DOI: 10.17469/O2108AISV000022

CLAUDIO ZMARICH, ANNAROSA BIONDI, SERENA BONIFACIO,
MARIA GRAZIA BUSÀ, BENEDETTA COLAVOLPE,
MARIAVITTORIA GAIOTTO, FRANCESCO OLIVUCCI

Coarticulation and VOT in an Italian child from 18 to 48 months of age

This study brings new data to two understudied topics in Italian child language development: VOT and anticipatory C-V coarticulation. One female subject was recorded every three months from 18 to 48 months, while interacting with the clinician in front of some toys, repeating several times each bisyllabic pseudo-word beginning with voiceless and voiced stops. The acoustic signals were annotated using Praat and scripts were created for the automatic extraction of VOT and F2 values (Hz) for the Locus of Equations methods. RESULTS: Voiced stops appeared more difficult to produce than voiceless stops, but voicing contrast was finally achieved from 30 months of age. The degree of coarticulation increased with age, but at 48 months bilabials and alveolars were still less coarticulated than in adults.

Keywords: VOT acquisition, Anticipatory Coarticulation development, Acquisition of Italian, Acoustic analysis, Locus Equations.

1. *Introduction*

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 that language and performance are separate, and language systems are often regarded as developing independently from other cognitive and sensorimotor systems (see Fodor, 1983; for a developmental perspective, see Redford, Oh, 2017). Methodologically, it has been difficult to trace neurobiological markers of the interaction between language and sensorimotor systems; this has been accompanied by an apparent lack of one-to-one correspondences between linguistic units and measures of executive behavior (see Smith, 2010; Laganaro, 2019). Traditional explanations in Linguistics have emphasized the acquisition of phonology, lexicon and morphosyntax, but *not* the acquisition of motor processes related to speech production (Goffman, 2015). The interconnection between linguistic and motor factors in the phonetic development is made complex by the continuous changes in the anatomo-physiological structures (for morphology, size and muscle innervation) and in the neural substrate of cognition (Callan, Kent, Guenther & Vorperian, 2000). Yet the neural organization for sensorimotor and cognitive-linguistic aspects is highly interactive: for instance, 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 motor control, and neurophysiological organization are currently available. According to Redford (2019), these proposals can be assigned to one of the two main existing frameworks: the *Information-Processing Approach* (*DIVA* model, Guenther, 1995; *State Feedback Model*, Parrell, Houde, 2019) and the *Ecological Dynamics Approach* (*Task Dynamics*, Saltzman, Munhall, 1989). When these two approaches were adapted into a developmental perspective, they inherited the preference for the basic units of speech representation and processing that were proper of the original models based on the adults' speech production, i.e. the segment, for the *Information-Processing Approach* (Guenther, Vladusich, 2012) and the word, for the *Ecological Dynamics Approach* (Best, Goldstein, Nam, & Tyler, 2016).

In order to study the acquisition of motor control in early stages of language development, empirical analyses preferably rely on acoustic data, since acoustic analysis offers the possibility to quantify the phonetic continuum in the time-frequency domain, and to derive information – by inference – on the underlying movements. Until a generation ago, only acoustic analysis could be used to infer physiological processes; however, in recent years new technology and methods have been developed for the analysis of physiological processes, which are non-invasive and compatible with the analysis of evolutionary subjects (Goffman, 2015), although they can be used only from the subjects' fourth year of age. Among these, the most successful ones are the *Ultrasound Tongue Imaging-UTI* (Abakarova, Iskarous & Noiray, 2018; Noiray, Abakarova, Rubertus, Krüger & Tiede, 2018; Zarkhova, 2018; Noiray, Wieling, Abakarova, Rubertus & Tiede, 2019a; Noiray, Popescu, Killmer, Rubertus, Krüger & Hintermeier, 2019b; Barbier, Perrier, Payan, Tiede, Gerber, Perkell & Menard, 2020; Cychosz, Munson & Edwards, 2021) and *Optical-tracking* devices like the *Optotrack* (Stone, 2012), which have sometimes been combined together. Nonetheless, some fields of inquiry in acquisition studies, namely Voice Onset Time (VOT) and the development of anticipatory coarticulation, have been shown to be easy to investigate by means of acoustic analysis only, thanks to the relatively clear and linear relation between speech movements and acoustic effects (Kent, Kim, 2008; Harrington, 2010; for VOT, see Grigos, Saxman & Gordon, 2005; Solé, 2018, as well as the contributions in the special number the of *Journal of Phonetics* devoted to VOT and edited by Cho, Whalen & Docherty, 2019; for anticipatory coarticulation, see Iskarous, Fowler & Whalen, 2010; Lindblom, Sussman, 2012). Last but not least, acoustic analysis is non-invasive, inexpensive and relatively simple to perform, and becomes the most viable solution for collecting and analysing huge amounts of data.

