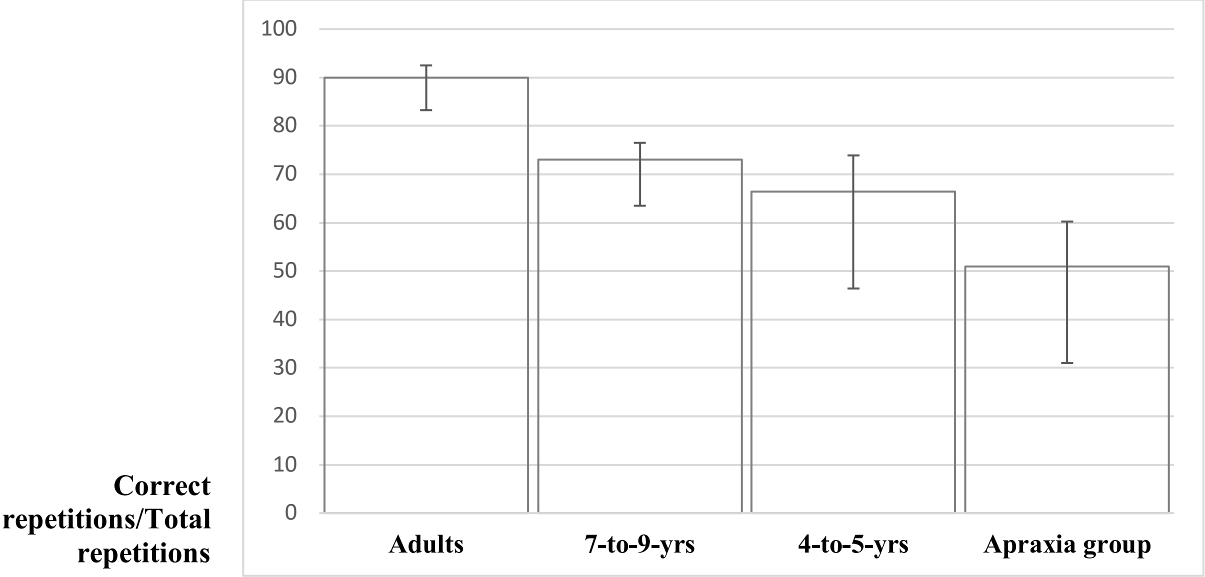


Figure 10 - Group comparison, mean scores, inconsistency (normal mode of production)

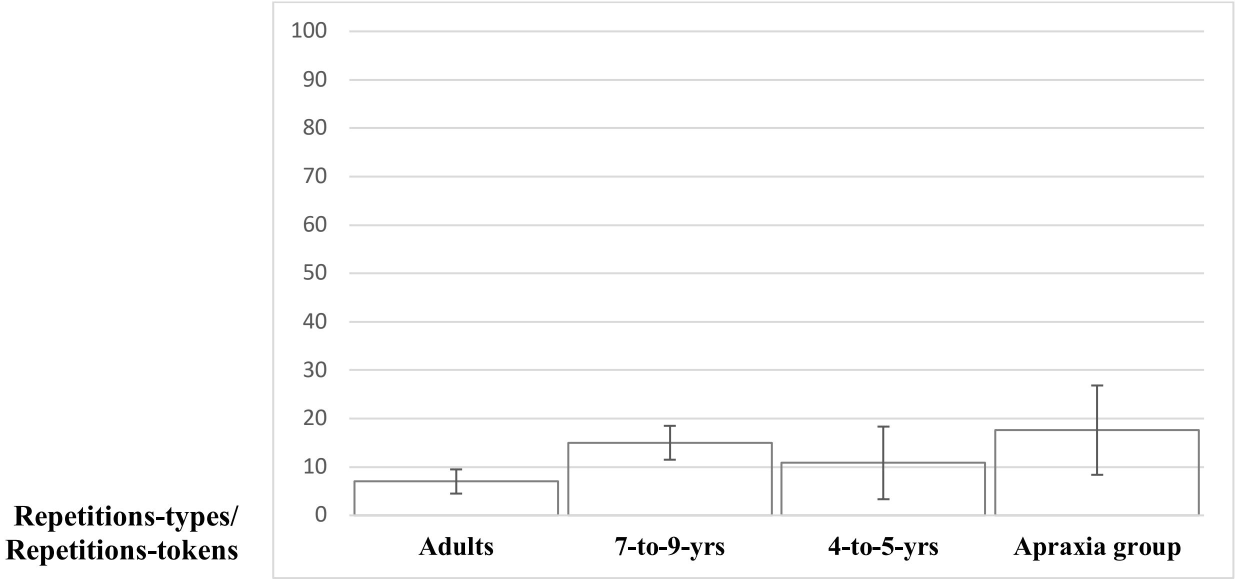


Table 23 – Group comparison, mean scores (normal mode of production)

	Two-sided, unpaired Wilcoxon test
Diadochokinetic rate	
Adults – Typical children (4-to-5-years-olds)	W = 447; p-value = 9.406e-08
Adults – Typical children (7-to-9-years-olds)	W = 625; p-value = 4.749e-08
Adults – Apraxia group	W = 328; p-value = 1.708e-06
Typical children 4-to-5-years-olds – 7-to-9-years-olds	W = 51.5; p-value = 0.0004
Typical children (4-to-5-years-olds) – Apraxia group	N.S.
Typical children (7-to-9-years-olds) – Apraxia group	W = 192, p-value = 0.007
Accuracy	
Adults – Typical children (4-to-5-years-olds)	N.S.
Adults – Typical children (7-to-9-years-olds)	W = 499; p-value = 0.0004
Adults – Apraxia group	W = 230; p-value = 0.01
Typical children 4-to-5-years-olds – 7-to-9-years-olds	N.S.
Typical children (4-to-5-years-olds) – Apraxia group	N.S.
Typical children (7-to-9-years-olds) – Apraxia group	N.S.
Inconsistency (outliers removed)	
Adults – Typical children (4-to-5-years-olds)	W = 69, p-value = 0.0003
Adults – Typical children (7-to-9-years-olds)	W = 104, p-value = 0.000108
Adults – Apraxia group	W = 75, p-value = 0.04
Typical children 4-to-5-years-olds – 7-to-9-years-olds	N.S.
Typical children (4-to-5-years-olds) – Apraxia group	N.S.
Typical children (7-to-9-years-olds) – Apraxia group	N.S.

Figure 11 - Group comparison, mean scores, diadochokinetic rate (auditory masking)

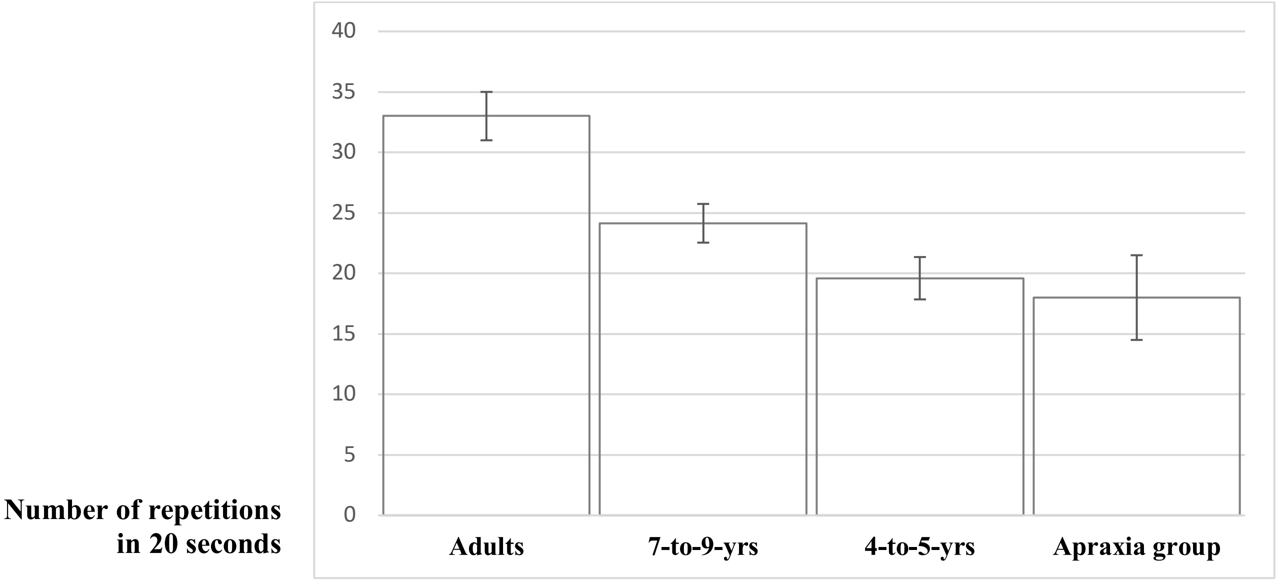


Figure 12 - Group comparison, mean scores, accuracy (auditory masking)

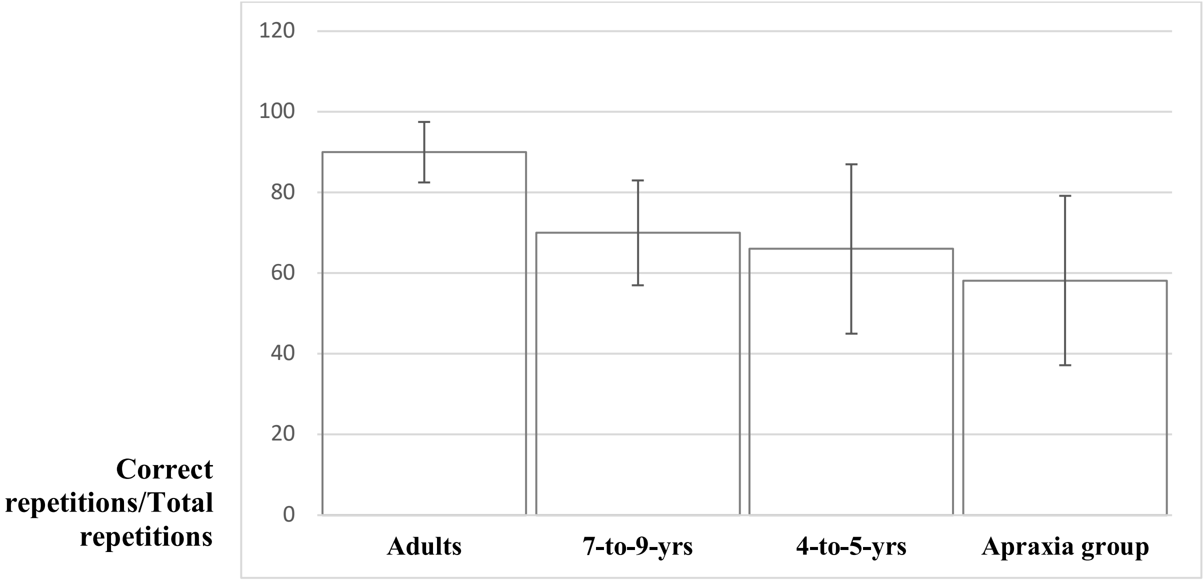


Figure 13 - Group comparison, mean scores, inconsistency (auditory masking)

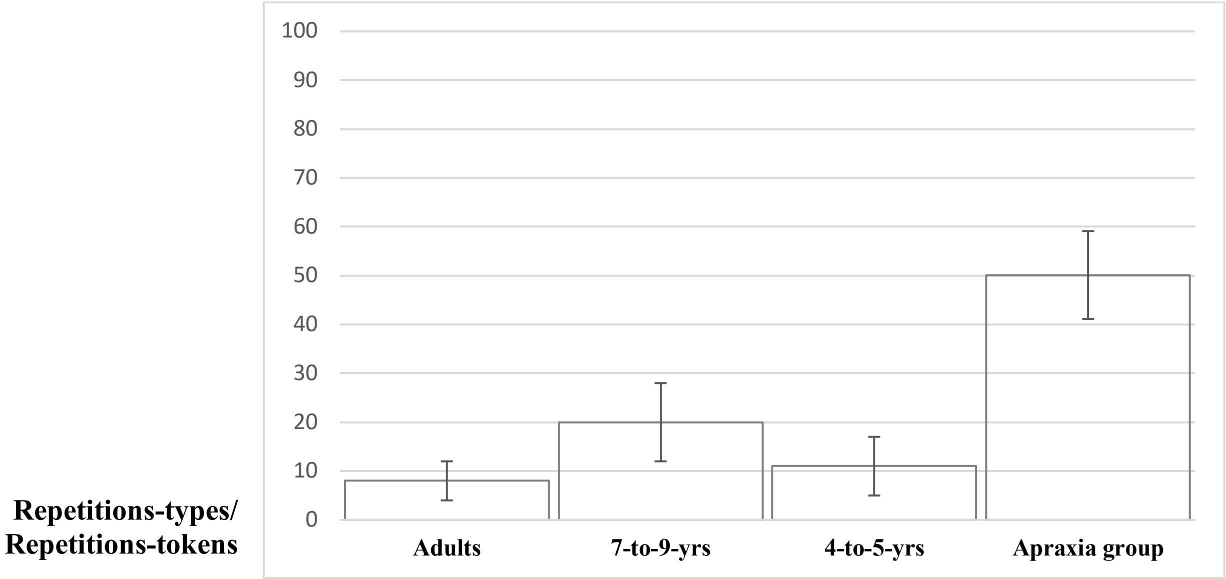


Table 24 - Group comparison, means scores (auditory masking mode of production)

	Two-sided, unpaired Wilcoxon test
Diadochokinetic rate	
Adults – Typical children (4-to-5-years-olds)	W = 464, p-value = 3.913e-08
Adults – Typical children (7-to-9-years-olds)	W = 605.5, p-value = 3.388e-09
Adults – Apraxia group	W = 377, p-value = 3.07e-07
Typical children 4-to-5-years-olds – 7-to-9-years-olds	W = 377, p-value = 3.07e-07
Typical children (4-to-5-years-olds) – Apraxia group	N.S.
Typical children (7-to-9-years-olds) – Apraxia group	W = 227, p-value = 0.001
Accuracy	
Adults – Typical children (4-to-5-years-olds)	N.S.
Adults – Typical children (7-to-9-years-olds)	V = 435, p-value = 2.695e-06
Adults – Apraxia group	V = 435, p-value = 2.695e-06
Typical children 4-to-5-years-olds – 7-to-9-years-olds	N.S.
Typical children (4-to-5-years-olds) – Apraxia group	N.S.
Typical children (7-to-9-years-olds) – Apraxia group	N.S.
Inconsistency (outliers removed)	
Adults – Typical children (4-to-5-years-olds)	W = 22, p-value = 6.859e-06
Adults – Typical children (7-to-9-years-olds)	W = 51, p-value = 2.919e
Adults – Apraxia group	W = 138, p-value = 0.004
Typical children 4-to-5-years-olds – 7-to-9-years-olds	N.S.
Typical children (4-to-5-years-olds) – Apraxia group	N.S.
Typical children (7-to-9-years-olds) – Apraxia group	N.S.

6.2. Associations between variables

The factors potentially influencing the groups' performances were analyzed in a series of generalized linear mixed-effects models (Models 7 to 9), separately targeting each of the production indexes (DIADOCHOKINETIC RATE, ACCURACY and INCONSISTENCY) as the predicted variable. The predictors were represented for all groups by: CONDITION (either normal mode of production or auditory masking); ORDER OF PRESENTATION of the conditions; AGE (z-scores) and GENDER. In addition, the possible effect of the TRIAL-TYPE was evaluated for the adults group (either easy or difficult motor schema), while the LENGTH and the TYPE OF SPEECH THERAPY were considered for the apraxia group. Participants were treated as the random factor.

As shown in Table 25, the considered variables carried statistically significant effects only in the adults group: diadochokinetic rate was diminished by both the auditory masking condition and the trials entailing complex motor schemas (i.e./ɲaradʒa/); the latter also increased inconsistency.

While the logic of such results is self-evident, their absence in the groups of children needs to be discussed. In particular, the lack of significant effects of the auditory masking paradigm could be explained either by a very reduced articulatory capacity, or (as previously mentioned with reference to the mean scores – Section 6.1, page 90) by a difference in cognitive strategies: while adults (explicitly or implicitly) reduced their velocity of articulation in order to better control their performance in the absence of the auditory feedback, children did not and thus displayed unaltered inconsistency.

Table 25 – Model 7 to 9 (production tasks)

Group	**Predicted variable**	**Significant predictors**	**Coefficient estimate**	**Standard error**	**z value**	**p value**
Adults	RATE	TRIAL-TYPE (difficult *vs* easy motor scheme)	3,49	0,39	9,02	4,84e-14
		CONDITION (auditory masking *vs* normal mode)	0,85	0,39	2,18	0,03
Adults	INCONSISTENCY	TRIAL-TYPE (difficult *vs* easy motor scheme)	0,02	0,01	-2,12	0,03

7. Correlations between production and perception scores

The scores obtained by each group in the perception and production procedures were jointly analyzed in order to detect possible correlational patterns (Kendall's Tau correlation).

Perception skills were represented by the accuracy of discrimination for each trial-type of the experiment (with the addition of reaction times in adults): [RICH VISIBILITY, LOW MOTOR COMPLEXITY], i.e. [ba] *vs* [ga]; [POOR VISIBILITY, LOW MOTOR COMPLEXITY], i.e. [da] *vs* [na]; [POOR VISIBILITY, HIGH MOTOR COMPLEXITY], i.e. [dza] *vs* [dʒa] (cf. Table 11, page 71).

These values were analyzed in combination with the three production indexes, i.e. DIADOCHOKINETIC RATE, ACCURACY and INCONSISTENCY. As detailed in Section 5.1 and 5.2, the latter represent measures of sensorimotor production proficiency. Recalling the fact that speech sensorimotor knowledge stems from speech production, these indexes can legitimately be intended as representing the participants' degree of sensorimotor knowledge (and such perspective is particularly useful in order to interpret the results of the correlations reported below).

Supplementary information was considered for the apraxic group, i.e. NUMBER and TYPE OF VOCAL MOTOR SCHEMAS making part of their phonological inventories, plus DURATION and TYPE OF THE INDIVIDUAL SPEECH THERAPY (cf. Table 7, page 65).

To begin the examination of the results from the adult benchmark, in these participants a very specific and consistent pattern emerged, were proficiency in speech motor control (i.e. diadochokinetic rate) turned out to be strongly correlated with accuracy in the discrimination of the [(poor visibility), high motor complexity] trials (i.e. [dza] *vs* [dʒa] and [dʒa] *vs* [dza]). This result held in connection with both the normal and the auditory masking paradigm ($p = 0.008$ and $p = 0.002$, respectively – cf. Table 26). In young adults, the degree of sensorimotor proficiency correlated with the capacity to deal with complex speech motor schemas during in-noise discrimination.

From a general perspective, this result confirms the known association between production and perception in young adults.

From a more specific point of view, its selectivity (i.e. the fact that the correlation was only significant in connection with complex motor schemas) tentatively suggests that the particularly vivid sensorimotor contribution that is observed in noisy contexts might be especially strong, during adulthood, for the sounds produced by complex articulatory movements. The actual existence of such effect and its possible reasons remain to be verifyied and explained by future research.

For what pertains to typical children, 4-to-5-years-olds displayed the same specific association between proficiency in speech motor control (i.e. diadochokinetic rate) and discrimination accuracy for [(poor visibility), high motor complexity] trials when the normal modality of production was considered ($p = 0.03$ - cf. Table 27).

Conversely, when perception scores were measured against the auditory masking condition, a different trend emerged. Namely, the diadochokinetic rate only correlated with accuracy for [low motor complexity] stimuli, irrespectively of their degree of visibility (Table 27). It is worth noting that the performance obtained in the absence of auditory feedback represents a stricter index of speech motor control abilities, focusing on the bare capacity to rely *solely* upon sensorimotor knowledge of the speech movements. Hence, once production skills were rigorously measured, the association between production and perception in this group pertained exclusively to the easy motor schemas.

This finding can be interpreted on developmental grounds: since speech sensorimotor knowledge is not mature in young children, a significant connection between production and perception is only truly observable for the easy motor schemas of their target language. This data supports the hypothesis in Section 4.5 (pages 86-87), i.e. that the lower discrimination accuracy displayed by these participants for [high complexity] *vs* [low complexity] speech motor schemas is connected with their immature sensorimotor development. In other words: if no connection is observed between production and perception for [high complexity] trials, then sensorimotor knowledge could not have significantly contributed to their processing: their degree of sensorimotor development did not help them to discriminate the affricates.

Overall, older typical children confirmed this logic. However, their performances appeared related in a more idiosyncratic fashion, i.e. perception resulted associated to the general cognitive effects which characterized their production performance (cf. Section 6.1).

Specifically, accuracy discriminating the [(low visibility), high motor complexity] trials did not correlate with the production variables (not even in the normal modality of production, as in 4-to-5-years-olds), while accuracy with the [low motor complexity] trials, irrespectively of rich or poor visibility, was inversely correlated with the score of production inconsistency (cf. Table 28).

In order to clarify this point, it has to be recalled that the production performance in this group was strongly influenced by a particular factor, connected with age-related increased articulation rapidity: namely, their difficulty to find a good trade-off between velocity and accuracy/inconsistency (cf. page 90). In connection with this point, the associations among inconsistency and [low motor complexity, richly and poorly visible] motor schemas show that the participants who could find a better compromise between rapidity and consistency also obtained better scores in discrimination. This capacity may depend on either an explicit strategy or on implicit difficulties to deal with recently augmented articulation skills. Whatever the case, these findings underline the important role that general cognitive control can play in children research.

