COARTICULATION AND VOT IN FOUR ITALIAN CHILDREN FROM 18 TO 48 MONTHS OF AGE

Zmarich C. [1,2], Bonifacio S., Busà M. G. [2], Colavolpe B. [2], Gaiotto M. [2], Olivucci F.[3]
¹CNR-ISTC, Padova (I), [2]Università di Padova (I), [3]Università di Bologna (I)
claudio.zmarich@cnr.it

STATE OF THE ART. Over the past decades linguistic theories and theories of speech motor control have proceeded separately. This has been due to both theoretical and methodological reasons. From the theoretical point of view, there has been a pervasive influence of Generative Grammar, claiming a separation between Language and Performance, where Language systems are often regarded as developing independently of other cognitive and senso-motor systems, see Fodor, 1983). Methodologically, there has been a difficulty to trace neurobiological markers of the interaction between language and senso-motor systems, accompanied by an apparent lack of one-to-one correspondence between linguistic units and measures of executive behavior, see Smith (2010); Laganaro, (2018). Traditional explanations have emphasized the acquisition of phonology, lexicon and morphosyntax, but NOT of the motor processes related to speech production. The interconnection between linguistic factors and motor factors in the phonetic development is made complex by the continuous changes of the anatomo-physiological structures (for morphology, size and muscle innervation), of the neural substrate and of cognition (Callan et al., 2000). Yet the neural organization for sensorimotor and cognitive-linguistic aspects is highly interactive: for instance, the behavioral evidence shows a high degree of co-occurrence between cognitive-linguistic deficits and motor deficits (Goffman, 2015; McAllister-Byun, Tessier; 2016), and theoretical proposals linking together language planning, speech production and perception, and neurophysiological organization are currently available, covering also the developmental perspective (*DIVA* model, Guenther, 1995; Guenther & Vladusich, 2012; *State Feedback Model*, Parrell & Houde, 2019).

In order to study the acquisition of motor control in the early stages of language development, the empirical analyses preferably rely on acoustic data, as acoustic analysis has the advantage to quantify the phonetic continuum in the time-frequency domain, deriving information - by inference - on the underlying movements. More importantly, acoustic analysis is preferred because it is non-invasive, cheap and relatively simple to perform. Acoustic analysis is also useful for the study of coarticulation, which refers to the temporal overlapping of gestures belonging to neighboring phones (Hardcastle, Hewlitt, 2006; Farnetani, Recasens, 2010).

The present investigation aims to contribute some experimental data to two topics that have been understudied in Italian child language development, namely the acquisition of VOT (see Bortolini et al., 1995; Zmarich et al., 2009) and of anticipatory CV coarticulation (Petracco, Zmarich, 2006; Zmarich et al., 2009), though they have been thoroughly studied in other languages such as English, French or German (see Macken, Barton, 1980a; Eilers et al., 1984; Allen, 1985 for VOT; Sussman et al., 1999, for anticipatory coarticulation).

The best parameter for quantifying and classifying the voicing contrast is Voice Onset Time or VOT, which measures the time elapsed since the release of the consonantal occlusion to the beginning of the vibration of the vocal folds. Early in phonetic development, the voiced and unvoiced consonants tend to be realized as voiceless unaspirated, which allows the synchronization between glottal and supraglottal events; only after the acquisition of articulatory maneuvers children come to achieve the VOT categories that characterize the native language. In different languages the phonemic contrast between sonority categories corresponds to distinct temporal intervals along the VOT continuum (Abramson, Whalen, 2017; Cho, Whalen & Docherty, 2019).

While progressive coarticulation has been explained mainly as a result of articulatory inertia, anticipatory coarticulation has a cognitive base, because it must be planned in advance. According to the most influential hypothesis (Sussman et al., 2009), in the development of anticipatory coarticulation in a CV syllable, the child progressively narrows the domain of the articulatory organization from the syllable to the individual C and V gestures, with the consequence of decreasing coarticulation and increasing phonemic distinctiveness, but the process is not linear and depends on physiological constraints on the articulators. In fact, anticipatory coarticulation varies according to the articulatory place of the consonant. In the case of bilabial consonants, the articulation of the lips for the production of the consonant in CV syllable is not affected by the tongue dorsum during the production of the following vowel; this allows for maximum temporal overlap of the articulators (coarticulation as co-production). As for dental\alveolar consonants, the child must learn to differentiate and coordinate the apex (for the consonant) and the dorsum of the tongue (for the vowel), which are largely independent. As for velar consonants, the biomechanical constraints are maximum (both C and V are articulated with the tongue dorsum), and in this case the child must learn to mutually adapt the articulatory places for C and V (coarticulation as mutual adaptation).