2. *Aims of this study*

The present investigation aims to contribute with experimental data, by means of acoustical analysis, to two topics in Italian child language development, namely the acquisition of VOT and the development of anticipatory CV coarticulation. These topics have been thoroughly studied in other western languages such as English, Spanish or French (see, respectively, Macken, Barton, 1980a; Eilers, Oller & Benito-Garcia, 1984; Allen, 1985, for VOT; and Sussman, Duder, Dalston & Cacciatore, 1999, for anticipatory coarticulation in American English), but rarely investigated in Italian, a so-called "true voicing language" (for VOT, see Bortolini, Zmarich, Fior, & Bonifacio, 1995; Zmarich, Bortone, Vayra & Galatà, 2013; for coarticulation, see Petracco, Zmarich, 2006; Zmarich et al., 2013).

2.1 The acquisition of VOT

The best parameter for quantifying and classifying voicing contrasts is VOT, which measures the time elapsed from the release of the consonant occlusion to the beginning of the vibration of the vocal folds. VOT provides an inferential estimate of speech motor control, requiring fine motor coordination of the respiratory, phonatory and articulatory structures. Early in phonetic development the voiced and unvoiced consonants tend to be realized as voiceless unaspirated, which allows for the synchronization between glottal and supraglottal events. It is only after the acquisition of additional articulatory manoeuvres that children come to achieve all the VOT categories that characterize their native language (see below). In different languages the phonemic contrast between sonority categories corresponds to distinct temporal intervals along the VOT continuum. Extensive cross-language studies (Abramson, Whalen, 2017; Cho, Whalen & Docherty, 2019) have shown that three categories of stops, having a rough correspondence across languages, emerge along the VOT continuum:

1. "voicing lead": characterised by negative VOT values, ranging from about -125 to -75 ms. Italian voiced stops belong to this category.
2. "short voicing lag": characterised by positive VOT values, ranging from 0 to +30 ms. Italian voiceless stops and English voiced stops belong to this category.
3. "long voicing lag": characterised by highly positive VOT values, ranging from +60 to +100 ms. English voiceless stops belong to this category.

Languages make use of these categories by having particular mean and range VOT values (Cho, Whalen & Docherty, 2019). Although many languages select only two among these voicing categories, there are languages which make use of more than two categories: among the languages with three-way contrasts are Thai, Vietnamese, Khmer; among the languages with more than three-way contrasts are Urdu, with a four-way contrast, and Sindhi with a five-way contrast. In fact, VOT is by no means the only mechanism available to languages for contrasting voicing: Cho et al. (2019) report other phonetic dimensions contributing to voicing contrast, such as consonant-induced F0 and voice quality.

VOT values have also been shown to depend on the specific stop and following-vowel sequence, due to aerodynamic reasons (Rothenberg, 1968): duration increases gradually as vowel openness decreases and the constriction for the plosive moves farther back in the vocal tract.

From a developmental point of view, in short lag/long lag VOT languages such as English, the acquisition of VOT contrast is usually accomplished earlier than in short lag/long lead languages such as Spanish, French and Arabic (Allen, 1985). Children acquire the short lag/long lag contrast around the age of 2;0 (Macken, Barton, 1980b). In languages that have VOT contrasts with short lag/long lead similar to that of Italian, children develop VOT contrasts around the age of 4;0 (Allen, 1985; Eilers et al., 1984; Al-Tamimi, Tarawneh & Howell, 2021).