Finally, no significant association among discrimination ability and the production scores emerged in the participants with apraxia of speech. Furthermore, no relevant interactions were detected among perceptual skills and the number or type of consonants mastered by each participant. In a nutshell, production and perception were not connected in this group. The interpretive hypothesis put forward in Section 4.5 (page 87) appears supported: sensorimotor knowledge could not have yield any significant contribution to perception in this group.

On the other hand, their discrimination accuracy with [poor visibility, low motor complexity] trials was significantly associated with the duration and type of speech therapy. Precisely, discrimination was more accurate when therapy began earlier ($p < 0.01$) and when the P.R.O.M.P.T. method had been more used ($p < 0.0001$ – Table 29). This latter finding has a critical relevance since, as previously mentioned (page 61), P.R.O.M.P.T. is a type of rehabilitation crucially based on the explicit effort to establish a connection between articulatory movements, proprioceptive sensations and auditory consequences. Hence, the more these aspects were re-connected, the greater the contribution that sensorimotor knowledge could grant during the speech-in-noise task (limitedly to easy motor schemas, as in typical children).

Worthwile to remark, these data also indirectly confirm the broad-spectrum effectiveness of the P.R.O.M.P.T. therapy.

Table 26 – Adults, production-perception correlations

Normal production

Association	Kendall Tau
Rate (simple motor scheme) – discrimination [+motor complexity]	z = 2.6, p-value = 0.008
Rate (complex motor scheme) – discrimination [+motor complexity]	z = 2, p-value = 0.03
Auditory masking	
Rate (simple motor scheme) – discrimination [+motor complexity]	z = 3, p-value = 0.002

Table 27 Typically developing children (4-to-5-years-old), production-perception correlations

Normal production

Association	Kendall Tau
Rate – discrimination [+motor complexity]	z = 2, p-value = 0.03
Auditory masking	
Rate – discrimination [poor visibility]	z = 2, p-value = 0.02
Rate – discrimination [rich visibility]	z = 2, p-value = 0.02

Table 28 – Typically developing children (7-to-9-years-old), production-perception correlations

Normal production

Association	Kendall Tau
Rate – discrimination [rich visibility]	z = 2, p-value = 0.009
Inconsistency – discrimination [poor visibility]	z = -2, p-value = 0.009
Auditory masking	
Rate – discrimination [rich visibility]	z = 2, p-value = 0.006
Inconsistency – discrimination [rich visibility]	z = -2, p-value = 0.04

Table 29 – Childhood Apraxia of Speech, clinical history – perception correlations

Normal production

Association	Kendall Tau
Years of therapy discrimination [poor visibility]	z = 3, p-value = 0.004
Prompt – discrimination [poor visibility]	z = 4, p-value = 3.76e-05
Years of therapy discrimination [poor visibility]	z = 2, p-value = 0.005
Prompt – discrimination [poor visibility]	z = 4, p-value = 3.76e-05

GENERAL DISCUSSION

1. Critical evaluation of the results

While some specific issues pertaining to the production task have been exhaustively discussed in Chap. 4 (Section 6.1, 6.2 and 7), the perception task deserves a specific discussion.

As described in Section 3.1 of Chapter 4, some relevant constraining factors determined a series of admittedly limiting choices, essentially aiming at reducing the risk of a massive loss of data with the apraxic participants. In this connection, the use of a single audiovisual procedure caused a major limitation of the extendibility of the results, since the measurement of the audiovisual benefit for each participant was not provided (as assessable against an acoustic baseline, i.e. the administration of the same stimuli in the auditory-only modality). Nevertheless, this did not prevent the attainment of some relevant results.

Albeit addressed on a case-by-case basis during the presentation of the results (Chapter 4, Sections 4.4 and 4.5), the relationship between such methodological limitation and the data obtained deserves a more systematic and clarifying discussion, targeting each condition of the task separately.

Table 30 – Condition 1: Summary

	Parameter	Description	
Condition 1	Visibility	Speech-in-noise perception of CV syllables as supported by rich *vs* poor visual cues	
Trials ('same' trials excluded)			
Low motor complexity		Rich visibility	Poor visibility
Set 1		[ba] *vs* [ga]	[da] *vs* [na]
Set 2		[ga] *vs* [ba]	[na] *vs* [da]

In Condition 1 (cf. Table above), consonants characterized by low motor complexity were administered at 0 dB white noise and with the support of rich *vs* poor visual cues.

In this context, the actual comparison of audiovisual *vs* auditory processing in the various types of participants was left to a future experiment, which will consist in the measurement (in comparable participants) of the auditory-only baseline in the perception of the relevant speech sounds (allowing for an estimation of the visual enhancement). At any rate, giving the sharp contrast in visibility characterizing the stimuli, together with the fact that perceptibility of the visual cues for [da] *vs* [na] has been discarded for children (cf. Lalonde & Holt, 2015; Jerger et al., 2018) and has not been demonstrated for adults (Files et al., 2015), it was possible to assume that, if a facilitation for [ba] *vs* [ga] as opposed to [da] *vs* [na] emerged, this would have been due to its comparatively (very) rich visibility.

A coherent pattern emerged:

(i) Adults did not display significant effects of the trial-type on discrimination accuracy, signaling the fact that they were fully able to process the stimuli, irrespectively of the visual support given. On the other hand, they demonstrated significantly more rapid reaction times in correspondence with rich visual cues, confirming perceptual enhancement in conditions of sharp visibility.

(ii) Typical children performed significantly better with richly visible as opposed to poorly visible oppositions, confirming the trend observed in former literature.

(iii) Like the typical children, the apraxic participants displayed facilitation for rich as opposed to poor visibility trials. However, this held for Set 1 only, due to significant drop in attentional level (hence, discrimination accuracy) in Set 2 (cf. Chap. 4, Section 4.4, page 83-84).

Differently from the typical controls, they scored uniformly below chance levels when the visual cues were poor and the only source of information was represented by the residual auditory signal (47% in both set 1 and 2, SD 51% and 49% respectively, cf. page 84).

Taken together, such results lead to conclude that the apraxic group: (a) unlike typical children, displayed limited (albeit detectable) support from rich visual cues; (b) were not able to discriminate the poorly visible stimuli on mere auditory bases.

With reference to the former aspect, it is however important to underline that, if an audiovisual task provides visual cues as salient and macroscopic as those characterizing [ba] *vs* [ga], one cannot unproblematically assume that the children participants elaborated them on linguistic grounds. In other words, such cues are so sharply evident as to be detectable on the basis of sheer domain-general information, i.e. information that can be used to distinguish any type of visual contrast, irrespectively of its nature (e.g. overt differences in the mouth's shape). Therefore, albeit highlighting an at least partial capacity to rely on the rich visual cues administered, the results of this dissertation do not grant that the apraxic children were actually able to process them *via* analysis of the articulatory movements. This latter aspect encourages further research explicitly aiming at assessing whether and to what extent this population can exploit the articulatory information.

Moreover, the apraxic group's generalized difficulties with the poor visibility trials display that, differently from typical children, the auditory processing of consonants masked by a 0 dB white noise is essentially ineffective in this population.

To sum up the data pertaining to condition 1, our results add new evidence on speech perception in Childhood Apraxia of Speech, as tested for the first time with a degraded auditory signal and with the support of visual cues. Apraxic children appeared unable to discriminate consonants presented at 0 dB white noise when rich visual cues were not provided. They were helped by rich visual cues, but this facilitation was less robust than observed with typically developing language- and age-matched control participants. Finally, its connection with the elaboration of articulatory information should not be taken for granted.

While condition 1 mainly produced data pertaining to Apraxia of Speech, condition 2, taken together with the correlations emerging between the perception and the productions scores, also provided information on typical developing children.

Precisely:

(i) Adults obtained high discrimination scores for both trial-types and no effect of the trial-type on accuracy and reaction times was observed, i.e. they were not affected by the lower or higher motor complexity of the consonants. This means that they were able to exploit the residual acoustic input[64]. In addition, a highly reliable index of sensorimotor proficiency (i.e. the diadochokinetic rate, cf. Chap. 4, Sections 5.1, 5.2 and 6.1) selectively correlated with their capacity to discriminate complex motor schemas (as measured in both the normal and the auditory masking condition of production).

(ii) Both groups of typical children displayed a significant effect of trial-type, obtaining better discrimination scores with easy *vs* complex motor schemas. Interestingly, this effect was not significant in the second set of trials (i.e. when identical trials were repeated in the reverse combination, cf. Table 11, page 71), suggesting that the previous occurrence determined this recovery in discrimination precision.

Besides, the same correlation observed in adults emerged, but limitedly to 4-to-5-years-olds, as tested in the normal modality of production. On the other side, when sensorimotor proficiency was measured in the auditory masking modality (i.e. with a

---

[64] It is worthwhile to recall that, as to former literature (Files et al., 2015), adults could also have benefitted from some of the poor visual cues characterizing the contrasts in condition 2 (i.e. [dza] *vs* [dʒa], more similar than [da] *vs* [na] to the contrasts used in the cited study (cf. Chap. 4, Section 3.2). Therefore, their high accuracy should also be seen as the product of a more redundant perceptual capacity. At any rate, this point does not affect the results of the task.

more accurate measurement), it only correlated with the capacity to discriminate easy motor schemas in both groups.

(iii) The apraxic children obtained below-chance-level scores for both easy and complex motor schemas (40.3% mean correct answers across sets, cf. page 87). In addition, no measure of sensorimotor proficiency correlated with perception in this group. By contrast, and crucially, perception for easy motor schemas did correlate with the duration of the speech therapy and its type (P.R.O.M.P.T.).

The results obtained in condition 2 agree with the former literature targeting the sensorimotor contribution to speech production in development (e.g. McAllister Byun & Tiede, 2017; Ross et al., 2011). In accordance with this field of study, they point to gradual ontogenetic emergence of the sensorimotor contribution to speech perception, which is attained thanks to progressive connection between the production-related sensorimotor knowledge and perception (cf. Chapter 2, Section 4.3 and 3.3).

With reference to typically developing children, this dissertation adds evidence of a late sensorimotor contribution to the elaboration of some complex sounds, i.e. affricates.

For what pertains to apraxia of speech, the data argue for a dissociation between production and perception abilities, coherently with the clinical characteristics of this pathology. In this respect, studies targeting the neurofunctional activation of the motor areas during speech processing are strongly recommended in this population.

Differently from former literature (Rvachew & Jamieson, 1989; Groenen et al., 1996; Nijland, 2009; McAllister Byun, 2012 – cf. Chap. 3, Section 2.3), this study did not detect a correlation between the participants’ capacity to produce and perceive affricates (it is worthwhile to recall that only 3 subjects in the group were able to produce those sounds, cf. Table 7, page 65). However, this seems to be due to differences in the experimental design. In fact, the cited investigations were based on the participants’ individual error patterns (i.e. the comparison production/perception focused the consonants with which the participants made more errors), while the values used in this study were less fine-grained, consisting in the speech therapists’ measurements of the consonants present or absent from the expressive phonological inventory of the child.

Moreover, differently from previous studies, this investigation targeted in-noise speech perception, a condition which is known to elicit enhanced reliance upon the sensorimotor perceptual component. Hence, the results obtained highlight that, once the listening conditions require a crucial contribution of sensorimotor processing, its severe deficiency in the apraxic participants becomes manifest.

Finally, this dissertation highlights an association between better perceptual skills and the use of the P.R.O.M.P.T. therapy, a rehabilitation method (re-)educating the integration between articulatory movements and auditory consequences. This point further highlights the sensorimotor origin of the perceptual deficits displayed by the apraxic participants.

2. Additional factors potentially contributing to the results

After having analyzed the aspects that explicitly affected the results, it is worthwhile to examine which elements could have been at play but were not factorized in the experimental procedure. The potential influence of three families of factors should be evaluated, i.e. acoustic, linguistic and attention-related ones.

Starting from the former, a specific feature of the experimental stimuli has to be mentioned: the frequency of usage of the stimuli used. While plosives and nasals have comparable type frequency in Italian[65] ($p = .54$, $t = 0.7$ – unpaired t-test), affricates are rarer than both plosives and nasals ($p = 0.01$, $t = 7.7$, respectively). This points to a common issue in studies addressing the relationship between production and perception of articulatory difficult sounds, i.e. the fact that, in general, complex motor schemas are cross linguistically ‘marked’ and rarer. Thus, it seems highly probable that lesser experience in hearing and producing these sounds could have contributed to these and other results.

It is worthwhile to underline, in this respect, that former theoretical frameworks addressing this topic (such as the ‘Cue weighting hypothesis’, cf. Chap. 1, page 16) have pointed out the joint role of experience in hearing and producing ‘difficult’ phonemes. Interpreting her own studies on the developmental differences between the intra-categorical processing of plosives as compared with that of fricatives, Nittrouer (2002) framed the problem as follows:

> « This hypothesis suggests simply that the informational aspects of the signal that get weighted greatly (or, attended to greatly) change as children gain experience listening to and speaking their native language. Initially, children seem to attend largely to those acoustic properties that specify a changing vocal tract, which means the dynamic spectral changes known as formant transitions. Then, as the child becomes more skilled, attention shifts to acoustic properties that do not involve spectral change, properties such as silent gaps (specifying periods of vocal-tract closure),

---

[65] PhonItalia: a phonological lexicon for Italian *Behavioral Research Methods*, vol. 46(3), 872-86. *PhonItalia* is an open access corpus providing the frequency of use for Italian phonemes, syllables, syllable onsets and codas (available at: http://www.phonitalia.org/).

differences in voicing duration (specifying the voicing of syllable-final stops), or periods of stable spectral information (specifying place of consonantal constrictions) »[66].

As one can notice, this account is based on both listening and production experience with the sounds of one's language. The relevance of the latter component emerges with special evidence when considering a previous study from the same author (cf. Nittrouer, Studderd-Kennedy & McGowan, 1989), showing a tendency to coarticulate the fricative with the vocalic context in production in a way that parallels that observed in perception (later – cf. Nittrouer, 2002).

In brief, sensorimotor complexity and acoustic low frequency are connected in development. Hence, future child studies specifically aiming at disentangling the two aspects are needed. A valuable possibility in this direction would be represented by neurofunctional techniques allowing to assess the *loci* of activation underlining the difficulty in the elaboration of complex/rare phonemes.

An indirect way to assess the incidence of the phonemes' frequency of occurrence resides in the evaluation of the level of lexical development of the participants (as measured with respect to comprehension); the logic being that, the wider the vocabulary of a child, the higher the probability that she/he has made significant experience with a variegated range of phonemes. The level of linguistic development, indeed, represents the second major factor that this study did not fully evaluate.

As a matter of fact, this level was only taken into account for the apraxic children (and it did not yield significant effects on discrimination accuracy, cf. Chap. 4, Section 4.4 and 4.5). By contrast, it could not be calculated for the typical participants, as parents did only allow a single 30-minutes experimental session. The only analyzable index, in this respect, was the sheer difference in phonological awareness existing between preschool and school-aged children, as determined by the acquisition of reading and writing skills. As no relevant difference was observed in the perceptual performance of the two groups, this do not seem to have directly influenced the results.

Finally, the generally high cognitive costs characterizing the experimental procedure (where the signal was degraded and an explicit metalinguistic judgement was required) have to be considered. In this respect, carefulness is required in comparing these results with the outcomes of investigations entailing a less intense cognitive and attentional effort. However, in relation to the type of signal degradation used (white noise) and to the experimental

---

[66] Nittrouer (2002), page 713.

setting provided (granting the vision of the mouth, although this was not uniformly informative), this experimental procedure was ecological enough to be (at least partially) comparable with some real-life situations, as listening in noisy classrooms. Speech in noise discrimination *can be* cognitively demanding in children's daily life, especially when rich visual cues are not provided and the linguistic input contains infrequent/complex sounds.

3. On the status of poor visual cues and their potential heuristic relevance

As previously mentioned (Chap. 4, Section 3.2, page 69), the perceptibility of poor visual cues constitutes a quite unexplored issue in children research. This is due to two basic facts: (i) their detectability was until recently discarded even for adult speakers; (ii) children are not very sensitive to visual speech cues (cf. Chap. 2, Section 4).

However, recent research has refuted the former assumption (Files et al., 2015; Bernstein, 2012) and demonstrated more nuanced results in relation to the latter (e.g. Jerger et al., 2014, 2017 and 2018; Lalonde & Holt, 2015). Specifically, the ability to rely upon rich visual cues in identification and discrimination tasks, starting from the preschool years (i.e. 4-to-5 years of age), has been repeatedly reported (e.g. Jerger, 2018 and Lalonde & Holt, 2015). Moreover, sensitivity to poor visual cues has emerged, in older children (8-to-9-years-olds), in correspondence with an experimental procedure which does not demand a heavy cognitive effort, i.e. audiovisual priming (Jerger et al., 2014).

In this respect, it is worthwhile to remark that, when compared with those used by Files and colleagues with adults (e.g. [da] vs [ʒa]), the poor visual cues tested in the above mentioned studies are definitely poor, namely: the silent articulation of [g] as a visual prime for auditory stimuli with excised auditory onsets in Jerger et al. (2014); [ba] *vs* [ma] in the speech-in-noise discrimination and identification task carried out by Lalonde & Holt. Interestingly, the [da] *vs* [na] pairing used in this dissertation was of the same level of difficulty (while the perceptual distance between [dza] and [dʒa], intuitively more informative, should be measured by means of rigorous methods, e.g. Jiang et al., 2007).

These kinds of contrasts could be too challenging for children (and, indeed, for adults too). But what about intermediate – i.e. subtle, rather than poor – distinctions such as those tested by Files and colleagues?

A systematic psychometric assessment of the sensitivity to perceptibly near visual cues during childhood, if based on a previous measure of the perceptual distance between the stimuli and conceived with an aim to compare the childrens' perceptual acuity in implicit (e.g. the visual priming procedures used by Jerger and colleagues) *vs* explicit tasks (such as discrimination and identification), could substantially contribute to solve the problem of visual cues detection during ontogenesis.

This kind of investigation could have an additional heuristic value. As previously observed, the use of rich visual cues does not grant that children process them on a language-specific base (Section 1, pages 101-102). Conversely, the detection of subtle visual cues could hardly be carried out by simply relying on visual information such as the physical dissimilarities in the mouth’s shape or movement. Probably, the most relevant source of information in this case would be articulatory in nature, i.e. would entail the recruitment of speech sensorimotor knowledge as automatically activated by the vision of the movement.

As such, if proved to be feasible, procedures based on subtle visible speech cues could allow to compare children and adults on tasks assuring that the elicited modality of processing is linguistic in nature.

APPENDIX

EXPERIENCE-RELATED DEVELOPMENTAL DIFFERENCES IN THE PERCEPTION OF VOWELS AND CONSONANTS: THE CASE OF FAMILIAR WORDS

I. General Rationale

It has long been known that consonants and vowels play different roles in lexical processing and learning, during development and in adulthood. In particular, a vast amount of investigations has displayed how contoids seem to be more informative than vocoids in these respects (cf. Chap. 1, Section 3.2).

This notwithstanding, recent research targeting such issue from a developmental point of view has shown that a universal and innate 'consonant bias' should not be taken for granted. Rather, robust connections between processing biases and language-specific characteristics of the input have been revealed, encouraging to explore the multiple ways in which the interactions between innate predispositions and environmental input could shape the phenomenon (cf. Mani & Plunkett, 2007; 2010; Floccia et al., 2014; Højen & Nazzi, 2016 – Chap. 1, pages 13-14).

Under this perspective, particularly interesting is the recently reported finding that French-learning infants (who are known to display a robust preference for consonants *vs* vowels in lexical processing) seemingly switch from a former bias for vowels to a later bias for consonants around the eighth month of life (Nishibayashi & Nazzi, 2016).

Evaluating an experience-based explanation of these data, the authors have underlined a factor of primary importance, i.e. the intrinsic difference characterizing the auditory exposure to the two classes of segments in early infancy (as given by the fact that vowels are significantly better perceivable than consonants during the prenatal period).

However, starting from the second half of the first year of life, auditory elaboration is not the only source of experience with the patterns of one's native language. In particular, current research highlights that, when infants begin to produce linguistic sounds, the sensorimotor knowledge accumulated through this activity begins to support perception, making it more efficient (cf. Chap. 2, Section 3.3). Relating these investigations to the results in Nishibayashi & Nazzi (2016), it could be worthwhile to recall that the production of vowels precedes the production of consonants in ontogenesis and that the onset of the latter approximates the moment when the switch from the vowel-bias to the consonant-bias has been observed (the 8th month of life). Hence, the possibility that experience with production could play a role in the development of the consonant-bias alongside with auditory factors is not excluded.

In more general terms, as argued in Chap. 2 (Section 3.2), it seems relevant to investigate further the role of linguistic experience in shaping the processing of consonants as compared to vowels by enlarging the investigation to sensorymotor experience.

II. Production-related effects in familiar words processing: a pilot study

As an attempt to contribute to the line of research described in the former section, a study is currently ongoing with the aim of evaluating the possibility that production-related effects could emerge in the early stages of lexical processing. The investigation was carried out at the Laboratoire de Psychologie de la Perception (Université Paris Descartes), in the context of a three-months LabEx Mobility Grant.

Two groups of 11- and 14-month-old French-learning monolingual infants with no familial history of speech and language disorders were recruited.