AIMS OF THE PRESENT WORK. On the basis of both literature' results and our previous investigations, our working hypotheses are the following. VOT: Since the production of the initial voiced stop consonants require some mechanism external to the larynx in order to sustain an adequate transglottal pressure drop during the stop closure (as an active lowering of the glottis, Westbury, 1983), children will be more advanced in the acquisition of the appropriate VOT values for the voiceless than for the voiced consonants. Developmental studies on the acquisition of voicing in languages having voiced stops with negative VOT values (like Spanish (Macken & Barton, 1980b; Eilers et al., 1984) and French (Allen, 1985)), show that two-year-old children have not acquired the VOT values for the initial voiced stops. Italian data from Zmarich et al. (2009) confirm the difficulty for some children even at the beginning of the fourth year of age. The acquisition criteria are the attainment of mean and standard deviation adult values, differing for the consonants place of articulation and the vocalic context (Bortolini et al., 1995). As for anticipatory coarticulation, we agree with Noiray et al. (2018): although coarticulation degree decreases with age, children will not organize consecutive articulatory gestures with a uniform organizational scheme (e.g. segmental or syllabic). Instead coarticulatory organization will be sensitive to the underlying articulatory properties of the combined segments (different lingual coarticulatory resistance and aggressiveness for consonants and vowels according to DAC model (Recasens, 1985), and subject to different articulatory constraints according to Sussman et al. (1999)). Criteria for acquisition are the attainment of mean adult values of coarticulation, differing on the basis of the consonants place of articulation.

METHOD. The data are part of a longitudinal corpus of ten children collected with the aims to study the typical speech development of Italian children. Four female subjects were recruited by one of the authors in Trieste (I), in two kindergarten centers from 2007 to 2009. The parents compiled the MacArthur CDI surveys for their children's lexical productions (Caselli et al., 2007) and filled out a questionnaire reporting information on normal psycho-physical development and monolingual language development (Italian). When they were 18-months-old, the children underwent audiometric screening (*Ling Six Sound Test*, Ling, 1976), to exclude the presence of hearing impairments. The children were recorded every three months from 18 to 48 months (11 sessions). The organization of the session was semi-structured (Schmitt, Meline, 1990), where the child interacted with the clinician in front of a set of toys. The objects were chosen based on the list of words compiled by the parent on the MacArthur CDI. In order to conduct a study on the development of VOT, in addition to saying "common" words, children were invited to repeat each of the 12 VOT test words at least three times. The test items were the following minimal pair pseudo-words, contrasting labial, dental and velar voiced and voiceless stops: *papa*, *baba*, *pipi*, *bibi*, *tata*, *dada*, *titi*, *didi*, *kaka*, *gaga*, *kiki*, *gigi*.

The acoustic files were annotated using *Praat*, through the use of *TextGrids*. All the productions with CV or CVC syllabic structures were selected. The borders marking the beginning of the consonant and the beginning of the vowel were also functional to the measurement of the VOT interval. Target and actual C and V individual segments were labeled. The syllable status as to lexical stress, the position in the word and the style of production (spontaneous or repeated) were also categorized. Finally, a number of exogenous events (like noise or uncertain transcription) or endogenous events, like a number of phonological processes, altering the syllable target, were also categorized.

The *Scriptgart.praat* developed by one of the authors was used to extract the VOT values (ms). For all the vowels present in the corpus, the script then proceeds to extract the values necessary for the calculation of the coarticulation, i.e. the F1 and F2 values in a number of points along the vowel, including the mid. Finally, for the consonant, the F2 values at the beginning of the formant transition are calculated. A number of other variables were obtained as a result of operations among the column of the final matrix, resulting in the duration values of segments and syllables (allowing to estimate the speech rate) and, most importantly, in the exclusion of syllables characterized by VOT values greater than +20 ms for the analysis of coarticulation. Only the syllables located at the beginning of the word that were preceded by silence (equal to or greater than 250 ms) were used for VOT analysis. As for anticipatory coarticulation, it is measured by means of *Locus Equations* (LE, see Lindblom, 1963; Krull, 1989). A LE describes a 1st order regression fit to a scatter of vowel steady-state frequency values predicting the onset of F2 transition values in CV sequences with a fixed C, of the form $F2_{cons} = k * F2_{vow} + b$. This measure provides an overall estimation of coarticulation, provided that LE slopes (indexed by k values) be calculated on CV sequences with vowel pooling and voiced plosives (Tabain, 2000). A nice characteristic of this method consists in an intrinsic normalization of k values, which could vary between 0 (no coarticulation at all) and 1 (maximal coarticulation), allowing the direct comparison of the production of different children at different ages those of adults.

RESULTS. The analysis of VOT from four children (Gaiotto, 2020; Colavolpe, 2020) revealed that differences exist in the chronology of the acquisition of voicing contrast according to the three places of consonant articulation. The variability of the VOT values within a single articulatory place could be an important index of the reorganization of the articulatory system. Most of the children made use of nasalization and fricativisation processes in order to start voicing in an attempt to reproduce voiced targets (like the French and Spanish children, respectively). The contrast of sonority does not assume important variations depending on stress or vowel context, but it seems to be sensitive to the style of production. As for coarticulation, the results confirm that the development of coarticulation can be better described by taking into account the biomechanical characteristics of articulatory gestures. The bilabials allow the maximum temporal overlap of the articulatory movements for which the development gradually proceeds towards a greater temporal synchronization, so that the back of the tongue can be found already in position for the vowel at the moment in which the lips open. Alveolars similarly require independent motor control maturation of two distinct portions of the same articulator (the tongue) to achieve a high degree of coarticulation. And finally velars show high levels of coarticulation from the beginnings, since the consonant and the vowel use the same articulator (the tongue dorsum).