At these stages, the amount of linguistic sounds that infants (normally) produce is sufficient to allow the detection of individual variability, i.e. to differentiate subjects who produce a wider or lower variety of speech sounds (‘high-producers’ *vs* ‘high-producers’). Given that the sensorimotor support to speech perception arises from experience with production, the comparison between low- and high-producers allows to assess if the perceptual performance differs in the two cases (e.g. De Paolis et al., 2010; Majorano et al., 2013).

A younger and an older group of infants were enrolled in order to estimate, in the event that production-related effects emerge in development, whether the latter evolve after a longer and more variegated experience with production.

II.I Stimuli and procedure

The experimental procedure entailed an evaluation of production skills and a perception task.

To begin with the former, a production score was calculated for each participant, corresponding to the quantity of consonants that the infant produced as assessed by means of a parental questionnaire (cf. Table 1 and 2, pages 115-116). The interview had to be preferred to more objective and informative methodologies, such as the recording and analysis of production samples based on the LENA system, due to the temporal limitations characterizing this preliminary investigation.

On the basis of such scores, the two groups were internally subdivided into high-producers and low-producers.

Given the reduced number of participants so far composing the groups, a rough subdivision was carried out, based on the positioning of each infant with respect to the group's mean production scores. Infants whose scores felt above or were identical to the group's average were assigned to the high-producers group; infants whose scores felt below the group's average were assigned to the low-producers group. More rigorous methods are applicable in this context (e.g. Majorano et al., 2013) and a new computation of this aspect is envisaged once the groups will be completed (i.e. approximately 30 infants per group).

The perception task consisted in a head-turn preference procedure replicating the methodology in Poltrock & Nazzi (2015).

The stimuli were constituted by familiar words, chosen from the French version of the MacArthur Bates Test (Kern, 2003) as to be understood by at least 30% of French-learning infants of 11 and 14 months (average value; range: 15% - 56%). This evaluation was based on a previous survey (S. Kern, personal communication); its reliability was verified case by case, by asking parents to evaluate if the infant heard the words in the lists on a regular basis.

Controlled for frequency and acoustic characteristics, the stimuli varied along a production-related dimension: one half of the items (10) was exclusively composed by early-acquired, easy-to-articulate consonants (i.e. plosives and nasals); the other half (10) was exclusively composed by late-acquired, difficult-to-articulate consonants (fricatives). The vowel context and the syllabic length were varied within lists and balanced across lists (cf. Table 3 to 6, pages 117-119).

The words were spoken by a French female native speaker in an infant-directed style and recorded in a sound-attenuating booth. Three tokens were selected for each item and twelve pseudo-randomly ordered lists were constructed, six containing the words composed by the 'easy consonants' and six containing the words composed by the 'difficult consonants'. Each list was subdivided into two ‘blocks’, containing two different tokens of the stimuli. All lists were equal in duration (23.3 seconds for the 11-months-olds, 22.065 seconds for the 14-months-olds), with an interstimulus interval varying between 550 and 650 ms.

For what pertains to the procedure, infants were held on the caregiver’s lap in a sound-attenuated booth which had a red light and a loudspeaker on each side (left and right) and a green light at the center (i.e. in front of the infant). The infants’ behavior was monitored through a video camera connected to a PC computer terminal, a TV screen and a response box located outside the booth.

At the beginning of each trial, the green light at the center of the cabin blinked until the infant looked in that direction. Once the participant was oriented towards the center, the green light was extinguished and the red light above the loudspeaker located at on one of the sides of the cabin began to flash. When the infant made a turn of at least 30° in the direction of the loudspeaker, the experimenter initiated the presentation of the sound file for the given trial and the red light continued to flash for the entire duration of the trial; if the infant turned away by 30° for less than 2 seconds and then turned back again, the trial continued but the time spent looking away was not recorded. Conversely, if the infant’s orientation time for a trial was shorter than 1.5 seconds, the trial was repeated.

The experimenter conducted an on-line coding of the infants’ looking times by pressing the buttons of the response box according to the direction of the infants’ head.

Each session began with a training phase, constituted by two non-experimental musical trials allowing the participant to learn the association between her/his own head turn and the administration of the stimuli. The test phase consisted of 12 trials (i.e. 12 lists), 6 for each condition (‘easy consonants’ *vs* ‘difficult consonants’). No familiarization phase with the test trials was entailed.

II.II Preliminary results and discussion

As reported in Table 1 and 2 (pages 115-116), a total of 15 11-months-olds and 10 14-months-olds completed the experiment so far. An additional 6 infants were tested but were excluded from the analyses due to fussiness (3) or because they did not meet the inclusion criteria (2 bilinguals, 1 infant coming from a family with a high familiarity rate of speech and language disorders). On the basis of the data collected on this limited number of participants, some trends emerged.

Starting with the 11-months-olds, the Low-producers did not display significant differences in their mean orientation times to the lists containing ‘easy consonants’ *vs* ‘difficult consonants’ (paired t-test: *t(x)* = 0.09, *p* = .9). Conversely, High-producers oriented significantly longer to the lists containing ‘easy consonants’ *vs* the lists containing ‘difficult consonants’ (paired t-test: *t(x)* = 2.96, *p* = .02).

Figure 1 – Orientation times; Low-producers (11-months-olds)

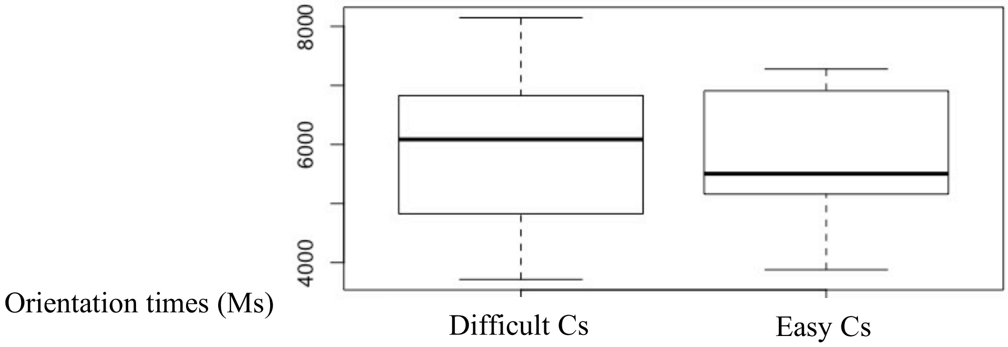
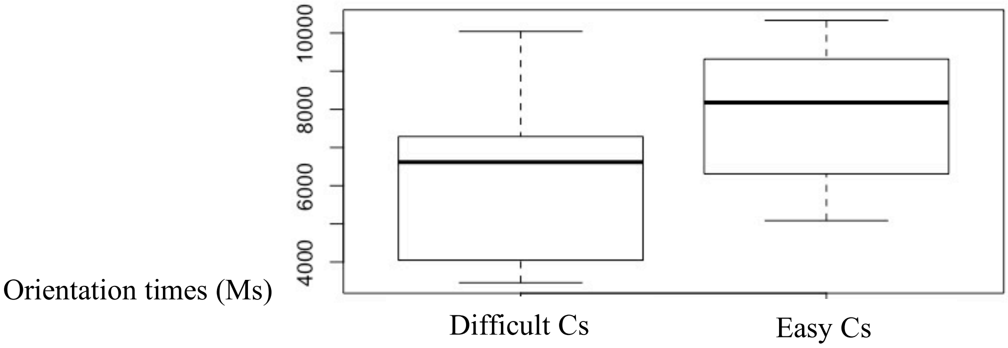


Figure 2 – Orientation times; High-producers (11-months-olds)



At this very preliminary stage, the tendencies displayed by the 14-month-old group (as composed by 10 infants, i.e. 5 High-producers and 5 Low-producers) displayed a reverse pattern, as the Low-producers oriented more to the lists containing ‘easy consonants’ *vs* the lists containing ‘difficult consonants’, while no specific trend was observed in High-producers (cf. Figures 3 and 4).

However, this pattern stems from an extremely limited number of participants and actual tendencies will only be detectable when the size of the group will be adequate.

Figure 3 – Orientation times; Low-producers (14-months-olds)

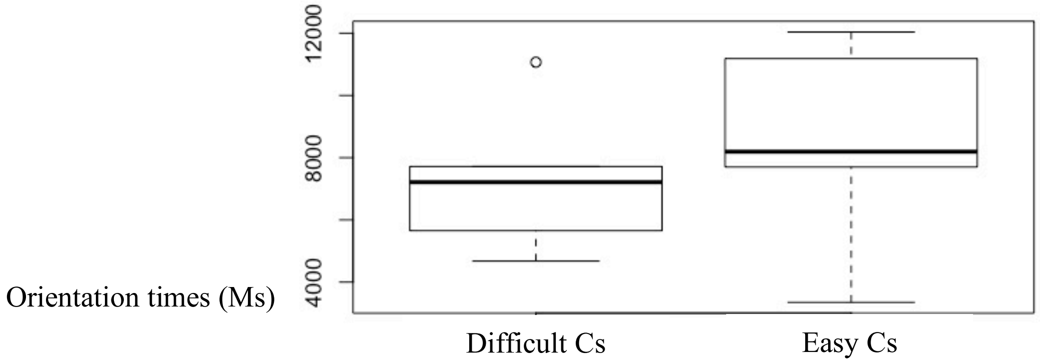
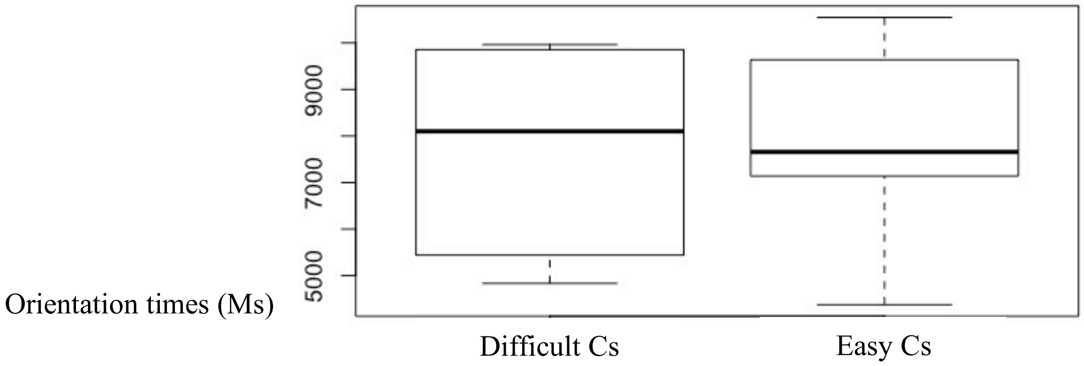


Figure 4 – Orientation times; High-producers (14-months-olds)



To sum-up, on the basis of a very preliminary analysis of the data, a significant orientation preference for the trials containing ‘easy consonants’ emerged in High-producers at 11 months. Such preliminary trends point to possible production-related differences, which will have to be evaluated *via* statistical modelling when a sufficient number of participants will have been collected (and the group of 14-month-old will have been added).

In these connection, some operative hypothesis can be put forward. Specifically, the preference for words made up by easy consonants displayed by the High-producers at 11 months and the (hypothetical) analogous preference showed by the Low-producers at 14 months (if confirmed) may be connected with the fact that such infants were accomplishing their capacity to produce this set of consonants in the period when the testing took place. This is suggested by the participants’ individual production patterns since, as Table 1 (page 115) shows, in both cases, none of the participants produced the whole set of French plosives and nasals (with the exception of subject 15).

Such possibility would be consistent with the results obtained by De Paolis, Vihman & Keren-Portnoy (2010), where infants' looking times «were influenced *both* by their own consonant production experience [i.e. the variety of consonants that each infant was able to produce] and by whether or not they produced the particular consonant featured in the non-words embedded in the passages»[67]. In other words, our participants could have listened more to plosives and nasals than to fricatives because they were still learning consonants of this type.

Overall, this preliminary analysis seems to encourage a continuation of the study. If our future results confirm the present tendencies, further analyses targeting the individual production patterns would be worthwhile.

---

[67] De Paolis, Vihman & Keren-Portnoy (2010), page 598.

III. Tables

Table 1 – 11-months-old participants

Id	**Age (mm; dd)**	**Gender**	**Stably produced Vs**	**Sporadically produced Vs**	**Stably produced Cs**	**Sporadically produced Cs**	**Production score**	**Reduplicated babbling (yes/no)**	**Variegated babbling (yes/no)**
1	11	M	a ɛ		p t b m	g d	Low-producer	y	n
2	11	M	a ɛ ɔ		b k g l		Low-producer	y	n
3	11;21	F	a ɛ		p b d m	l	Low-producer	y	n
4	11	M	a e o		p b k m l		Low-producer	y	n
5	11	M	a e		b t d m l	g n	Low-producer	y	n
6	11;20	M	a ɛ ɔ e		p b t m l	d	Low-producer	y	y
7	11;12	M	a	ɛ ɔ	p d g m n		Low-producer	y	n
8	11	F	a e		p b t k g m	l	High-producer	y	y
9	11	F	a ɛ		p b t k m n		High-producer	y	n
10	10;28	F	a ɛ e		p b t d g m l		High-producer	y	n
11	11	M	a ɛ e i		p b t d k m l		High-producer	y	y
12	11;16	F	a ɛ ɔ e o	i	p b t d g m n		High-producer	y	y
13	11	F	a ɛ u		p b t d k g m l	n	High-producer	y	n
14	11	F	a ɛ		p b t d k m n l		High-producer	y	n
15	11;29	M	a ɛ u		p b t d k g m n l		High-producer	y	y

Table 2 – 14-months-old participants

Id	**Age (mm; dd)**	**Gender**	**Stably produced Vs**	**Sporadically produced Vs**	**Stably produced Cs**	**Sporadically produced Cs**	**Production score**	**Reduplicated babbling (yes/no)**	**Variegated babbling (yes/no)**
1	14;9	M	a		t d m		Low-producer	y	n
2	14	F	a ɛ e o		p b k m l	g	Low-producer	y	y
3	14;30	M	a ɛ o u		p b m n l	k g	Low-producer	y	y
4	14	F	a ɛ o u		p b t d g m	l	Low-producer	y	y
5	14;18	M	a ɛ e i		p b d k g m	n l	Low-producer	y	y
6	14	F	a ɛ ɔ u		b t d k g m n l		High-producer	y	n
7	14	M	a ɛ ɔ u	e o	p b t d m n l ʃ		High-producer	N.A.	N.A.
8	14	F	a ɛ e u		p b t d k m n l R		High-producer	y	y
9	14;18	M	a ɔ e o		p b t d k g m n l		High-producer	y	y
10	14;7	M	a ɛ ɔ		p b t d k g m n z		High-producer	y	y

Table 3 – Test stimuli used in the perception experiment (11-months-olds)

Familiar words – easy consonants		**Syllabic structure**
pain	/pɛ̃/	CV
nez	/ne/	CV
doigt	/dwa/	C[dip]
pied	/pje/	C[dip]
pomme	/pɔm/	CVC
tête	/tɛt/	CVC
cube	/kyb/	CVC
gâteau	/ga'to/	CVCV
body	/bo'di/	CVCV
banane	/ba'nan/	CVCVC
Familiar words – difficult consonants		**Syllabic structure**
chat	/ʃa/	CV
lit	/li/	CV
lait	/le/	CV
verre	/vɛR/	CVC
soeur	/soeR/	CVC
vache	/vaʃ/	CVC
chaise	/ʃɛz/	CVC
avion	/avjɔ̃/	VC[dip]
chausson	/ʃosɔ̃/	CVCV
chaussure	/ʃosyR/	CVCVC

Table 4 – Test stimuli used in the perception experiment (14-months-olds)

Familiar words – easy consonants		**Syllabic structure**
pain	/pɛ̃/	CV
nez	/ne/	CV
pied	/pje/	C[dip]
pot	/po/	CV
pomme	/pɔm/	CVC
tête	/tɛt/	CVC
cube	/kyb/	CVC
gâteau	/ga'to/	CVCV
body	/bo'di/	CVCV
banane	/ba'nan/	CVCVC

Familiar words – difficult consonants		**Syllabic structure**
chat	/ʃa/	CV
lit	/li/	CV
lait	/le/	CV
lion	/ljɔ̃/	C[dip]
four	/fuR/	CVC
vache	/vaʃ/	CVC
chaise	/ʃɛz/	CVC
vélo	/velo/	CVCV
chausson	/ʃosɔ̃/	CVCV
cheval	/ʃəval/	CVCVC

Table 5 – Acoustic characteristics of the stimuli (11-months-olds)

	Condition		**t-value and significance level (two-sided)**
	Easy consonants	Difficult consonants	
Duration (Ms)	570 (70)	600 (90)	$t = 0.73$, $p = .50$
Amplitude RMS (dB)	74.25 (5.1)	74.96 (4.1)	$t = -0.32$, $p = .75$
F0 Mean (Hz)	268 (13.4)	273 (14.1)	$t = -0.79$, $p = .43$
F0 Max (Hz)	332 (17)	273 (14.1)	$t = 0.95$, $p = .35$
F0 Min (Hz)	203.6 (26.2)	221 (24)	$t = -1.2$, $p = .23$

Table 6 – Acoustic characteristics of the stimuli (14-months-olds)

	Condition		**t-value and significance level (two-sided)**
	Easy consonants	Difficult consonants	
Duration (Ms)	540 (10)	570 (0.06)	*t* = -0.83, *p* = .41
Amplitude RMS (dB)	75 (6)	77 (2)	*t* = -0.99, *p* = .33
F0 Mean (Hz)	269.75 (14)	269.3 (15)	*t* = -0.02, *p* = .97
F0 Max (Hz)	330.1 (19)	318.2 (20)	*t* = 1.21, *p* = .24
F0 Min (Hz)	210.2 (30)	212.6 (32)	*t* = -0.15, *p* = .87

BIBLIOGRAPHICAL REFERENCES

- A -

Abboub N., Boll-Avetisyan N., Bhatara A., Hoehle B. & Nazzi T. (2016). An exploration of rhythmic grouping of speech sequences by French- and German-learning infants. *Frontiers in Human Neuroscience*, vol. 10(292), 10, 292.

Aldridge M.A., Braga E.S., Walton G.E. & Bower T.G.R. (1999). The intermodal representation of speech in newborns. *Developmental Science*, vol. 2, 42-46.

Altvater-Mackensen N. & Grossmann T. (2016). The role of left inferior frontal cortex during audiovisual speech perception in infants. *NeuroImage*, vol. 133, 14-20.

Altvater-Mackensen N., Mani N. & Grossmann T. (2016). Audiovisual speech perception in infancy: The influence of vowel identity and infants' productive abilities on sensitivity to (mis)matches between auditory and visual speech cues. *Developmental Psychology*, vol. 52(2), 191-204.

Altvater-Mackensen N. & Grossmann T. (2015). Learning to Match Auditory and Visual Speech Cues: Social Influences on Acquisition of Phonological Categories. *Child Development*, vol. 86(2), 362-378.

Altvater-Mackensen N., van der Feest S.H. & Fikkert P. (2014). Asymmetries in Early Word Recognition: The Case of Stops and Fricatives. *Language, Learning and Development*, vol. 10 (2), 149-178.

Altvater-Mackensen N. & Fikkert P. (2010). The acquisition of the stop-fricative contrast in perception and production. *Lingua*, vol. 120 (8), 1898-1909.

American Speech, Language and Hearing Association, *2007 Technical Report on Childhood Apraxia of Speech*. Available at: http://www.asha.org/public/speech/disorders/childhoodapraxia.htm

Arai T., Ohki E. & Iitaka K. (2008). Perception of long vowels in Japanese by Children. *Acoustic Science & Technology*, vol. 29(1), 106-109.

Aslin R.N., Saffran J.R. & Newport, E.L. (1999). Computation of Conditional Probability Statistics by 8-Month-Old Infants. *Psychological Science*, vol. 9(4), 321-324.

Ayneto A. & Sebástian-Gallés N. (2016). The influence of bilingualism on the preference for the mouth region of dynamic faces. *Developmental Science*, vol. 20(1), e12446.

Aziz A.A., Shohdi S., Osman D.M., Habib E.I. (2010). Childhood apraxia of speech and multiple phonological disorders in Cairo-Egyptian Arabic speaking children: Language, speech, and oro-motor differences. *International Journal of Pediatric Otorhinolaryngology*, vol. 74, 578-585.

- B -

Baart M., Bortfeld H. & Vroomen J. (2015). Phonetic matching of auditory and visual speech develops during childhood: Evidence from sine-wave speech. *Journal of Experimental Child Psychology*, vol. 129, 157-164.

Baart M., Vroomen J., Shaw K. & Bortfeld (2014). Degrading phonetic information affects matching of audiovisual speech in adults, but not in infants. *Cognition*, vol. 130(1), 31-43.

Bahrick L.E., Lickliter R. & Flomm R. (2004). Intersensory Redundancy Guides the Development of Selective Attention, Perception, and Cognition in Infancy. *Current Directions in Psychological Science*, vol. 13(3), 99-102.

Bailey T.M. & Plunkett K. (2002). Phonological specificity in early words. *Cognitive Development*, vol. 17, 1265-1282.

Baker Palmer S., Fais L., Michnick Golinkoff R. & Werker J. (2012). Perceptual Narrowing of Linguistic Sign Occurs in the 1st Year of Life. *Child Development*, vol. 83(2), 543-553.

Barbier G., Boë L.J., Captier G. & Laboissière R. (2015). Human vocal tract growth: A longitudinal study of the development of various anatomical structures. 16th Annual Conference of the International Speech Communication Association (Interspeech 2015), Sep 2015, Dresden, Germany.

Barnaud M-L., Bessière P., Diard J. & Schwartz J-L. (2017). Reanalyzing neurocognitive data on the role of the motor system in speech perception within COSMO, a Bayesian perceptuo-motor model of speech communication. *Brain and Language*, doi: 10.1016/j.bandl.2017.12.003.

Basirat A., Dehaene S. & Dehaene-Lambertz G. (2014). A hierarchy of cortical responses to sequence violations in three-month-old infants. *Cognition*, vol. 132(2), 137-150.

Benasich A.A. & Tallal P. (2002). Infant discrimination of rapid auditory cues predicts later language impairment. Behavioural Brain Research, vol. 136(1), 31-49.

Benson R.R., Whalen D.H., Richardson M., Swainson B., Clark V.P., Lai S. & Liberman A.M. (2001). Parametrically Dissociating Speech and Nonspeech Perception in the Brain Using fMRI. *Brain & Language*, vol. 78(3), 364-396.

Berken J., Gracco V. & Klein D. (2015). The timing of language learning shapes brain structure associated with articulation. *Brain, Structure and Function*. 221(7):3591-600.

Bernstein N.A. (1967). *The coordination and regulation of movements*. Oxford: Pergamon.

Bernstein L.E. (2012). Visual speech perception. In: Bailly G., Perrier P. & Vatikiotis-Bateson E. *Audiovisual Speech Processing*. Cambridge: Cambridge University Press, 21-39.

Bertoncini J., Nazzi T., Cabrera L. & Lorenzi C. (2011). Six-month-old infants discriminate voicing on the basis of temporal envelope cues. *Journal of the Acoustical Society of America*, vol. 129(5), 2761–2764.

Bertoncini J., Serniclaes W. & Lorenzi C. (2009). Discrimination of speech sounds based upon temporal envelope versus fine structure cues in 5-to-7-year-old children. *Journal of Speech, Language and Hearing Research*, vol. 52(3), 682-695.

Bertoncini J., Morais J., Bijeljac-Babic R., McAdams S., Peretz I. & Mehler J. (1989). Dichotic perception and laterality in neonates. *Brain & Language*, 37, 591–605.

Bertoncini B., Bijeljac-Babic R., Jusczyk P.W., Kennedy L.J. & Mehler J. (1988). An investigation of young infants' perceptual representations of speech sounds. *Journal of experimental psychology: general*, vol. 117(1), 21-33.

Bertoncini J., Bijeljac-Babic R., Blumstein S.E. & Mehler J. (1987). Discrimination in neonates of very short CVs. *The journal of the acoustic society of America*, vol. 82(1), 31-37.

Best C.T. & McRoberts G.W. (2003). Infant perception of non-native consonant contrasts that adults assimilate in different ways. *Language and Speech*, vol. 46(2), 183-216.

Betz S.K., Stoel-Gammon K. (2005). Measuring articulatory error consistency in children with developmental apraxia of speech. *Clinical Linguistics and Phonetics*, vol.19(1), 53-66.

Bever T.G. & Poeppel D. (2010). Analysis by synthesis: A (re-) emerging program of research for language and vision. *Biolinguistics*, vol. 4(2-3), 174-200.

Bahrick L. E., Lickliter R. & Flom R. (2004). Intersensory redundancy guides the development of selective attention, perception, and cognition in infancy. *Current Directions in Psychological Science*, vol. 13(3), 99-102.

Bion R.A.H., Benavides-Varela S. & Nespor M. (2011). Acoustic markers of prominence influence infants' and adults' segmentation of speech sequences. *Language and Speech*, vol. 54(1), 123-140.

Boliek C.A., Hixon T.J., Watson P.J. & Morgan W.J. (1997). Vocalization and breathing during the second and third years of life. *Journal of Voice*, vol. 11(4), 373-390.

Bonatti L.L., Peña M., Nespor M. & Mehler J. (2005a). Linguistic constraints on statistical computations: the role of consonants and vowels in continuous speech processing. *Psychological science*, vol. 16(6), 451-459.

Bosch L. & Sebastian-Gallés N. (1997). Native-language recognition abilities in 4-month-old infants from monolingual and bilingual environments. Cognition, vol. 65(1), 33-69.

Bouchon C., Floccia C., Fux T., Adda-Decker M., & Nazzi T. (2015). Call me Alix, not Elix: Vowels are more important than consonants in own name recognition at 5 months. *Developmental Science*, vol. 18(4), 587-598.

Bridgeman E. & Snowling M. (1988). The perception of phoneme sequence: A comparison of dyspraxic and normal children. *British Journal of Disorders of Communication*, vol. 23(3), 245-252.

Bristow D., Dehaene-Lambertz G., Mattout J., Soares C., Gliga T., Baillet S. & Manging J-F. (2009). Hearing faces: how the infant brain matches the face it sees with the speech it hears. *Journal of Cognitive Neuroscience*, vol. 21(5), 905-921.

Bruderer A.G., Danielson D.K., Kandhadai P. & Werker J. (2015). Sensorimotor influences on speech perception in infancy. *PNAS*, vol. 112(44), 13531-13536.

Burgmeier R., Desai R.U., Farner K.C., Tiano B., Lacey R., Volpe N.J., Mets M.B. (2016). The Effect of Amblyopia on Visual-Auditory Speech Perception Why Mothers May Say “Look At Me When I’m Talking To You”. *JAMA Ophthalmology*, vol. 133(1), 11-16.

Burnham D.K. & Sekiyama (2012). Investigating auditory-visual speech perception. In: Bailly G., Perrier P. & Bateson-Vatikiotis E. *Audiovisual Speech Processing*. Cambridge: Cambridge University Press, 62-75.

Burnham D.K. & Dodd B. (2004). Auditory-Visual Speech Integration by Prelinguistic Infants: Perception of an Emergent Consonant in the McGurk Effect. *Developmental Psychobiology*, vol. 45(4), 204-20.

Burnham D.K., Earnshaw L. J., & Clark J. E. (1991). Development of categorical identification of native and non-native bilabial stops: Infants, children and adults. *Journal of Child Language*, vol. 18, 231-260.

Burns T.C., Yoshida K.A., Hill K. & Werker J.F. (2007). The development of phonetic representation in bilingual and monolingual infants. *Applied Psycholinguistics*, vol. 28(3), 455-74.

- C -

Cabrera L., Lorenzi C. & Bertoncini J. (2015). Infants discriminate voicing and place of articulation with reduced spectral and temporal modulation cues. *Journal of Speech, Language and Hearing Research*, vol. 58(3), 1033–1042.

Cabrera L., Bertoncini J. & Lorenzi C. (2013). Perception of speech modulation cues by 6-month-old infants. *Journal of Speech, Language, and Hearing Research*, vol. 56(6), 1733–1744.

Calvert G., Brammer M.J., Bullmore E.T., Iversen S.D. & David A.S. (1999). Response amplification in sensory-specific cortices during crossmodal binding. *NeuroReport*, vol. 10(12), 2619–2623.

Calvert G.A., Bullmore E.T., Brammer, M.J., Campbell R., Williams S.C.R., McGuire P.K., Woodruff P.W.R, Iversen S.D. & David A.S. (1997). Activation of auditory cortex during silent lipreading. *Science*, vol. 276(5312), 593–596.

Cantiani C., Riva V., Piazza C., Bettoni R., Molteni M., Choudhury N., Marino C. & Benasich A.A. (2016). Auditory discrimination predicts linguistic outcome in Italian infants with and without familial risk for language learning impairment. *Developmental Cognitive Neuroscience*, vol. 20, 23-34.

Caselli M.C. & Casadio P. (1995). *Il primo vocabolario del bambino*. Milano: Franco Angeli Edizioni.

Charles-Luce J. & Luce P.A. (1995). An examination of similarity neighbourhoods in young children's receptive vocabularies. *Journal of Child Language*, vol. 22(3), 727-735.

Chen X., Striano T. & Rakoczy H. (2004). Auditory-oral matching behavior in newborns. *Developmental Science*, vol. 7, 42-47.

Chen Y-C., Shore D.I., Lewis T.L. & Maurer D. (2016). The development of the perception of audiovisual simultaneity. *Journal of Experimental Child Psychology*, vol. 146, 17-33.

Cheng H.Y., Murdoch B.E., Goozé J.V. & Scott D. (2007). Electropalatographic Assessment of Tongue-to-Palate Contact Patterns and Variability in Children, Adolescents, and Adults. *Journal of Speech, Language and Hearing Research*, vol. 50, 375-392.

Cheung C., Liberty S.H., Johnson K. & Chang E.F. (2016). The auditory representation of speech sounds in human motor cortex. *Elife*, vol. 5, e12577.

Chilosi A.M., Lorenzini I., Fiori S., Graziosi V., Rossi G., Pasquariello R., Cipriani P. & Cioni G. (2015). Behavioral and neurobiological correlates of Childhood Apraxia of Speech in Italian Children. *Brain & Language*, vol. 150, 177-185.

Choi D., Black A.K. & Werker J. (2018). Cascading and Multisensory Influences on Speech Perception Development. *Mind, Brain and Education*, doi: 10.1111/mbe.12162.

Christophe A., Mehler J., Sebastian-Gallos N. (2001). Perception of prosodic boundary correlates by newborn infants. *Infancy*, vol. 2, 385-394.

Christophe A., Dupoux E., Bertoncini J., Mehler J. (1994). Do infants perceive word boundaries? An empirical study of the bootstrapping of lexical acquisition. *Journal of the Acoustic Society of America*, vol. 95(3), 1570-80.

Chunming L., Long Y., Zheng L., Shi G., Li L., Ding G. & Howell P. (2016). Relationship between Speech Production and Perception in People Who Stutter. *Frontiers in Human Neuroscience*, vol. 10, doi: 10.3389/fnhum.2016.00224.

Colin C., Radeau M., Soquet A., Demolin D., Colin F. & Deltenre P. (2002). Mismatch negativity evoked by the McGurk–MacDonald effect: A phonetic representation within short-term memory. *Clinical Neurophysiology*, vol. 113(4), 495-506.

Conant L.L., Liebenthal E., Desai A. & Binder J.R. (2014). FMRI of phonemic perception and its relationship to reading development in elementary- to middle-school-age children. *Neuroimage*, vol. 89, 192-202.

Conboy B.T., Sommerville J.A. & Kuhl P.K. (2008). Cognitive control factors in speech perception at 11 months. *Developmental Psychology*, vol. 44(5), 1505-1512.

Cooper A., Fecher N. & Johnson E.K. (2018). Toddlers' comprehension of adult and child talkers: Adult targets versus vocal tract similarity. *Cognition*, vol. 173, 16-20.

Corbera S., Corral M.J., Escera C. & Idiazábal M.A. (2005). Abnormal speech sound representation in persistent developmental stuttering. *Neurology*, vol. 65(8), 1246-52.

Coulon M., Hemimou C. & Streri A. (2013). Effects of seeing and hearing vowels on neonatal facial imitation. *Infancy*, vol. 18, 782-796.

Creel S. C., Aslin R. N. & Tanenhaus M. K. (2006). Acquiring an artificial lexicon: Segment type and order information in early lexical entries. *Journal of Memory and Language*, vol. 54(1), 1-19.

- D -

D'Ausilio A., Bufalari I., Salmas P. & Fadiga L. (2012). The role of the motor system in discriminating normal and degraded speech sounds. *Cortex*, vol. 48(7), 882-887, 10.1016/j.cortex.2011.05.017.

D'Ausilio A., Pulvermüller F., Salmas P., Bufalari I., Begliomini C. & Fadiga L. (2009). The Motor Somatotopy of Speech Perception. *Current Biology*, vol. 19(5), 381-385.

Danielson D.K., Bruderer A., Kandhadai P., Vatikiotis-Bateson E. & Wercker J. (2017) The organization and reorganization of audiovisual speech perception in the first year of life. *Cognitive development*, vol. 42, 37-48.

Danielson D.K., Tam C., Kandhadai P. & Werker J.F. (2016). Infants' use of temporal and phonetic information in the encoding of audiovisual speech. *Journal of the Canadian Acoustical Association - Révue de l'Association canadienne d'acoustique*, vol. 44(3).

Dautriche I., Swingley D. & Cristophe A. (2015). Learning novel phonological neighbors: Syntactic category matters. *Cognition*, vol. 143, 77-86.

Davis B.L., Jacks A. & Marquardt T. (2005). Vowel patterns in developmental apraxia of speech: Three longitudinal case studies. *Clinical Linguistics & Phonetics*, vol. 19(4), 249–274.

Davis B.L. & Velleman S.L. (2000). Differential diagnosis and treatment of Developmental Apraxia of Speech in infants and toddlers. Infant Toddler Intervention, vol. 10(3), 177-192.

Davis B., Jakielski K. & Marquardt T. (1998). Developmental apraxia of speech: Determiners of differential diagnosis. Clinical Linguistics and Phonetics, vol. 12, 25-45.

Dawson M. (2014). Embedded and Situated Cognition. In: Shapiro L., ed. The Routledge Handbook of Embodied Cognition. New York: Routdlege Handbooks.

DeCasper A. J., Lecanuet J.P., Busnel M.-C., Granier-Deferre C. & Maugeais R. (1994). Fetal reactions to recurrent maternal speech. *Infant Behavior and Development*, vol. 17(2), 159-164.

DeCasper A. J. and Spence M. J. (1986) Prenatal maternal speech influences newborn's perception of speech sounds. *Infant Behavior and Development*, vol. 9, 133–150

DeCasper A. J. and Fifer W. P. (1980) Of human bonding: newborns prefer their mother's voices. *Science*, vol. 208, 1174–1176.

De Paolis R., Vihman M.M. & Keren-Portnoy T. (2014). When do infants begin recognizing familiar words in sentences? *Journal of Child Language*, 41, 226-239.

De Paolis R.A., Vihman M.M., Nakai S. (2013). The influence of babbling patterns on the processing of speech. *Infant Behavior & Development*, vol. 36(4), 642-649.

De Paolis R.A., Vihman M.M., Keren-Portnoy T. (2010). Do production patterns influence the processing of speech in prelinguistic infants? *Infant Behavior and Development*, vol. 34(4), 590-601.

Dehaene-Lambertz G. & Spelke E.S. (2015). The Infancy of the Human Brain. *Neuron*, vol. 88(1), 93-109.

Dehaene-Lambertz G., Montavont A., Jobert A., Alliroi L., Dubois J., Hertz-Pannier L. & Dehaene S. (2010). Language or music, mother or Mozart? Structural and environmental influences on infants' language networks. *Brain & Language*, vol. 114(2), 53-65.

Dehaene-Lambertz G., Hertz-Pannier L. & Dubois J. (2006a). Nature and nurture in language acquisition: anatomical and functional brain-imaging studies in infants. *Trends in neurosciences*, vol. 29(7), 367-373.

Dehaene-Lambertz G., Hertz-Pannier L., Dubois J., Mériaux S., Roche A., Sigman M. & Dehaene S. (2006b). Functional organization of perisylvian activation during presentation of sentences in preverbal infants. *PNAS*, vol. 103(38), 14240-14245.

Dehaene-Lambertz G., Pallier C., Serniclaes W., Sprenger-Charolles L., Jobert A. & Dehaene S. (2005). Neural correlates of switching from auditory to speech perception. *Neuroimage*, vol. 24(1), 21-33.

Dehaene-Lambertz G. & Gliga T. (2004). Common neural basis for phoneme processing in infants and adults. *Journal of Cognitive Neuroscience*, vol. 16(8), 1375-1387.

Dehaene-Lambertz G., Peña M., Christophe A. & Landrieu P. (2004). Phoneme perception in a neonate with a left sylvian infarct. *Brain & Language*, vol. 88(1), 26-38.

Dehaene-Lambertz G., Dehaene S. & Hertz-Pannier L. (2002). Functional neuroimaging of speech perception in infants. *Science*, vol. 298(5600), 2013-2015.

Dehaene-Lambertz G. & Peña M. (2001). Electrophysiological evidence for automatic phonetic processing in neonates. *Neuroreport*, vol. 12(14), 3155-3158.

Dehaene-Lambertz G. (2000) Cerebral specialization for speech and non-speech stimuli in infants. *Journal of Cognitive Neuroscience*, vol. 12, 449–460.

Dehaene-Lambertz G. & Baillet S. (1998). A phonological representation in the infant brain. *NeuroReport*, vol. 9, 1885–1888.

Dehaene-Lambertz G., Dehaene S. (1994). Speed and cerebral correlates of syllable discrimination in infants. *Nature*, vol. 370(6487), 292–95.

Desjardins R.N. & Werker J.F. (2004). Is the integration of heard and seen speech mandatory for infants? *Developmental Psychobiology*, vol. 45(4), 187-203.

Dejardins R.N., Rogers J. & Werker J. (1997). An exploration of why preschoolers perform differently than do adults in audiovisual speech perception tasks. *Journal of Experimental Child Psychology*, vol. 66(1), 85-110.

De Boisferon A.H., Tift A.H., Minar N.J. & Lewkowicz D.J. (2016). Selective attention to a talker's mouth in infancy: role of audiovisual temporal synchrony and linguistic experience. *Developmental Science*, vol. 20(3), doi: 10.1111/desc.12381.

Dick A.S., Solodkin A. & Small S.L. (2010). Neural development of networks for audiovisual speech comprehension. *Brain and Language*, vol. 114, 101-114.

Dietrich C., Swingley D. & Werker J.F. (2007). Native language governs interpretation of salient speech sound differences at 18 months. *PNAS*, vol. 104(41), 16027-31.

Di Pellegrino G., Fadiga L., Fogassi L., Gallese V. & Rizzolatti G. (1992). Understanding motor events: a neurophysiological study. *Experimental brain research*, vol. 91(1), 176-180.

Dodd B., McIntosch B., Erdener D. & Burnham D. (2008). Perception of the auditory-visual illusion in speech perception by children with phonological disorders. *Clinical Linguistics & Phonetics*, vol. 22(1), 69-82.

Dodd, B. (1987). The acquisition of lip-reading skills by normally hearing children. In: Dodd B., Campbell R., eds.. *Hearing by eye: The psychology of lip-reading*, pp. 163-175. Hillsdale, NJ: Lawrence Erlbaum Associates

Dove G., ed. (2016). *Beyond the body? The Future of Embodied Cognition*. Lausanne: Frontiers Media. doi: 10.3389/978-2-88919-797-2

Dubois J., Poupon C., Thirion B., Simonnet H., Kulikova S., Leroy F., Hertz-Pannier L. & Dehaene-Lambertz G. (2016). Exploring the Early Organization and Maturation of Linguistic Pathways in the Human Infant Brain. *Cerebral Cortex*, vol. 26, 2283-2298.

Dubois J., Hertz-Pannier L., Cachia A., Mangin JF., Le Bihan D., Dehaene-Lambertz G. (2009). Structural asymmetries in the infant language and sensori-motor networks. *Cerebral cortex*, vol. 19(2), 414-423.

Duffy J. R. (2000). Motor speech disorders: clues to neurologic diagnosis. In: *Parkinson's Disease and Movement Disorders*. Humana Press, Totowa, NJ, pages 35-53.

D'Souza D. & Karmiloff-Smith A. (2011). When modularization fails to occur: a developmental perspective. *Cognitive Neuropsychology*, vol. 28(3-4), 276-287.

- E -

Eimas P.D. (1975). Auditory and phonetic coding of the cues for speech: Discrimination of the [r-l] distinction by young infants. *Perception & Psychophysics*, vol. 18(5), 341-347.

Eimas P.D., Siqueland E.R., Jusczyk P.W., Vigorito J. (1971). Speech perception in infants. *Science*, vol. 171(968), 303-306.

Eisenberg L.S., Shannon R.V., Martinez A.S., Wygonski J. & Boothroyd A. (2000). Speech recognition with reduced spectral cues as a function of age. *The Journal of the Acoustical Society of America*, vol. 107, 2704–2710.

Elliott L.L., Longinotti C., Meyer D., Raz I. & Zucker K. (1981). Developmental differences in identifying and discriminating CV syllables. *Journal of the Acoustical Society of America*, vol. 70, 669-677.

Elliott L.L. (1979). Performance of children aged 9 to 17 years on a test of speech intelligibility in noise using sentence material with controlled word predictability. *The Journal of the Acoustical Society of America*, vol. 66(3), 651-3.

Elsabbagh M., Hohenberger A., Campos R., Herwegen J.V., Serres J., de Schonen S., Aschersleben G. & Karmiloff-Smith A. (2013). Narrowing Perceptual Sensitivity to the Native Language in Infancy: Exogenous Influences on Developmental Timing. *Behavioral Sciences*, vol. 6;3(1), 120-132.

Erdener D. & Burnham D. (2013). The relationship between auditory-visual speech perception and language-specific speech perception at the onset of reading instruction in English-speaking children. *Journal of Experimental Child Psychology*, vol. 116(2), 120-38.

- F -

Fadiga L. & Craighero L. (2006). Hand actions and speech representation in Broca's area. *Cortex*, vol. 42(4), 486-90.

Fadiga L., Craighero L., Buccino G. & Rizzolatti G. (2002). Speech Listening Specifically Modulates the Excitability of Tongue Muscles: A TMS Study. *European Journal of Neuroscience*, vol. 15, 399-402.

Fais L., Kajikawa S., Amano S. & Werker J.F. (2010). Now you hear it, now you don’t: Vowel devoicing in Japanese infant-directed speech. *Journal of Child Language*, vol. 37(2), 319-40.

Fausey C.M., Jayaraman S. & Smith L.B. (2016). From faces to hands: Changing visual input in the first two years. *Cognition*, vol. 152, 101-107.

Fennell C.T. & Waxman S.R. (2010). What Paradox? Referential Cues Allow for Infant Use of Phonetic Detail in Word Learning. *Child Development*, vol. 81(5), 1376-1383.

Fennell C.T., Byers-Heinlein K. & Werker J.F. (2007). Using speech sounds to guide word learning: The case of bilingual infants. *Child Development*, vol. 78(5), 1510-25.

Fennel C.T. & Werker J.F. (2003). Early Word Learners' Ability to Access Phonetic Detail in Well-Known Words. *Language and Speech*, vol. 46(2-3), 245-264.

Ferrari P.F., Gallese V., Rizzolatti G. & Fogassi G. (2003). Mirror neurons responding to the observation of ingestive and communicative mouth actions in the monkey ventral premotor cortex. *European Journal of Neuroscience*, vol. 17(8), 1703-14.

Fitch W.T. & Giedd J. (1999). Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *The Journal of the Acoustical Society of America*, vol. 106(3), 1511-1522.

Files B.T., Bosco S.T., Jiang J. & Bernstein L.E: (2015). Visual speech discrimination and identification of natural and synthetic consonant stimuli. *Frontiers in Psychology*, vol. 6, 878.

Fiori S., Guzzetta A., Mitra J., Pannek K., Pasquariello R., Cipriani P., Tosetti M., Cioni G., Rose S.E. & Chilosi A. (2016). Neuroanatomical correlates of childhood apraxia of speech: A connectomic approach. *Neuroimage: Clinical*, vol.12, 894-901.

Fitch W.T. & Giedd J. (1999). Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *The Journal of the Acoustical Society of America*, vol. 106(3), 1511-22.

Flinkers A., Korzeniewska A., Shestyuk A.Y., Franaszczu P.J., Dronkers N.F., Knight R.T. & Crone N.E. (2015). Redefining the role of Broca's area in speech. *PNAS*, vol. 112(9), 2871-2875.

Floccia C., Nazzi T., Delle Luche C., Poltrock S. & Goslin J. (2014). English-learning one-to two-year-olds do not show a consonant bias in word learning. *Journal of Child Language*, vol. 41(5), 1085-1114.

Floccia C., Nazzi T. & Bertoncini J. (2000). Unfamiliar voice discrimination for short stimuli in newborns. *Developmental Science*, vol. 3(3), 333-343.

Fowler A.E. (1991). How early phonological development might set the stage for phoneme awareness. In: Brady S.A. & Shankweiler D.P. (Eds.), *Phonological processes in literacy: A tribute to Isabelle Y. Liberman*, 97-117. Hillsdale, New Jersey: Erlbaum.

Froud K. & Khamis-Dakwar R. (2012). Mismatch Negativity Responses in children with a diagnosis of Childhood Apraxia of Speech (CAS). *American Journal of Speech-Language Pathology*, vol. 21(4), 302-312.

- G -

Gadesmann M. & Miller N. (2007). Reliability of speech diadochokinetic test measurement. *International Journal of Language & Communication Disorders*, vol. 43, 41-54.

Gervain J., Werker J.F., Black A. & Geffen M.N. (2016). The neural correlates of processing scale-invariant environmental sounds at birth. *Neuroimage*, vol. 133, 144-150.

Gervain J. & Werker J.F. (2013). Prosody cues word order in 7-month-old bilingual infants. *Nature Communications*, 4, 1490.

Gervain J., Macagno F., Cogoi S., Peña M. & Mehler J. (2008). The neonate brain detects speech structure. *PNAS*, vol. 105(37), 14222-14227.

Gervain J. & Werker J. (2008). How infant speech perception contributes to language acquisition. *Language and Linguistics Compass*, vol. 2(6), 1149–1170.

Ghanzafar AA. (2013). Multisensory vocal communication in primates and the evolution of rhythmic speech. *Behavioral Ecology and Sociobiology*, vol. 67(9), 1381-1388.

Gliga T. & Dehaene-Lambertz G. (2005). Structural encoding of body and face in human infants and adults. *Journal of cognitive neuroscience*, vol. 17(8), 1328-1340.

Goffman L. (2010). Dynamic interaction of motor and language factors in normal and disordered development. In: Maassen B., van Lieshout P. (2010), eds., *Speech Motor Control*. Oxford University Press, 137-152.

Goffman L. & Smith A. (1999). Development and phonetic differentiation of speech movement patterns. *Journal of Experimental Psychology: Human Perception and Performance*, vol. 25(3), 649.

Gómez D.M., Berent I., Benavides-Varela S., Bion R.H., Cattarossi L., Nespor M. & Mehler J. (2014). Language universals at birth. PNAS, vol. 111(16), 5837-41.

Gonzalez-Gomez N. & Nazzi T. (2012). Acquisition of Nonadjacent Phonological Dependencies in the Native Language During the First Year of Life. *Infancy*, vol. 17(5), pp. 498-524.

Goren C. C., Sarty M. & Wu P. Y. (1975). Visual following and pattern discrimination of face-like stimuli by newborn infants. *Pediatrics*, vol. 56(4), 544-549.

Green J.R. & Nip I.S.B. (2010). Some organizational principles in early speech development. In: Maassen B. & van Lieshout P., eds. *Speech Motor Control: New developments in basic and applied research*. Oxford: Oxford University Press.

Green J.R., Moore C.A. & Reilly K.J. (2002). The Sequential Development of Jaw and Lip Control for Speech. *Journal of Speech, Language, and Hearing Research*, vol. 45, 66-79.

Green J.R., Moore C.A., Higashikawa M., Steeve R.W. (2000). The Physiological Development of Speech Motor Control: Lip and Jaw Coordination. *Journal of Speech, Language, and Hearing Research*, vol. 43, 239-255.

Griffits S.K., Brown W.S. Jr., Gerhardt K.J. (1994). The perception of speech sounds recorded within the uterus of a pregnant sheep. *The Journal of the Acoustical Society of America*, vol. 96.

Grigos M.I. (2009). Changes in Articulator Movement Variability During Phonemic Development: A Longitudinal Study. *Journal of Speech, Language and Hearing Research*, vol. 52, 164-177.

Grigos M.I., Kolenda N. (2010). The relationship between articulatory control and improved phonemic accuracy in childhood apraxia of speech: a longitudinal case study. *Clinical linguistics and phonetics*, vol. 24, 17-40.

Grigos M.I., Moss A. & Lu Y. (2015). Oral Articulatory Control in Childhood Apraxia of Speech. *Journal of Speech, Language, and Hearing Research*, vol. 58(4), 1103-1118.

Groenen P., Maassen B., Crul T., Thoonen G. (1996). The Specific Relation Between Perception and Production Errors for Place of Articulation in Developmental Apraxia of Speech. *Journal of Speech and Hearing Research*, vol. 39(3), 468-482.

Guellaï B., Streri A., Chopin A., Rider D. & Kitamura C. (2016). Newborns' sensitivity to the visual aspects of infant-directed speech: Evidence from point-line displays of talking faces. *Journal of Experimental Psychology: Human Perception and Performance*, vol. 42(9), 1275-1281

Guellaï B., Mersad & Streri (2015). Suprasegmental information affects processing of talking faces at birth. *Infant Behavior and Development*, vol. 38, 11-19.

Guellaï B., Streri A. & Yeung H.H. (2014). The development of sensorimotor influences in the audiovisual speech domain: Some critical questions. *Frontiers in Psychology*, vol. 5(812), 1-7.

Guellaï B., Coulon M. & Streri A. (2011). The role of motion and speech in face recognition at birth. *Visual Cognition*, vol. 19(9), 1212-1233.

Guenther F.H. & Perkell J.S. (2004). A neural model of speech production and its application to studies of the role of auditory feedback in speech. In: Maassen B., Kent R., Peters

H., Van Lieshout P., Hulstijn W., eds. *Speech Motor Control in Normal and Disordered Speech*. Oxford: Oxford University Press.

- H -

Hallé P.A. & de Boysson-Bardies B. (1994). Emergence of an early receptive lexicon: Infants' recognition of words. *Infant Behavior and Development*, vol. 17(2), 119-129.

Hallé P.A. & de Boysson-Bardies B. (1996). The format of representation of recognized words in infants' early receptive vocabulary. *Infant Behavior and Development*, vol. 19, 463-481.

Halle M. & Stevens K. (1962). Speech recognition: A model and a program for research. *IRE transactions on information theory*, vol. 8(2), 155-159.

Hauser M.D., Chomsky N. & Fitch W.T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science*, vol. 298(5598), 1569-1579.

Havy M., Foroud A., Fais L. & Werker, J.F. (2016). The role of auditory and visual speech in word-learning at 18 months and in adulthood. *Child Development*, vol. 88(6), 2043-2059.

Havy M., Serres J. & Nazzi T. (2014). A Consonant/Vowel Asymmetry in Word-form Processing: Evidence in Childhood and in Adulthood. *Language and Speech*, vol. 57(Pt2), 254-81.

Havy M. & Nazzi T. (2009). Better Processing of Consonantal Over Vocalic Information in Word Learning at 16 Months of Age. *Infancy*, vol. 14(4), 439-456.

Hayden D.A. & Square P.A. (1986). Motor Speech Treatment Hierarchy: a systems approach. *Clinics in Communication Disorders*, vol. 4(3), 162-174.

Hayes E.A., Tiippana K., Nicol T.G., Sams M. & Kraus N. (2003). Integration of heard and seen speech: a factor in learning disabilities in children. *Neuroscience Letters*, vol. 351(1), 46-50.

Hazan V. & Barrett S. (2000). The development of phonemic categorization in children aged 6–12. *Journal of Phonetics*, vol. 28(4), 377-96.

Hickok G. (2010). The role of mirror neurons in speech perception and action word semantics. *Language and Cognitive Processes*, vol. 25(6), 749-776.

Hickok G., Costanzo M., Capasso R. & Miceli G. (2011). The role of Broca's area in speech perception: evidence from aphasia revisited. *Brain & Language*, vol. 119(23), 214-220.

Hickok G. & Poeppel D. (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition*, vol. 92(1-2), 67-99.

Hickok G. & Poeppel D. (2007). The cortical organization of speech processing. *Natural reviews. Neuroscience*, vol. 8(5), 393-402.

Highman C., Leitão S., Hennessey N., Piek J. (2012). Prelinguistic communication development in children with childhood apraxia of speech: A retrospective analysis. *International Journal of Speech-Language Pathology*, vol. 14(1), 35-47.

Hillock-Dunn A., Grantham D.W. & Wallace M.T. (2016). The temporal binding window for audio visual speech: children are like little adults. *Neuropsychologia*, vol. 88, 74-82.

Hillock A.R., Powers A.R. & Wallace M.T. (2011). Binding of sights and sounds: Age-related changes in multisensory temporal processing. *Neuropsychologia*, vol. 49(3), 461-467.

Hochmann J-R., Benavides S., Nespor M. & Mehler J. (2011). Consonants and Vowels: different roles in early language acquisition. *Developmental Science*, vol. 14(6), 1445-1458.

Hockley N.S. & Polka L. (1994). A developmental study of audiovisual speech perception using the McGurk paradigm. *The Journal of the Acoustical Society of America*, vol. 96(5), 3309-3309.

Hoffman P.R., Daniloff R.G., Bengoa D. & Schuckers G.H. (1985). Misarticulating and normally articulating children’s identification and discrimination of synthetic [r] and [w]. Journal of Speech and Hearing Disorders, vol. 50(1), 46-53.

Höhle B., Bijeljac-Babic R., Herold B., Weissenborn J. & Nazzi T. (2009). Language specific prosodic preferences during the first half year of life: evidence from German and French infants. *Infant behavior & development*, vol. 32(3), 262-274.

Højen A. & Nazzi T. (2016). Vowel bias in Danish word-learning: processing biases are language-specific. *Developmental Science*, vol. 19(1), 41-49.

Holden-Pitt L.D., Hazan V., Revoile S.G., Edward D.M. & Droge J. (1995). Temporal and spectral cue-use for initial plosive voicing perception by hearing-impaired children and normal-hearing children and adults. European Journ of Disorders of Communication: the journal of the College of Speech and Language Therapists, London, vol. 30(4), 417-34.

Hollich G., Newman R.S. & Jusczyk P.W. (2005). Infants' Use of Synchronized Visual Information to Separate Streams of Speech. *Child Development*, vol. 76(3), 598-613.

Hyde D.C., Jones B.L., Flom R. & Porter C.L. (2011). Neural signatures of face-voice synchrony in 5-month-old human infants. *Developmental Psychobiology*, vol. 53(4), 359-370.

- I -

Idemaru K. & Holt L.L. (2013). The developmental trajectory of children's perception and production of English /r/-/l/. *The Journal of the Acoustical Society of America*, vol. 133(6), 4232-46.

Imada T., Zhang Y., Cheour M., Taulu S., Ahonen A. & Kuhl P.K. (2006). Infant speech perception activates Broca's area: a developmental magnetoencephalography study. *Neuroreport*, vol. 17(10), 957-962.

Iuzzini-Seigel J., Hogan P.T., Guarino A.J., Green J.R. (2015). Reliance on auditory feedback in children with childhood apraxia of speech. *Journal of Communication Disorders*, vol. 54, 32-42.

Iuzzini-Seigel J., Hogan P.T., Rong P. & Green J.R. (2015). Longitudinal development of speech motor control: Motor and linguistic factors. *Journal of Motor Learn Development*, vol. 3(1), 53-68.

- J -

Jacks A., Marquardt T.P., Davis B.L. (2006). Consonant and syllable structure patterns in childhood apraxia of speech: developmental change in three children. *Journal of communication disorders*, vol. 39, 424-441.

Jääskeläinen I.P. (2010). The Role of the Speech Production System in Audiovisual Speech Perception. *The Open Neuroimaging Journal*, vol. 10(4), 30-36.

Jerger S., Damian M.F., McAlpine R.P. & Abdi H. (2018). Visual speech fills in both discrimination and identification of non-intact auditory speech in children. *Journal of Child Language*, vol. 45(2), 392-414.

Jerger S., Damian M.F., Tie-Murray N. & Abdi H. (2017). Children perceive speech onsets by ear and eye*. *Journal of Child Language*, vol. 44(1), 185-215.

Jerger S., Damian M.F., Tie-Murray N. & Abdi H. (2014). Children use visual speech to compensate for non-intact auditory speech. *Journal of Experimental Child Psychology*, vol. 126, 295-312.

Jiang J., Auer E.T. Jr., Alwan A., Keating P.A. & Bernstein L.E. (2007). Similarity structure in visual speech perception and optical phonetic signals. *Perception & Psychophysics*, vol. 69(7), 1070-83.

Johnson C.E. (2000). Children's phoneme identification in reverberation and noise. *Journal of Speech, Language and Hearing Research*, vol. 43, 144-57.

Johnson E.K. & Seidl A.H. (2009). At 11 months, prosody still outranks statistics. *Developmental Science*, vol. 12(1), 131-141.

Jones P.R., Moore D.R. & Sygal A. (2015). Development of auditory selective attention: Why children struggle to hear in noisy environments. *Developmental Psychology*, vol. 51(3), 353-69.

Jusczyk P. W., Houston D. M. & Newsome M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology*, vol. 39, 159-207.

Jusczyk P.W. & Aslin R.N. (1995). Infants′ Detection of the Sound Patterns of Words in Fluent Speech. *Cognitive Psychology*, vol. 29(1), 1-23.

Jusczyk P.W., Cutler A. & Redanz N. (1993). Preference for the predominant stress patterns of English words. *Child Development*, vol. 64, 675-687.

- K -

Kaganovich N. (2016). Development of sensitivity to audiovisual temporal asynchrony during mid-childhood. *Developmental Psychology*, vol. 52(2), 232-241.

Kaganovich N. & Schumaker J. (2014). Audiovisual integration for speech during mid-childhood: electrophysiological evidence. *Brain & Language*, vol. 139, 36-48.

Kaganovich N., Schumaker J., Macias D. & Gustafson D. (2014). Processing of audiovisually congruent and incongruent speech in school-age children with a history of specific language impairment: a behavioral and event-related potentials study. *Developmental Science*, vol. 18(5), 751-770.

Kajikawa S., Fais L., Mugitani R., Werker J.F. & Amano S. (2006). Cross-language sensitivity to phonotactic patterns in infants. *Journal of the Acoustical Society of America*, vol. 120(4), 2278-84.

Karmiloff-Smith A. (2015). An Alternative to Domain-general or Domain-specific Frameworks for Theorizing about Human Evolution and Ontogenesis. *AIMS Neuroscience*, vol. 2(2), 91-104.

Karmiloff-Smith A. (2013b). Challenging the use of adult neuropsychological models for explaining neurodevelopmental disorders: Developed versus developing brains. *The Quarterly Journal of Experimental Psychology*, vol. 66, 1-14.

Karmiloff-Smith, A. (2009b). Nativism Versus Neuroconstructivism: Rethinking the Study of Developmental Disorders. *Developmental Psychology*, vol. 45, Issue 1, 56-63.

Karmiloff-Smith, A. (2006). The tortuous route from genes to behaviour: A neuroconstructivist approach. *Cognitive, Affective and Behavioural Neuroscience*, vol. 6(1), 9-17.

Kent R.D. (1984). Psychobiology of speech development: coemergence of language and a movement system. *American Journal of Physiology - Regulatory, Integrative and Comparative Physiology*, vol. 246(6), R888-R894.

Kisilevsky B.S., Hains S.M.J., Lee K., Xie X., Huang H., Ye H.H., Zhang K. & Wang Z. (2003). Effects of experience on fetal voice recognition. *Psychological Science*, vol. 14(3), 220-4.

Kislyuk D.S., Möttonen R. & Sams M. (2008). Visual Processing Affects the Neural Basis of Auditory Discrimination. *Journal of Cognitive Neuroscience*, vol. 20(12), 2175-2184.

Kitamura C., Guellaï B. & Kim J. (2014). Motherese by eye and ear: Infants perceive visual prosody in point-line displays of talking heads. *PLoS ONE*, vol. 9(10), e111467.

Knowland V.C.P., Mercure E., Karmiloff-Smith A., Dick F. & Thomas M.S.C. (2014). Audio-visual speech perception: a developmental ERP investigation. *Developmental Science*, vol. 17(1), 110-124.

Koenig L.L., Lucero J.C. & Perlman E. (2008). Speech production variability in fricatives of children and adults: Results of functional data analysis. *The Journal of the Acoustical Society of America*, vol. 124(5).

Kolinsky R., Morais J., Cohen L., Dehaene-Lambertz G. & Dehaene S. (2014). L'influence de l'apprentissage du langage écrit sur les aires du langage / The impact of literacy on the language brain areas. *Neuropsychological Review*, vol. 6(3), 173-181.

Kooijman V., Hagoort P. & Cutler A. (2005). Electrophysiological evidence for prelinguistic infants' word recognition in continuous speech. *Cognitive Brain Research*, vol. 24(1), 109-116.

Krause S.F. (1982). Vowel duration as a perceptual cue to post-vocalic consonant voicing in young children and adults. *Journal of the Acoustical Society of America*, vol. 71, 990-995.

Kubicek C., Hillairet de Boisferon A., Dupierrix E., Pascalis O., Lœvenbruck H., Gervain J. & Schwarzer G. (2014). Cross-modal matching of audio-visual German and French fluent speech in infancy. *PloSOne*, vol. 9(2), e89275.

Kubicek C., de Boisferon A.H., Dupierrix E., Lœvenbruck H., Gervain J. & Schwarzer G. (2013). Face-scanning behavior to silently-talking faces in 12-month-old infants: The impact of pre-exposed auditory speech. *International Journal of Behavioral Development*, vol. 00(0), 1-5.

Kuhl P.K., Ramírez R.R., Bosseler A., Lotus J-F. & Imada T. (2014). Infants' brain responses to speech suggest Analysis by Synthesis. *PNAS*, vol. 111(31), 11238-11245.

Kuhl P.K., Conboy B.T., Coffey-Corina S., Padden D., Rivera-Gaxiola M. & Nelson T. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). *Philosophical transactions of the Royal Society B - Biological Science*, vol. 363(1493), 979-1000.

Kuhl P.K., Stevens E., Hayashi A., Deguchi T., Kiritani S. & Iverson P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science*, vol. 9(2), F13-F21.

Kuhl P.K., Conboy B.T., Padden D., Nelson T. & Pruitt J. (2005). Early Speech Perception and Later Language Development: Implications for the “Critical Period”. *Language learning and development*, vol. 1(3&4), 237-264.

Kuhl P.K. (2004). Early language acquisition: cracking the speech code. *Nature Reviews: Neuroscience*, vol. 5(11), 831-43.

Kuhl P.K., Andruski J.E., Chistovich I.A., Chistovich L.A., Kozhevnikova E.V., Ryskina V.L., Stolyarova E.I., Sundberg U. & Lacerda F. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science*, vol. 277(5326), 684-686.

Kuhl P.K. & Meltzoff A.N. (1996). Infants’ vocalizations in response to speech: Vocal imitation and developmental change. *The Journal of the Acoustic Society of America*, vol. 100(4 Pt 1), 2425-2438.

Kuhl P.K., Williams K.A., Lacerda F., Stevens K.N. & Lindblom B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, vol. 255(5044), 606-608.

Kuhl P.K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics*, vol. 50(2), 93-107.

Kuhl P.K., Williams K.A. & Meltzoff A.N. (1991). Cross-modal speech perception in adults and infants using non-speech auditory stimuli. *Journal of Experimental Psychology. Human perception and performance*, vol. 17(3), 829-840.

Kuhl P.K. & Meltzoff A.N. (1982). The bimodal perception of speech in infancy. *Science*, vol. 218(4577), 1138-1141.

Kushnerenko E., Tomalski P., Ballieux H., Potton A., Birtles D., Frostick C. & Moore D.G. (2013a). Brain responses and looking behavior during audiovisual speech integration in infants predict auditory speech comprehension in the second year of life. *Frontiers in Psychology*, 4:432.

Kushnerenko E., Tomalski P., Ballieux H., Ribeiro H., Potton A., Axelsson E.L., Murphy E. & Moore D.G. (2013b). Brain responses to audiovisual speech mismatch in infants are associated with individual differences in looking behaviour. *European Journal of Neuroscience*, vol. 38(9), 3363–3369.

Kushnerenko E., Teinonen T., Volein A., Csibra G. (2008). Electrophysiological evidence of illusory audiovisual speech percept in human infants. *PNAS*, vol. 105(32), 11442-114445.

- L -

Ladefoged P. & Maddieson I. (1996). The sounds of the world's languages. Oxford (UK): Blackwell Publishers Inc.

Lalonde K. & Holt F. R. (2016). Audiovisual speech perception development at varying levels of perceptual processing. *The Journal of the Acoustic Society of America*, vol. 139(4), 1713-1723.

Lalonde K. & Holt R.F. (2015). Preschoolers benefit from visually salient speech cues. *Journal of Speech, Language, and Hearing Research*, vol. 58(1), 135-150.

Lalonde C., & Werker J.F. (1995). Cognitive influences on cross-language speech perception in infancy. *Infant Behavior and Development*, vol. 18(4), 459-476.

Langus A. & Nespor M. (2013). Language development in infants: what do humans hear in the first months of life? *Hearing, Balance and Communication*, vol. 11(3), 121-129.

Lasky R.E., Syrdal-Lasky A. & Klein R.E. (1975). VOT discrimination by four to six and a half month old infants from Spanish environments. *Journal of Experimental Child Psychology*, vol. 20, 215-225.

Law I.I.F. & Edwards J.R. (2015). Effects of vocabulary size on online lexical processing by toddlers. *Language Learning and Development*, vol. 11, 331–355.

Lecanuet J., Granier-Deferre C., Jacquet A.-Y., Capponi I. & Ledru L. (1993). Prenatal discrimination of a male and a female voice uttering the same sentence. *Infant and Child Development*, vol. 2(4), 217-228.

Lecanuet J., Granier-Deferre C., DeCasper A.J., Maugeais R., Andrieu A., Busnel M. (1987). Perception et discrimination foetales de stimuli langagiers; mise en evidence a partir de la reactivité cardiaque; resultats preliminaires. *Comptes Rendus de l'Académie des Sciences*, vol. 305, 161-164.

LeNormand M.T., Vaivre-Douret L., Payan C., Cohen H. (2000). Neuromotor development and language processing in developmental dyspraxia: A follow-up case study. *Journal of Clinical and Experimental Neurophsychology*, vol. 22(3), 408-417.

Leroy F., Glasel H., Dubois J., Hertz-Pannier L., Thirion B., Mangin J-F. & Dehaene-Lambertz G. (2011). Early Maturation of the Linguistic Dorsal Pathway in Human Infants. *The journal of neuroscience*, vol. 31(4), 1500-1506.

Lewkowicz D.J., Minar N.J., Tift A. H. & Brandon M. (2015). Perception of the Multisensory Coherence of Fluent Audiovisual Speech in Infancy: Its Emergence & the Role of Experience. *Journal of Experimental Psychology*, vol. 130, 147-162.

Lewkowicz D.J. (2014). Early Experience & Multisensory Perceptual Narrowing. *Developmental Psychobiology*, vol. 56(2), 292-315.

Lewkowicz D.J. & Flom R. (2014). The Audiovisual Temporal Binding Window Narrows in Early Childhood. *Child Development*, vol. 85(2), 685-694.

Lewkowicz D.J. & Hansen-Tift A.M. (2012). Infants deploy selective attention to the mouth of a talking face when learning speech. *PNAS*, vol. 109(5), 1431-1436.

Lewkowicz D.J., Leo I. & Simion F. (2010). Intersensory Perception at Birth: Newborns Match Nonhuman Primate Faces and Voices. *Infancy*, vol. 15(1), 46-60.

Lewkowicz D.J. (2010). Infant Perception of Audio-Visual Speech Synchrony. *Developmental Psychology*, vol. 46(1), 66-77.

Lewkowicz D.J. & Ghazanfar A.A. (2009). The emergence of multisensory systems through perceptual narrowing. *Trends in Cognitive Sciences*, vol. 13(11), 470-8. *PNAS*, vol. 106(26), 10598-602.

Lewkowicz D.J. & Ghazanfar A.A. (2006). The decline of cross-species intersensory perception in human infants. *PNAS*, vol. 103(17), 6771-4.

Lewkowicz D.J. & Kraebel K.S. (2004). The Value of Multisensory Redundancy in the Development of Intersensory Perception. In: Calvert G., Spence, C. & Stein B.E., eds. *Handbook of multisensory processing*. MIT Press.

Lewkowicz D.J. (2003). Learning and discrimination of audiovisual events in human infants: The hierarchical relation between intersensory temporal synchrony and rhythmic pattern cues. *Developmental Psychology*, vol. 39(5), 795-804.

Lewkowicz D.J. (2000). The development of intersensory temporal perception: an epigenetic systems/limitations view. *Psychological bulletin*, vol. 126(2), 281-308.

Lewkowicz D. J. (1998). Infants' response to the audible and visible properties of the human face: II. Discrimination of differences between singing and adult-directed speech. *Developmental Psychobiology*, 32, 261-274.

Lewkowicz D.J. (1996a). Perception of auditory–visual temporal synchrony in human infants. *Journal of experimental psychology. Human perception and performance*, vol. 22(5), 1094-1106.

Lewkowicz D.J. (1996b). Infants' response to the audible and visible properties of the human face. I: Role of lexical-syntactic content, temporal synchrony, gender, and manner of speech. *Developmental Psychology*, vol. 32(2), 347-366.

Lewis B., Freebairn L., Hansen A., Iyengar S., Taylor H. (2004a). School-age follow-up of children with Childhood Apraxia of Speech. *Language Speech and Hearing Services in Schools*, vol. 35(2), 122-140.

Li N. & Loizou P.C. (2008). The contribution of obstruent consonants and acoustic landmarks to speech recognition in noise. *The Journal of the Acoustical Society of America*, vol. 124(6).

Liberman A.M. & Mattingly I.G. (1985). The motor theory of speech perception revised. *Cognition*, vol. 21(1), 1-36.

Liégeois F.J. (2017). Neurobiology of developmental coordination disorder: time to raise our game. *Developmental Medicine & Child Neurology*, vol. 59(11), 1103-1103

Liégeois F.J., Morgan A.T. (2011). Neural bases of childhood speech disorders: lateralization and plasticity for speech functions during development. *Neuroscience and biobehavioral reviews*, vol. 36, 439-458.

Liu H-M., Chen Y. & Tsao F-M (2014). Developmental Changes in Mismatch Responses to Mandarin Consonants and Lexical Tones from Early to Middle Childhood. *PLOSOne*, vol. 9(4).

Liu H-M., Kuhl P.K. & Tsao F-M. (2003). An association between mothers' speech clarity and infants' speech discrimination skills. *Developmental Science*, vol. 6(3), F1-F10.

Lotto A.J., Hickok G.S. & Holt L.L. (2009). Reflections on mirror neurons and speech perception. *Trends in cognitive sciences*, vol. 13(3), 110-114.

- M -

Maassen B., Nijland L. & Terband H. (2010). Developmental models of childhood apraxia of speech. In: Maassen B., van Lieshout P., eds., *Speech Motor Control*. Oxford University Press, 243-258.

Maassen B., Groenen P. & Crul T. (2003). Auditory and phonetic perception of vowels in children with apraxic speech disorders. *Clinical Linguistics and Phonetics*, vol. 17, 447-467.

Maassen B. (2002). Issues contrasting adult acquired versus developmental apraxia of speech. In: *Seminars in Speech and Language* (Vol. 23, No. 04, pp. 257-266). New York: Thieme Medical Publishers.

MacKain K., Studdert-Kennedy M., Spieker S. & Stern D. (1983). Infant intermodal speech perception is a left-hemisphere function. *Science*, vol. 219(4590), 1347-1349.

MacLeod, A., & Summerfield, Q. (1987). Quantifying the contribution of vision to speech perception in noise. *British journal of audiology*, vol. 21(2), 131-141.

Maddieson I. (2009). Calculating phonological complexity. In: Pellegrino F., Marsico E., Chitoran I. & Coupé C., eds. *Approaches to phonological complexity* (vol. 16). Walter de Gruyter. pages 85-110.

Magezi D.A. (2015). Linear mixed-effects models for within-participant psychology experiments: an introductory tutorial and free, graphical user interface (LMMgui). *Frontiers in Psychology*, vol. 6(22).

Mahmoudzadeh M., Dehaene-Lambertz G., Fournier M., Kongolo G., Goudjil S., Dubois J., Grebe R. & Wallois F. (2013). Syllabic discrimination in premature human infants prior to complete formation of cortical layers. *PNAS*, vol. 110(12), 4846-4851.

Maidment D.W., Kang H.J., Stewart H.J. & Amitay S. (2015). Audiovisual Integration in Children Listening to Spectrally Degraded Speech. *Journal of Speech, Language and Hearing Research*, vol. 58, 61-68.

Majorano M., Vihman M.M. & De Paolis R.A. (2013). The relationship between infants' production experience and their processing of speech. *Language Learning and Development*, vol. 10(2), 179-204.

Mani N. & Plunkett K. (2010). Twelve-month-olds know their cups from their keps and tups. *Infancy*, vol. 15(5), 445–70.

Mani N. & Plunkett K. (2007). Phonological specificity of consonants and vowels in early lexical representations. *Journal of Memory and Language*, vol. 57, 252–272.

Marion M.J., Sussman H.M. & Marquardt T.P. (1993). The perception and production of rhyme in normal and developmentally apraxic children. *Journal of Communication Disorders*, vol. 26(3), 129-160.

Marquardt T.P., Jacks A., Davis B.L. (2004). Token-to-Token Variability in Developmental Apraxia of Speech: Three Longitudinal Case Studies. *Clinical Linguistics and Phonetics*, vol. 18(2), 127-144.

Marquardt, T., Sussman, H. M., Snow, T., & Jacks, A. (2002). The integrity of the syllable in developmental apraxia of speech. *Journal of Communication Disorders*, vol. 35, 31–49.

Masapollo M., Polka L. & Ménard L. (2017). A universal bias in adult vowel perception – By ear or by eye. *Cognition*, vol. 166, 358-370.

Masapollo M., Polka L. & Ménard L. (2015). When infants talk, infants listen: Pre-babbling infants prefer listening to speech with infant vocal properties. *Developmental Science*, vol. 19(2), 318-28.

Massaro D.W., Thompson L.A., Barron B. & Laren E. (1986). Developmental changes in visual and auditory contributions to speech perception. *Journal of Experimental Child Psychology*, vol. 41(1), 93-113.

Massaro, D. W. (1984). Children's perception of visual and auditory speech. *Child Development*, vol. 55, 1777-1788.

Maurer D. & Werker J.F. (2014). Perceptual Narrowing During Infancy: A Comparison of Language and Faces. *Developmental Psychobiology*, vol. 56(2), 154-78, doi: 10.1002/dev.21177.

Mazuka R., Cao Y., Dupoux E. & Cristophe A. (2011). The development of a phonological illusion: a cross-linguistic study with Japanese and French infants. *Developmental Science*, vol. 14(4), 693-699.

May L., Byers-Heinlein K., Gervain J. & Werker J.F. (2011). Language and the Newborn Brain: Does Prenatal Language Experience Shape the Neonate Neural Response to Speech? *Frontiers in psychology*, 2, 222.

Maye J., Werker J.F. & Gerken L.A. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, vol. 82(3), B101-11.

Maye J., Weiss D.J. & Aslin R.N. (2008). Statistical phonetic learning in infants: facilitation and feature generalization. *Developmental Science*, vol. 11(1), 122-134.

Mayo C., Scobbie J.M., Hewlett N. & Waters D. (2003). The Influence of Phonemic Awareness Development on Acoustic Cue Weighting Strategies in Children's Speech Perception. *Journal of Speech, Language and Hearing Research*, vol. 46, 1184-96.

McAllister Byun T. & Tiede M. (2017). Perception-production relations in later development of American English rhotics. *PLOSOne*, vol. 12(2), e0172022,

McAllister Byun T. (2012). Bidirectional perception-production relations in phonological development: evidence from positional neutralization. *Clinical Linguistics & Phonetics*, vol. 26(5), 397-413.

McGurk H. & MacDonald J. (1976). Hearing lips and seeing voices. *Nature*, vol. 264, 746-748.

McMurray B. & Aslin R.N. (2005). Infants are sensitive to within-category variation in speech perception. *Cognition*, vol. 95(2), B15-B26.

McMurray B., Tanenhaus M. & Aslin R.N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition*, vol. 86(2), B33–B42.

McNeill B.C. & Gillon G.T. (2015). Expressive morphosyntactic development in three children with Childhood Apraxia of Speech. *Speech, Language and Hearing*, vol. 16(1), 9-17.

Medina V. & Serniclaes W. (2013). Late development of the categorical perception of speech sounds in pre-adolescent children. Universitätsbibliothek Johann Christian Senckenberg.

Medina V., Hoonhorst I., Bogliotti C. & Serniclaes W. (2010). Development of voicing perception in French: Comparing adults, adolescents, and children. *Journal of Phonetics*, vol. 38(4), 493-503.

Mehler J., Jusczyk P.W., Lambertz G., Halsted N., Bertoncini J. & Amiel-Tison C. (1988). A precursor of language acquisition in young infants. *Cognition*, vol. 29(2), 143–78.

Mehler J., Dehaene-Lambertz G. Jusczyk P. & Amiel-Tison C. (1986). Discrimination of the mother tongue by newborn infants. *Comptes rendus de l'Académie des sciences. Série III, Sciences de la vie*, vol. 303(15), 637-640.

Mehler J., Bertoncini J., Barriere M., Gerschenfeld D.J. (1977). Infant recognition of mother's voice. *Perception*, vol. 7, 491–97.

Meltzoff A.N. & Moore M.K. (1989). Imitation in newborn infants: Exploring the range of gestures imitated and the underlying mechanisms. *Developmental psychology*, vol. 25(6), 954.

Meltzoff A.N. & Moore M.K. (1977). Imitation of facial and manual gestures by human neonates. *Science*, vol. 198(4312), 75-78.

Meltzoff A.N. & Kuhl P.K. (1989). Infants' perception of faces and speech sounds: Challenges to developmental theory. In: Zelazo P.R. & Barr R.G. (Eds.), Challenges to developmental paradigms: Implications for theory, assessment and treatment, 67-91. Hillsdale, NJ: Erlbaum.

Mersad K. & Dehaene-Lambertz G. (2015). Electrophysiological evidence of phonetic normalization across coarticulation in infants. *Developmental science*, vol. 19(5), 710-722.

Messum P. & Howard I.S. (2015). Creating the cognitive form of phonological units: The speech sound correspondence problem in infancy could be solved by mirrored vocal interactions rather than by imitation. *Journal of Phonetics*, vol. 53, 125-140.

Messum, P. R. (2008). The role of imitation in learning to pronounce. PhD Thesis, University of London, University College London (United Kingdom).

Metsala J.L. & Walley A.C. (1998). Spoken vocabulary growth and the segmental restructuring of lexical representations: Precursors to phonemic awareness and early reading ability. In: Metsala J.L. & Ehri L.C., eds., *Word recognition in beginning literacy*, pp. 89-120. Hillsdale, NJ: Erlbaum.

Miller C. T., Dimauro A. Pistorio, A. Hendry, S. & Wang X. (2010). Vocalization induced CFos expression in marmoset cortex. *Frontiers in Integrative Neuroscience*, vol. 4, 128.

Miller C. T., Thomas A.W., Nummela S.U. & de la Mothe L.A. (2015). Responses of primate frontal cortex neurons during natural vocal communication. *Journal of Neurophysiology*, vol. 114(2), 1158-1171.

Mills D.L., Prat C., Zangl R., Stager C.L., Neville H.J., Werker J.F. (2004). Language experience and the organization of brain activity to phonetically similar words: ERP evidence from 14- and 20-month-olds. *Journal of Cognitive Neuroscience*, vol. 16(8), 1452-1464.

Monzalvo K. & Dehaene-Lambertz G. (2013). How reading acquisition changes children's spoken language network. *Brain & Language*, vol. 127(3), 356-365.

Moore C.A. (2004). Physiologic development of speech production. In: Maassen B., Kent R., Peters H., van Lieshout P. & Hulstijn W., eds. *Speech Motor Control in normal and disordered speech*. Oxford: Oxford University Press. pp. 191-210.

Moon C., Lagercrantz H. & Kuhl P.K. (2013). Language experienced in utero affects vowel perception after birth: a two-country study. *Acta Paediatrica*, vol. 102(2), 156-160.

Moon C., Cooper R.P., Fifer W.P. (1993). Two-day-olds prefer their native language. *Infant Behavioral Development*, vol. 16(4), 495-500.

Morgan A., Bonthrone A. & Liégeois F.J. (2016). Brain basis of childhood speech and language disorders: are we closer to clinically meaningful MRI markers? *Current Opinion in Pediatrics*, vol. 28(6), 725-730.

Möttönen R. & Watkins K.E. (2009). Motor representations of articulators contribute to categorical perception of speech sounds. *The journal of neuroscience: the official journal of the Society for Neuroscience*, vol. 29(31), 9819-25.

Möttönen R., Krause C.M., Tiippana K. & Sams M. (2002). Processing of changes in visual speech in the human auditory cortex. Cognitive Brain Research, vol. 13(3), 417–425.

Mugitani R. & Hiroa S. (2012). Development of vocal tract and acoustic features in children. *Acoustical Science and Technology*, vol. 33(4), 215-200.

Mugitani R., Pons F., Fais L., Werker J.F. & Amano S. (2009). Perception of vowel length by Japanese- and English-learning infants. Developmental Psychology, vol. 45(1), 236-47.

Mugitani R., Kobayashi T, & Hirakib K. (2008). Audiovisual matching of lips and non-canonical sounds in 8-month-old infants. *Infant Behavior and Development*, vol. 31(2), 307-310.

Mugitani R., Fais L., Kajikawa S., Werker J.F., & Amano S. (2007). Age-related changes in sensitivity to native phonotactics in Japanese infants. *Journal of the Acoustical Society of America*, vol. 122(3), 1332-5.

Murdoch B. E., Cheng H.Y. & Goozée J.V. (2011). Developmental changes in the variability of tongue and lip movements during speech from childhood to adulthood: An EMA study. *Clinical Linguistics & Phonetics*, vol. 26(3), 216-231.

Murdoch B. E., Attard M. D., Ozanne A. E. & Stokes P. D. (1995). Impaired tongue strength and endurance in developmental verbal dyspraxia: A physiological analysis. *European Journal of Disorders of Communication*, vol. 30(1), 51–64.

Murray M.M., Lewkowvicz D.J., Amedi A. & Wallace M.T. (2016). Multisensory Processes: A Balancing Act across the Lifespan. *Trends in Neurosciences*, vol. 39(8), pp. 567-579.

Murray E., McCabe P., Heard R. & Ballard K.J. (2015). Differential Diagnosis of Children with Suspected Childhood Apraxia of Speech. *Journal of Speech, Language, and Hearing Research*, vol. 58(1), 43-60.

Murray E., McCabe P. & Ballard K.J. (2014). A Systematic Review of Treatment Outcomes for Children With Childhood Apraxia of Speech. *American Journal of Speech-Language Pathology*, vol. 23(3), 486-504.

Murray E., McCabe P. & Ballard K.J. (2012). A comparison of two treatments for childhood apraxia of speech: methods and treatment protocol for a parallel group randomised control trial. BCM Pediatrics, doi: 10.1186/1471-2431-12-112.

Musacchia G., Sams M., Nicol T. & Kraus N. (2006). Seeing speech affects acoustic information processing in the human brainstem. *Experimental Brain Research*, vol. 168(1-2), 1-10.

- N -

Navarra J., Yeung H.H., Werker J.F. & Soto-Faraco S. (2012). Multisensory Interactions in Speech Perception. In: Stein B.E., ed., *The New Handbook of Multisensory Processes*, 435-452. Cambridge: the MIT Press.

Nath A.R., Fava E.E. & Beauchamp M.S. (2011). Neural Correlates of Interindividual Differences in Children's Audiovisual Speech Perception. *Journal of Neuroscience*, vol. 31(39), 13963-13971.

Nazzi T. & Gonzalez-Gomez N. (2011). Sensitivity to non-adjacent phonological dependencies in 10-month-old infants. *Proceedings of ICPhS XVII*, 1454-1457.

Nazzi T., Bertoncini J. & Bijeljac-Babic R. (2009a). A perceptual equivalent of the labial-coronal effect in the first year of life. *The journal of the acoustic society of America*, vol. 126(3), 1440-1446.

Nazzi T., Floccia C., Moquet B. & Butler J. (2009b). Bias for consonantal over vocalic information in French- and English-learning 30-month-olds: crosslinguistic evidence in early word learning. *Journal of Experimental Child Psychology*, vol. 102, 522-537.

Nazzi T. & New B. (2007). Beyond stop consonants: consonantal specificity in early lexical acquisition. *Cognitive Development*, vol. 22, 271-279.

Nazzi T. (2006). Spécificité phonétique : de la perception précoce à l’acquisition des premiers mots. *Recherches linguistiques de Vincennes*, 35/2006.

Nazzi T., Iakimova G., Bertoncini J., Fredonie S., & Alcantara C. (2006). Early segmentation of fluent speech by infants acquiring French: Emerging evidence for crosslinguistic differences. *Journal of Memory and Language*, vol. 54(3), 283-299.

Nazzi T. (2005). Use of phonetic specificity during the acquisition of new words: differences between consonants and vowels. *Cognition*, vol. 98, 13-30.

Nazzi T., Bertoncini J. & Mehler J. (1998a). Language discrimination by newborns: Toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance*, vol. 24(3), 756-766.

Nazzi T., Floccia C. & Bertoncini J. (1998b). Discrimination of pitch contours by neonates. *Infant behavior and development*, vol. 21(4), 779-784.

Nespor M., Peña M. & Mehler J. (2003). On the different roles of vowels and consonants in speech processing and language acquisition. *Lingue e Linguaggio*, vol. ii, 221-247.

Neuman A.C. & Hochberg I. (1983). Children's perception of speech in reverberation. *The Journal of the Acoustical Society of America*, vol. 73(6), 2145-9.

Newman R.S. (2003). Using links between speech perception and speech production to evaluate different acoustic metrics: A preliminary report. *The Journal of the Acoustical Society of America*, vol. 113(5), 2850-60.

New B., Araújo V. & Nazzi T. (2008). Differential Processing of Consonants and Vowels in Lexical Access through Reading. *Psychological Science*, vol. 19(12), 1223-7.

Newmeyer A.J., Aylward C., Akers R., Ishikawa K., Grether S., deGrauw T., Grasha C., White J. (2009). Results of the Sensory Profile in children with suspected childhood apraxia of speech. *Physical & occupational therapy in pediatrics*, vol. 29(2), 203-218.

Nijland L. (2009). Speech perception in children with output disorders. *Clinical Linguistics and Phonetics*, vol. 23(3), 222-239.

Nijland L., Maassen B., van der Meulen S., Gabreëls F., Kraaimaat F.W. & Schreuder R. (2002). Coarticulation patterns in children with developmental apraxia of speech. *Clinical Linguistics and Phonetics*, vol. 16(6), 461-483.

Nishibayashi L.L. & Nazzi T. (2016). Vowels, then consonants: Early bias switch in recognizing segmented word forms. *Cognition*, vol. 155, 188-203.

Ngong C., Martin A., Dupoux E., Cabrol D., Dutat M. & Peperkamp S. (2013). (Non)words, (non)words, (non)words: evidence for a protolexicon during the first year of life. *Developmental Science*, vol. 16(1), 24-34.

Nguyen N. & Delvaux V. (2016). Role of imitation in the emergence of phonological systems. *Journal of Phonetics*, vol. 53, 46-54.

Nittrouer S. & Studdert-Kennedy M. (1987). The role of coarticulatory effects in the perception of fricatives by children and adults. *Journal of Speech and Hearing Research*, vol. 30, 319– 329.

Nittrouer S., Studdert-Kennedy M., McGowan R.S. (1989). The emergence of phonetic segments: evidence from the spectral structure of fricative-vowel syllables spoken by children and adults. *Journal of Speech and Hearing Research*, vol. 32, 120-132.

Nittrouer S. (1992). Age-related differences in perceptual effects of formant transitions within syllables and across syllable boundaries. *Journal of Phonetics*, 20(3), 351-382.

Nittrouer S. (1993). The Emergence of Mature Gestural Patterns Is Not Uniform. *Journal of Speech, Language and Hearing Research*, vol. 36, 959-972.

Nittrouer S. (1996). Discriminability and Perceptual Weighting of Some Acoustic Cues to Speech Perception by 3-Year-Olds. *Journal of Speech, Language and Hearing Research*, vol. 39, 278-297.

Nittrouer S. & Miller M.E. (1997). Predicting developmental shifts in perceptual weighting schemes. *The Journal of the Acoustical Society of America*, vol. 101(4), 2253-66.

Nottebohm F., Alvarez-Buylla A., Cynx J., Kirn J., Ling C. Y., Nottebohm M., Suter R., Tolles A. & Williams, H. (1990). Song learning in birds: the relation between perception and production. Phil. Trans. R. Soc. Lond. B, 329(1253), 115-124.

- O -

Odell H.K., Shriberg L.D. (2001). Prosody-voice characteristics of children and adults with apraxia of speech. *Clinical Linguistics and Phonetics*, vol. 15(4), 275-307.

Ohde R.N., Haley K.L., Vorperian H.K. & McMahon C.W. (1995). A developmental study of the perception of onset spectra for stop consonants in different vowel environments. *The Journal of the Acoustic Society of America*, vol. 97, 3800-3812.

Ojanen V., Möttönen R., Pekkola J., Jääskeläinen I.P., Joensuu R., Autti T. & Sams M. (2005). Processing of audiovisual speech in Broca's area. *NeuroImage*, vol. 25, 333-338.

Okada K., Rogalsky C., O'grady L., Hanaumi L., Bellugi U., Corina D. & Hickok G. (2016). An fMRI study of perception and action in deaf signers. *Neuropsychologia*, vol. 82, 179-188.

Ozanne A.E. (2005). Childhood Apraxia of Speech. In: Dodd B., a cura di. *Differential Diagnosis and Treatment of Children with Speech Disorders* second edition, 71-82. London: Whurr Pub. Ltd.

- P -

Pascalis O, de Haan M. & Nelson C.A. (2002). Is face processing species-specific during the first year of life? *Science*, vol. 17, 1321-1323.

Pascalis O., Loevenbruck H., Quinn P.C., Kandel S., Tanaka J.W. & Lee K. (2014). On the Links Among Face Processing, Language Processing, and Narrowing During Development. *Child Development Perspectives*, vol. 8(2), 65-70.

Patterson M.L. & Werker J.F. (2003). Two-month-old infants match phonetic information in lips and voice. Developmental Science, vol. 6(2), 191-6.

Patterson M.L. & Werker J.F. (1999). Matching phonetic information in face and voice is robust in 4.5-month-old infants. *Infant Behavior and Development*, vol. 22(2), 237-47.

Pekkola J., Ojanen V., Autti T., Jääskeläinen I. P., Möttönen R., Tarkiainen A., et al. (2005). Primary auditory cortex activation by visual speech: An fMRI study at 3 T. *NeuroReport*, vol. 16, 125–128.

Peña M., Werker J. & Dehaene-Lambertz (2012). Earlier speech exposure does not accelerate speech acquisition. *Journal of Neuroscience*, vol. 32(33), 11159-63.

Peña M., Pittaluga E. & Mehler J. (2010) Language acquisition in premature and full-term infants. *PNAS*, vol. 107(8), 3823-3828.

Peña M., Maki A., Kovačić D., Dehaene-Lambertz G., Koizumi H., Bouquet F. & Mehler J. (2003). Sounds and silence: An optical topography study of language recognition at birth. *PNAS*, vol. 100(20), 11702-11705.

Perkell J.S., Matthies M.L., Tiede M., Lane H., Zandipour M., Marrone N., Stockmann E. & Guenther F.H. (2004a). The distinctness of speakers' /s/-/ʃ/ contrast is related to their auditory discrimination and use of an articulatory saturation effect. *Journal of Speech, Language & Hearing Research*, vol. 47(6), 1259-69.

Perkell J.S., Guenther F.H., Lane H., Matthies M.L., Stockmann E., Tiede M. & Zandipour M. (2004b). The distinctness of speakers' productions of vowel contrasts is related to their discrimination of the contrasts. *The Journal of the Acoustical Society of America*, vol. 116(4), 2338-44.

Phatak S.A. & Allen J.B. (2007). Consonant and vowel confusions in speech-weighted noise). *The Journal of the Acoustical Society of America*, vol. 121(4), 2312-26.

Phatak S.A., Lovitt A. & Allen J.B. (2007). Consonant confusions in white noise. *The Journal of the Acoustical Society of America*, vol. 124(2), 1220-1233. Pickering M.J. & Garrod S. (2013) An integrated theory of language production and comprehension. Behavioral and Brain Sciences, vol. 36, pp. 329-347.

Poeppel D. (2014). The neuroanatomic and neurophysiological infrastructure for speech and language. *Current Opinion in Neurobiology*, vol. 28, 142-149.

Polka L., Masapollo M. & Ménard L. (2014). Who's Talking Now? Infants' Perception of Vowels With Infant Vocal Properties. *Psychological Science*, vol. 25(7), 1448-56.

Polka L., Masapollo M. & Ménard L. (2013). Infants' categorization of vowels with infant vocal properties. In: *Proceedings of Meetings on Acoustics ICA2013* (Vol. 19, No. 1, p. 060053). ASA.

Polka L., Colantonio C. & Sundara M. (2001). A cross-language comparison of /d/-/th/ perception: evidence for a new developmental pattern. *The Journal of the Acoustical Society of America*, vol. 109, 2190-2201.

Polka L., & Werker J.F. (1994). Developmental changes in perception of non-native vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance*, vol. 20(2), 421-35.

Poltrock S. & Nazzi T. (2015). Consonant/vowel asymmetry in early word form recognition. *Journal of Experimental Child Psychology*, vol. 131, 135–148.

Pons F., Bosch L. & Lewkowicz D.J. (2015). Bilingualism Modulates Infants' Selective Attention to the Mouth of a Talking Face. *Psychological Science*, vol. 26(4), 490-498.

Pons F. & Lewkowicz D.J. (2014). Infant perception of audio-visual speech synchrony in familiar and unfamiliar fluent speech. *Acta Psychologica*, vol. 149, 142-147.

Pons F., Lewkowicz D.J., Soto-Faraco S., Sebastián-Gallés N. (2009). Narrowing of intersensory speech perception in infancy. *PNAS*, vol. 106(26), 10598-10602.

Prather J., Okanoya K. & Bolhuis J.J. (2017). Brains for birds and babies: Neural parallels between birdsong and speech acquisition. *Neuroscience & Biobehavioral Reviews*, 2017.

Preston J.L., Molfese P.J., Gumkowski N., Sorcinelli A., Harwood V., Irwin J.R. & Landi N. (2014). Neurophysiology of speech differences in childhood apraxia of speech. *Developmental neuropsychology*, vol. 39(5), 385-403.

Pulvermüller F. & Fadiga L. (2010). Active perception: sensorimotor circuits as a cortical basis for language. *Nature Reviews. Neuroscience*, vol. 11(5), 351-60.

Pulvermüller F., Huss M., Kherif F., Moscoso del Prado Martin F., Hauk O. & Shtyrov Y. (2006). Motor cortex maps articulatory features of speech sounds. *PNAS*, vol. 103, 7865–7870.

Pulvermüller F. (2005). Brain mechanisms linking language and action. *Nature Reviews. Neuroscience*, vol. 6(7), 576-582.

Putzar L., Hötting K. & Röder B. (2010). Early visual deprivation affects the development of face recognition and of audio-visual speech perception. *Restorative Neurology and Neuroscience*, vol. 28(2), 251-257.

- Q -

Quené H. & van der Bergh (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language*, vol. 59(4), 413-425.

- R -

Ramus F., Hauser M.D., Miller C., Morris D. & Mehler J. (2000). Language Discrimination by Human Newborns and by Cotton-Top Tamarin Monkeys. *Science*, vol. 288(5464), 349-351.

Ramus F., Nespor M., Mehler J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, vol. 73(3), 265-292.

Randazzo M. (2016). Audiovisual Integration in Apraxia of Speech: EEG Evidence for Processing Differences. *PHd Dissertation*, University of British Columbia. Available at: http://dx.doi.org/10.7916/D85D8S5H.

Rauschecker A.M., Pringle A., Watkins K.E. (2007). Changes in neural activity associated with learning to articulate novel auditory pseudo-words by covert repetition. *Human Brain Mapping*, vol. 29, 1231-1242.

Reisberg D., Mclean J. & Goldfield A. (1987). Easy to Hear but Hard to Understand: A Lip-Reading Advantage with Intact Auditory Stimuli. In: Dodd B. & Campbell R., eds., *Hearing by Eye: the Psychology of Lip-reading*. Hillsdale, NJ: Lawrence Erlbaum Associates publications, 97–113.

Rigler H., Farris-Trimble A., Greiner L., Walker J., Tomblin J. B., & McMurray B. (2015). The slow developmental time course of real-time spoken word recognition. *Developmental Psychology*, 51(12), 1690-1703.

Rivera-Gaxiola M., Silva-Pereira J. & Kuhl P.K. (2005). Brain potentials to native and non-native speech contrasts in 7- and 11-month-old American infants. *Developmental Science*, vol. 8(2), 162-172.

Rogalsky C., Love T., Driscoll D., Anderson S.W. & Hickok G. (2011). Are mirror neurons the basis of speech perception? Evidence from five cases with damage to the purported human mirror system. *Neurocase*, vol. 17(2), 178-187.

Rogers J.C., Möttönen R., Boyles R. & Watkins K. (2014). Discrimination of speech and non-speech sounds following theta-burst stimulation of the motor cortex. *Frontiers in Psychology*, vol. 5, https://doi.org/10.3389/fpsyg.2014.00754.

Rosenblum L.D. (2005). Primacy of Multimodal Speech Perception. In: Pisoni D.B. & Remez R.E., eds., *The Handbook of Speech Perception*. Malden (USA): Blackwell Publishing Ltd, 51-78.

Rosenblum L.D. & Saldaña H.M. (1996). An audiovisual test of kinematic primitives for visual speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, vol. 22(2), 318-331.

Rosenblum L.D., Schmuckler M.A. & Johnson J.A. (1997). The McGurk effect in infants. *Perception & Psychophysics*, vol. 59(3), 347-357.

Ross L.A., Molholm S., Blanco D., Gomez-Ramirez M., Saint-Amour D. & Foxe J.J. (2011). The development of multisensory speech perception continues into the late childhood years. *European Journal of Neuroscience*, vol. 33(12), 2329-2337.

Ross L.A., Del Bene V.A., Molholm S., Frey H-P. & Foxe J.J. (2015). Sex differences in multisensory speech processing in both typically developing children and those on the autism spectrum. *Frontiers in Neuroscience*, vol. 9, doi: 10.3389/fnins.2015.00185.

Roy A.C., Craighero L., Fabbri-Destro M., Fadiga L. (2008). Phonological and lexical motor facilitation during speech listening: A transcranial magnetic stimulation study. *Journal of Psychology*, vol. 102, 1-3.

Rvachew S. & Jamieson D.G. (1989). Perception of voiceless fricatives by children with a functional articulation disorder. *Journal of Speech and Hearing Disorders*, 54(2), 193-208.

- S -

Saffran J.R., Aslin R.N. & Newport, E.L. (1996). Statistical learning by 8-month-old infants. *Science*, vol. 274, 1926–1928.

Sai F.Z. (2005). The role of the mother's voice in developing mother's face preference: Evidence for intermodal perception at birth. *Infant and Child Development*, vol. 14(1), 29-50.

Saksida A., Langus A. & Nespor M. (2017). Co-occurrence statistics as a language dependent cue for speech segmentation. *Developmental Science*, doi: 10.1111/desc.12390.

Sansavini A., Bertoncini J., & Giovanelli G. (1997). Newborns discriminate the rhythm of multisyllabic stressed words. *Developmental Psychology*, vol. 33(1), 3-11.

Sams M., Aulanko R., Hämäläinen M., Hari R., Lounasmaa O. V., Lu S.T. & Simola J. (1991). Seeing speech: visual information from lip movements modifies activity in the human auditory cortex. *Neuroscience letters*, vol. 127(1), 141-145.

Schwartz J-L., Basirat A., Ménard L. & Sato M. (2012). The Perception-for-Action-Control Theory (PACT): A perceptuo-motor theory of speech perception. *Journal of Neurolinguistics*, vol. 25(5), 336-354.

Schwartz J-L., Moulin-Frier C. & Oudeyer P-Y. (2015). On the cognitive nature of speech sound systems. *Journal of Phonetics*, vol. 53, 1-4.

Schwartze M. & Koltz S. (2015). Contributions of cerebellar event-based temporal processing and preparatory function to speech perception. *Brain & Language*, vol. 161, 28-32.

Scotto M., Yeung H.H., Gick B. & Werker J.F. (2013). Inner speech captures the perception of external speech. *The Journal of the Acoustic Society of America*, vol. 133(4), EL286-EL292.

Sebastián-Gallés N., Albareda-Castellot B., Weikum W. & Werker J.F. (2012). A bilingual advantage in visual language discrimination in infancy. *Psychological Science*, vol. 23(9), 994-9.

Sekiyama K. & Burnham D. (2008). Impact of language on development of auditory-visual speech perception. *Developmental Science*, vol. 11(2), 306-320.

Sekiyama K. & Tohkura Y. I. (1993). Inter-language differences in the influence of visual cues in speech perception. *Journal of Phonetics*, vol. 21(4), 427-444.

Sekiyama K. & Tohkura Y. I. (1991). McGurk effect in non-English listeners: Few visual effects for Japanese subjects hearing Japanese syllables of high auditory intelligibility. *The Journal of the Acoustical Society of America*, vol. 90(4), 1797-1805.

Shaw K.E. & Bortfeld H. (2015). Sources of Confusion in Infant Audiovisual Speech Perception Research. *Frontiers in Psychology*, vol. 6, doi: 10.3389/fpsyg.2015.01844.

Shaw K.E., Depowski N. & Bortfeld H. (2015). Infants' preference for native audiovisual speech dissociated from congruency preference. *PLOSOne*, vol. 10(4), doi: 10.1371/journal.pone.0126059.

Shi R., Werker J.F. & Morgan J.L. (1999). Newborn infants’ sensitivity to perceptual cues to lexical and grammatical words. *Cognition*, vol. 72(2), B11-21.

Shinozaki J., Hiroe N., Sato M-a., Nagamine T. & Sekiyama K. (2016). Impact of language on functional connectivity for audiovisual speech integration. *Scientific Reports*, doi: 10.1038/srep31388.

Shriberg L.D., Lohmeier H.L., Strand E.A., Jakielski K.J. (2012). Encoding, memory, and transcoding deficits in Childhood Apraxia of Speech. *Clinical Linguistics and Phonetics*, vol. 26, 445-482.

Shriberg L.D. (2010). A neurodevelopmental framework for research in childhood apraxia of speech. In: Maassen B., van Lieshout P. (2010), eds., *Speech Motor Control*. Oxford University Press, 259-270.

Shriberg L.D., Campbell T.F., Karlsson H.B., Brown R.L. Mcsweeny J.L. and Nadler C.J. (2003a). A diagnostic marker for childhood apraxia of speech: the lexical stress ratio. *Clinical Linguistics and Phonetics*, vol. 17(7), 549-574.

Shriberg L.D., Green J.R., Campebell T.F, Mcsweeny J.L. and Scheer A.R. (2003b). A diagnostic marker for childhood apraxia of speech: the coefficient of variation ratio. *Clinical Linguistics and Phonetics*, vol. 17(7), 575-595.

Shriberg L. D., Aram D. M. & Kwiatkowski J. (1997b). Developmental apraxia of speech: II. Toward a diagnostic marker. *Journal of Speech, Language, and Hearing Research*, vol 40(2), 286-312.

Shriberg, L. D. Aram, D. M. & Kwiatkowski J. (1997c). Developmental apraxia of speech: III. A subtype marked by inappropriate stress. *Journal of Speech, Language, and Hearing Research*, vol. 40(2), 313-337.

Shukla M., Nespor M. & Mehler J. (2007). An interaction between prosody and statistics in the segmentation of fluent speech. *Cognitive Psychology*, vol. 54(1), 1-32.

Simon C. & Fourcin A.J. (1978). Cross-language study of speech-pattern learning. *Journal of the Acoustical Society of America*, vol. 63, 925-35.

Skipper J.I., Nusbaum H.C. & Small S.L. (2005). Listening to talking faces: motor cortical activation during speech perception. *NeuroImage*, vol. 25(1), 76-89.

Slawinski E.B. & Fitzgerald L. (1998). Perceptual development of the categorization of the /r-w/ contrast in normal children. Journal of Phonetics, vol. 26(1), 27-43.

Smith A. (2006). Speech motor development: Integrating muscles, movements, and linguistic units. *Journal of Communication Disorders*, vol. 39 (5), 331-349.

Smith A. & Zelaznik H.N. (2004). Development of functional synergies for speech motor coordination in childhood and adolescence. *Developmental Psychology*, vol. 45(1), 22-33.

Smith A. & Goffman L. (1998). Stability and patterning of speech movement sequences in children and adults. *Journal of Speech, Language, and Hearing Research*, vol. 41(1), 18-30.

Soto-Faraco S., Calabresi M., Navarra J., Werker J. & Lewkowicz D.J. (2012). The development of audiovisual speech perception. In: Bremner A.J., Lewkowicz D.J. & Spence C., eds., *Multisensory Development*. Oxford University Press, pp 207-228.

Souza T.N., Payão L.M., Costa R.C. (2009). Childhood speech apraxia in focus: theoretical perspectives and present tendencies. *Pró-Fono Revista de Atualizaçao Cientifica*, vol. 21, 75-80.

Sporns O. & Edelman G.M. (1993). Solving Bernstein's Problem: A Proposal for the Development of Coordinated Movement by Selection. *Child Development*, vol. 64(4), 960-981.

Stackhouse J. (1992). Developmental verbal dyspraxia: A review and critique. *European Journal of Disorders of Communication*, vol. 27, 19-34.

Stager C.L & Werker J.F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature*, vol. 388(6640), 381-382.

Strauß A. & Schwartz J-L. (2017). The syllable in the light of motor skills and neural oscillations. *Language, Cognition and Neuroscience*, vol. 32(5), 562-569.

Stollman M.H.P., van Velzen E.C.W., Simkens H.M.F., Snik A.F.M. & van der Broek P. (2009). Development of auditory processing in 6–12-year-old children: a longitudinal study. *International Journal of Audiology*, vol. 43(1), 33-44.

Sussman H. M., Marquardt T. P. & Doyle J. (2000). An acoustic analysis of phonemic integrity and contrastiveness in developmental apraxia of speech. *Journal of Medical Speech-Language Pathology*, vol. 8(4), 301–313.

Streri A., Coulon M., Marie J. & Yeung H.H. (2016). Developmental change in infants' detection of visual faces that match auditory vowels. *Infancy*, vol. 21(2), 177-198.

Sumby W.H. & Pollack I. (1954). The visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America*, vol. 26(2), p. 212-216.

Swingley D. (2017). The infant's developmental path in phonological acquisition. *British Journal of Psychology*, vol. 108(1), 28-30.

Swingley D. (2016). Two-Year-Olds Interpret Novel Phonological Neighbors as Familiar Words. *Developmental Psychology*, vol. 52(7), 1011-1023.

Swingley D. & Aslin R.N. (2007). Lexical competition in young children's word learning. *Cognitive Psychology*, vol. 54(2), 99-132.

Swingley, D. (2005). 11-month-olds' knowledge of how familiar words sound. *Developmental Science*, vol. 8(5), 432-443.

Swingley D. (2003). Phonetic detail in the developing lexicon. *Language and Speech*, vol. 46, 265-294.

Swingley D. & Aslin R.N. (2002). Lexical neighborhoods and the word-form representations of 14-month-olds. *Psychological Science*, vol. 13, 480-484.

Swingley D. & Aslin R.N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition*, vol. 76, 147-166.

Swoboda P., Morse P.A. & Leavitt L.A. (1976). Continuous vowel discrimination in normal and at-risk infants. *Child Development*. vol. 47, 459-65.

- T -

Taitelbaum-Swead R. & Fostick L. (2016). Auditory and visual information in speech perception: A developmental perspective. *Clinical Linguistics and Phonetics*, vol. 30(7), 531-545.

Tkach J.A., Chen X., Freebairn L.A., Schmithorst V.J., Holland S.K. & Lewis B.A. (2011). Neural Correlates of Phonological Processing in Speech Sound Disorder: A Functional Magnetic Resonance Imaging Study. *Brain & Language*, vol. 119(1), 42-49.

Teinonen T., Aslin R.N., Alku P. & Csibra G. (2008). Visual speech contributes to phonetic learning in 6-month-old infants. *Cognition*, vol. 108, 850-855.

Tenenbaum E.J., Sobel D.M., Sheinkopf S.J., Malle B.F. & Morgan J.L. (2015). Attention to the mouth and gaze following in infancy predict language development. *Journal of Child Language*, vol. 42(6), 1173-1190.

Terband H. & Maassen B. (2010a). Speech motor development in childhood apraxia of speech: generating testable hypotheses by neurocomputational modeling. *Folia phoniatrica et logopaedica: official organ of the International Association of Logopedics and Phoniatrics*, vol. 63 (3), 134-142.

Terband H. & Maassen B. (2010b). Testing hypothesis about the neurological mechanisms underlying childhood apraxia of speech. *Oral presentation at the International Conference on Motor Speech, Savannah, GA, March 2010.*

Teverovsky E.G., Bickel J.O., Feldman H.M. (2009). Functional characteristics of children diagnosed with Childhood Apraxia of Speech. *Disability and rehabilitation*, vol. 31, 94-102.

Ter Schure S., Junge C. & Boersma P. (2016). Discriminating Non-native Vowels on the Basis of Multimodal, Auditory or Visual Information: Effects on Infants' Looking Patterns and Discrimination. *Frontiers in Psychology*, vol. 7, doi: 10.3389/fpsyg.2016.00525.

Thierry G., Vihman M. & Roberts M. (2003). Familiar words capture the attention of 11-month-olds in less than 250ms. *Learning and Memory*, vol. 14(18), 2307-10.

Thiessen E.D. & Saffran J.R. (2003). When Cues Collide: Use of Stress and Statistical Cues to Word Boundaries by 7- to 9-Month-Old Infants. *Developmental Psychology*, vol. 39(4), 706-16.

Thompson E.C., Woodruff Carr K., White-Schwoch T., Otto-Meyer S. & Kraus N. (2017). Individual differences in speech-in-noise perception parallel neural speech processing and attention in preschoolers. *Hearing Research*. [Epub ahead of print], pii: S0378-5955(16)30233-7. doi: 10.1016/j.heares.2016.11.007.

Thoonen G., Maassen B., Gabreëls F. & Schreuder R. (1999). Validity of maximum performance tasks to diagnose motor speech disorders in children. *Clinical Linguistics and Phonetics*, vol. 13, 1-23.

Thoonen G., Maassen B., Gabreëls F. & Schreuder R. (1994). Feature analysis of singleton consonant errors in developmental verbal dyspraxia (DVD). *Journal of Speech, Language and Hearing Research*, vol. 37, 1424-1440.

Tiippana K., Möttönen R. & Schwartz J-L. (2015). Multisensory and sensorimotor interactions in speech perception. *Frontiers in Psychology*, doi: 10.3389/fpsyg.2015.00458.

Tomalski P. (2015). Developmental Trajectory of Audiovisual Speech Integration in Early Infancy. A Review of Studies Using the McGurk Paradigm. *Psychology of Language and Communication*, vol. 19(2), doi: 10.1515/plc-2015-0006.

Tomalski P., Ribeiro H., Ballieux H., Axelsson E.L., Murphy E., Moore D.G. & Kushnerenko E. (2013). Exploring early developmental changes in face scanning patterns during the perception of audiovisual mismatch of speech cues. *European Journal of Developmental Psychology*, vol. 10(5), 611-624.

Treiman R. & Zukowski A. Levels of phonological awareness. In: Brady S.A. & Shankweiler D.P. (2001). *Phonological Processes in Literacy*. NY: Lawrence Erlbaum Associates Inc., 67-83.

Tsang T., Atagi N. & Johnson S.P. (2018). Selective attention to the mouth is associated with expressive language skills in monolingual and bilingual infants. *Journal of Experimental Child Psychology*, vol. 169, 93-109.

Tsao F.M., Liu H.M. & Kuhl P.K. (2006). Perception of native and non-native affricate-fricative contrasts: cross-language tests on adults and infants. *The Journal of the Acoustic Society of America*, vol. 120(4), 2285-2294.

Tsao F.M., Liu H.M. & Kuhl P.K. (2004). Speech perception in infancy predicts language development in the second year of life: a longitudinal study. *Child Development*, vol. 75(4), 1067-1084.

Tubul-Lavy G. (2012). Intra-word inconsistency in apraxic Hebrew-speaking children. *Clinical linguistics and phonetics*, vol. 26, 502-517.

Turner A.C., McIntosh D.N. & Moody E.J. (2015). Don't Listen With Your Mouth Full: The Role of Facial Motor Action in Visual Speech Perception. *Language and Speech*, vol. 58(2), 267-278.

- V -

van Wassenhove V., Grant K.W. & Poeppel D. (2005). Visual speech speeds up the neural processing of auditory speech. *PNAS*, vol. 102(4), 1181-1186.

Vance M., Rosen S. & Coleman M. (2009). Assessing speech perception in young children and relationships with language skills. *International Journal of Audiology*, vol. 48, 708–717.

Velleman S.L. (2011). Lexical and phonological development in children with childhood apraxia of speech. A commentary on Stoel-Gammon's ‘Relationships between lexical and phonological development in young children’. *Journal of Child Language*, vol. 38(1), 82-86.

Venezia J.H., Fillmore P., Matchin W., Isenberg A.L., Hickok G. & Fridriksson J. (2016). Perception drives production across sensory modalities: A network for sensorimotor integration of visual speech. *Neuroimage*, vol. 126, 196-207.

Ventura P., Kolinsky R., Fernandesa S., Queridoa L. & Morais J. (2007). Lexical restructuring in the absence of literacy. *Cognition*, vol. 105(2), 334-61.

Vihman M.M. (2017). Learning words and learning sounds: Advances in language development. *British Journal of Psychology*, vol. 108(1), 1-27.

Vihman M.M. (2015). Perception and production in phonological development. In: MacWhinney B. & O'Grady W., eds. *Handbook of Language Emergence*. Malden, MA: Wiley-Blackwell, pp. 437-457.

Vihman M.M., De Paolis R.A. & Keren-Portnoy T. (2014). The Role of Production in Infant Word Learning. *Language Learning*, vol. 64, suppl. 2, 121–140.

Vihman M.M. & Keren-Portnoy T., eds. (2013). *The Emergence of Phonology: Whole word approaches, cross-linguistic evidence*. Cambridge: Cambridge University Press.

Vihman M.M. & Keren-Portnoy T. (2011). The role of production practice in lexical and phonological development. *Journal of Child Language*, vol.38, 41-45.

Vihman M. M. & Vihman V-A. (2011). From first words to segments: A case study in phonological development. In: Arnon I. & Clark E.V., eds. *Experience, Variation, and Generalization: Learning a first language.* (Trends in Language Acquisition Research 7., pp. 109-133. Amsterdam: John Benjamins.

Vihman M.M., Nakai S., De Paolis R.A. & Hallé P. (2004). The role of accentual pattern in early lexical representation. *Journal of Memory and Language*, vol. 50, 336-353.

Villacorta V.M., Perkell J.S. & Guenther F.H. (2007). Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception. *The Journal of the Acoustical Society of America*, vol. 122(4), 2306-19.

Volpe G., D'Ausilio A., Badino L., Camurri A. & Fadiga L. (2016). Measuring social interaction in music ensembles. *Philosophical Transactions of the Royal Society B - biological sciences*, vol. 371, p. 1693.

Vorperian H. K., Wang S., Chung M. K., Schimek E. M., Durtschi R. B., Kent R. D., Ziegert A. J. & Gentry L. R. (2009). Anatomic development of the oral and pharyngeal portions of the vocal tract: an imaging study. *The Journal of the Acoustic Society of America*, vol. 125(3), pp. 1666-1678.

Vorperian H.K. (2005). Development of vocal tract length during early childhood: A magnetic resonance imaging study. *The Journal of the Acoustical Society of America*, vol. 117(1), 338-50.

Vouloumanos A., Hauser M.D., Werker J.F. & Martin A. (2010). The tuning of human neonates' preference for speech. *Child Development*, vol. 81(2), 517-27.

Vouloumanos A., & Werker J.F. (2007). Listening to language at birth: Evidence for a bias for speech in neonates. *Developmental Science*, vol. 10(2), 159-64.

Vouloumanos A. & Werker J.F. (2004). Tuned to the signal: The privileged status of speech for young infants. *Developmental Science*, vol. 7(3), 270-6.

Vroomen J & Baart M. (2012). Phonetic Recalibration in Audiovisual Speech. In: Murray M.M., Wallace M.T., eds. *The Neural Bases of Multisensory Processes*. Boca Raton (FL): CRC Press/Taylor & Francis, Chapter 19 (Frontiers in Neuroscience).

- W -

Walker J.F. & Archibald L.M.D. (2005). Articulation rate in preschool children: a 3-year longitudinal study. *International Journal of Language & Communication Disorders*, vol. 41(5), 541-565.

Walker P.M. & Tanaka J.W. (2003). An encoding advantage for own-race versus other-race faces. *Perception*, vol. 32(9), 1117-25.

Walley A.C., Metsala J.L. & Garlock M. (2003). Spoken vocabulary growth: Its role in the development of phoneme awareness and early reading ability. *Reading and Writing*, vol. 16(1-2), 5-20.

Walsh B. & Smith A. (2002). Articulatory Movements in Adolescents – Evidence for Protracted Development of Speech Motor Control Processes. *Journal of Speech, Language and Hearing Research*, vol. 45, 1119-33.

Walton G.E. & Bower T.G.R. (1993). Amodal representation of speech in infants. *Infant Behavior and Development*, vol. 16(2), 233-243.

Weatherhead D. & White K.S. (2017). Read my lips: Visual speech influences word processing in infants. *Cognition*, vol. 160, 103-109.

Webb A.R., Heller H.T., Benson C.B. & Lahav A. (2015). Mother's voice and heartbeat sounds elicit auditory plasticity in the human brain before full gestation. *PNAS*, vol. 112(10), 3152-3157.

Weikum W.M., Vouloumanos A., Navarra J., Soto-Faraco S., Sebastián-Gallés N. & Werker J.F. (2013). Age-related sensitive periods influence visual language discrimination in adults. *Frontiers in Systems Neuroscience*, vol. 7, p. 86.

Weikum W., Vouloumanos A., Navarra J., Soto-Faraco S., Sebastián-Gallés N. & Werker J.F. (2007). Visual language discrimination in infancy. *Science*, vol. 316(5828), 1159.

Werker J.F. & Hensch T.K. (2015). Critical Periods in Speech Perception: New Directions. Annual Review of Psychology, vol. 66, 173-96.

Werker J.F. (2012). Perceptual Foundations of Bilingual Acquisition in Infancy. *Annals of the New York Academy of Sciences: The Year in Cognitive Neuroscience*, vol. 1251, 50-61.

Werker J.F., Yeung H.H. & Yoshida K.A. (2012). How do infants become experts at native-speech perception? *Current Directions in Psychological Science*, vol. 21(4), 221-226.

Werker J.F., Pons F., Dietrich C., Kajikawa S., Fais L. & Amano S. (2007). Infant-directed speech supports phonetic category learning in English and Japanese. *Cognition*, vol. 103(1), 147-62.

Werker J.F. & Curtin S. (2005). PRIMIR: A developmental framework of infant speech processing. *Language Learning and Development*, vol. 1(2), 197-234.

Werker J.F. & Tees R.C. (2005). Speech perception as a window for understanding plasticity and commitment in language systems of the brain. *Developmental Psychobiology*, vol. 46(3), 233-51.

Werker J.F. & Yeung H.H. (2005). Infant speech perception bootstraps word learning. *Trends in Cognitive Sciences*, vol. 9(11), 519-527.

Werker J.F., Fennell C.T., Corcoran K.M. & Stager C.L. (2002). Infants' ability to learn phonetically similar words: Effects of age and vocabulary size. *Infancy*, vol. 3, 1-30.

Werker J.F., Frost P. & McGurk H.E. (1992). La langue et les lèvres: Cross-language influences on bimodal speech perception. *Canadian Journal of Psychology*, vol. 46(4), 551-68.

Werker J.F. (1989). Becoming a native listener: A developmental perspective on human speech perception. *American Scientist*, vol. 77(1), 54-59.

Werker J.F. & Lalonde C.E. (1988). Cross-language speech perception: Initial capabilities and developmental change. *Developmental Psychology*, vol. 24(5), 672-683.

Werker J.F. & Tees R.C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, vol. 7(1), 49-63.

Werker J.F., Gilbert J.H.V., Humphrey K. & Tees R. (1981). Developmental aspects of Cross-Language Speech Perception. *Child Development*, vol. 52(1), 349-355.

White K.S. & Morgan J.L. (2008). Sub-segmental detail in early lexical representations. *Journal of Memory and Language*, vol. 59, 114-132.

White K.S., Yee E., Blumstein S.E. & Morgan, J.L. (2013). Adults show less sensitivity to phonetic detail in unfamiliar words, too. *Journal of Memory and Language*, 68, 362-378.

Whitehill T.L., Francis A.L. & Ching C.K. (2003). Perception of place of articulation by children with cleft palate and posterior placement. *Journal of Speech, Language and Hearing Research*, vol. 46(2), 451–61.

Williams P. & Stackhouse J. (2000). Rate, accuracy and consistency: diadochokinetic performance of young, normally developing children. *Clinical Linguistics and Phonetics*, vol. 14, 267-293.

Williams P. & Stackhouse J. (1998). Diadochokinetic Skills: Normal and Atypical Performance in Children Aged 3–5 Years. *International Journal of Language & Communication Disorders*, vol. 33 (suppl.), 481-486.

Wilson S.M. & Iacoboni M. (2006). Neural responses to non-native phonemes varying in producibility: Evidence for the sensorimotor nature of speech perception. *Neuroimage*, 33(1), 316-325.

Wilson S.M., Saygin A.P., Sereno I.M., Iacoboni M. (2004). Listening to speech activates motor areas involved in speech production. *Nature Neuroscience*, vol. 7(7), 701-702.

Woods D.L., Yund E.W., Herron T.J. & Cruadhlaoich M.A.I. (2009). Consonant identification in consonant-vowel-consonant syllables in speech-spectrum noise. *The Journal of the Acoustic Society of America*, vol. 127(3), doi: 10.1121/1.3293005.

- Y -

Yeung H.H., Scott M. & Granjon L. (2017). Breathing position and the perception of consonants. *Journal of the Canadian Acoustical Association*, vol. 44(3).

Yeung H.H. & Werker J.F. (2013). Lip movements affect infant audiovisual speech perception. *Psychological Science*, vol. 24(5), 603-612.

Yeung H.H., Chen K.H. & Werker J.F. (2013). When does native language input reorganize phonetic perception? The precocious case of lexical tone. *Journal of Memory and Language*, vol. 68(2), 123-139.

Yeung H.H. & Werker J.F. (2009). Learning words' sounds before learning how words sound: 9-month-old infants use distinct objects as cues to categorize speech information. *Cognition*, vol. 113(2), 234-243.

Yi H.G., Phepls J.E.B. & Chandrasekaran B. (2013). Reduced efficiency of audiovisual integration for nonnative speech. *The Journal of the Acoustic Society of America*, vol. 134(5), EL387-93.

Yoshida K. A., Pons F., Maye J. & Werker J.F. (2010a). Distributional phonetic learning at 10 months of age. *Infancy*, vol. 15(4), 420-33.

Yoshida K. A., Iversen J.R., Patel A.D., Mazuka R., Nito H., Gervain J. & Werker J.F. (2010b). The development of perceptual grouping biases in infancy: A Japanese-English cross-linguistic study. *Cognition*, vol. 115(2), 356-61.

Yoshida K., Fennell C., Swingley D. & Werker J.F. (2009). 14-month-olds learn similar-sounding words. *Developmental Science*, vol. 12, 412-418.

- Z -

Zamuner T.S., Yeung H.H. & Ducos M. (2017). The many facets of speech production and its complex effects on phonological processing. *British Journal of Psychology*, vol. 108, 37-39.

Zaretsky E., Velleman S.L., Curro K. (2010). Through the magnifying glass: Underlying literacy deficits and remediation potential in childhood apraxia of speech. *International journal of speech-language pathology*, vol. 12(1), 58-68.

Zekveld A.A., Kramer S.E., Kessens J.M., Vlaming M.S. & Houtgast T. (2009). The influence of age, hearing, and working memory on the speech comprehension benefit derived from an automatic speech recognition system. *Ear and hearing*, vol. 30(2), 262-272.

Zesiger P. & Jöhr J. (2011). Les representations phonologiques des mots chez le jeune enfant. *NecPlus enfance*, vol. 3, 293-309.

Zevin J.D. (2012). A sensitive period for shibboleths: The long tail and changing goals of speech perception over the course of development. *Developmental Psychology*, vol. 54(6), 632-42.

Zhao T.C. & Kuhl (2017). Effects of enriched auditory experience on infants' speech perception during the first year of life. *PROSPECTS*, vol. 46(2), 235-247.

Zimmer E.Z., Fifer W.P., Kim Y., Rey H.R., Chao C.R., Myers M.M. (1993). Response of the premature fetus to stimulation by speech sounds. *Hearly Human Development*, vol. 33, 207-215.

Zlatin M.A. & Koenigsknecht R. A. (1975). Development of the voicing contrast: Perception of stop consonants. *Journal of Speech and Hearing Research*, 18, 541–553.

Zobouyan C., Serniclaes W. & Bertoncini J. (2012). Troubles de la perception des phonèmes chez l’enfant dysphasique : déficit de perception catégorielle et perception allophonique. *ANAE - Approche Neuropsychologique des Apprentissages chez l’Enfant*, vol. 118, 367-374.

Zuk J., Iuzzini-Seigel J., Cabbage K., Green J.R. & Hogan T. (2017). Poor Speech Perception Is Not a Core Deficit of Childhood Apraxia of Speech: Preliminary Findings. *Journal of Speech, Language and Hearing Research*, doi:10.1044/2017_JSLHR-S-16-0106.