2.2 The development of coarticulation

Acoustic analysis is also useful for the study of coarticulation, which refers to the temporal overlap of gestures belonging to neighbouring phones (Hardcastle, Hewlitt, 2006; Farnetani, Recasens, 2010). In a detailed survey, Mildner (2018) explains that coarticulatory phenomena can be originated by both biomechanical and linguo-specific factors. Coarticulation is the result of a continuous modification and adaptation of articulation to the linguistic context. First, this is a function of the biomechanical constraints of the phono-articulatory system for speech motor control, which are supposed to be universal – since they are biologically determined. Secondly, modification and adaptation are related to linguo-specific factors, because central planning and organization processes are governed by linguistic rules that differ depending on the speaker's language. While carryover coarticulation has been mainly explained as a result of articulatory inertia, thus mantaining a universal character, anticipatory coarticulation, which is under the lens of the present research, has a cognitive basis, because it must be planned in advance, and it mantains a linguo-specific character. The development of coarticulation is a very debated issue (Noiray et al., 2019a), but general consensus has been reached now on the fact that young children generally coarticulate more than adults. A number of explanations have been offered for the greater children's coarticulation, including:

> (1) low-level pressures upon speech due to the protracted development of domain-general fine motor control, (2) the phonological reorganization of speech from more holistic units, such as syllables, into phonemes, (3) cognitive pressures from children's limited working memory and speech planning capacities, (4) children's inexperience articulating the sounds and words of their native language(s), or (5) a combination of these explanations (Cychosz et al., 2021: 367).

Recently, other factors have been discovered that possibly contribute to the acquisition of the degree of anticipatory coarticulation characterising adults' productions: increase in phonological awareness, vocabulary size (Noiray et al., 2019b; Cychosz et al., 2021) and degree of speech practice (Cychosz et al., 2021). All these factors contribute to lower the coarticulation degree, by separating the articulation of the consonant from that of the following vowel.

According to Noiray et al. (2019a), the three main hypotheses on the development of coarticulation are:

a. “Holistic Approach”
b. “Segmental Approach”
c. “Gestural Hypotesis”

According to the “Holistic Approach” in the first phase coarticulation units would be large, the articulators' movements would be interdependent (Nittrouer, Studdert-Kennedy & Neely, 1996), and lexical development would have an important role in defining which combinations the child is able to achieve with coarticulation (Vihman, Velleman, 1989). Therefore, large units with interdependence of movements would shrink, and show an even greater motor independence of each articulator to achieve new segment combinations.

The “Segmental Approach” hypothesis could be considered as the exact opposite of the “Holistic Approach”. Initially, the articulatory movements to produce consonants and vowels would be independent of each other, and the first articulatory skills would manifest themselves in the realization of individual segments. Subsequently, depending on motor maturation and the increase in cohesion between movements, coarticulation over segments would develop. According to this hypothesis, thanks to the incremental maturation of the motor control of each articulatory organ, the precision and spatial-temporal coordination between segments would also improve (Kunhert, Nolan, 1999).

Finally, the “Gestural Hypotesis” refers to the coarticulation models developed starting from the Articulatory phonology theory of Browman, Goldstein (1992). Based on a concept of coarticulatory gesture that extends over the phonological segment, the development of coarticulation would depend on the degree of coarticulation compatibility of each gesture. The coarticulatory organization would take place in a more holistic or segmental way depending on whether the articulatory organs necessary to carry out the adjacent articulatory gestures compete with each other or not.

According to the most important studies on anticipatory coarticulation using the Locus Equations' method (Sussman et al., 1992; Sussman et al., 1999; see also Gibson, Ohde, 2007), 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 that coarticulation decreases and phonemic distinctiveness increases. But the process is not linear and strongly depends on the physiological constraints on the articulators. The studies by Sussman and colleagues are important because they are among the first and the few to have investigated, with the use of acoustic analysis, the development of coarticulation starting from the beginning of babbling (at seven months of age), while the lowest age investigated using UTI devices starts from the fourth year.

As claimed by Sussman and colleagues, 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 a 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 tip (for the consonant) and the dorsum of the tongue (for the vowel), which are largely independent. As for velar consonants, the biomechanical constraints are the largest (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).

2.3 Experimental hypotheses

It is well-established that it is difficult to produce vocal fold vibration during voiced stops due to challenging aerodynamic conditions. Ohala (2011: 64) described the existence of an "Aerodynamic Voicing Constraint" [...]:

> voicing requires a sufficient airflow through the adducted vocal cords. The airflow requires a sufficient pressure difference (ΔP) between subglottal pressure (Ps) and oral pressure (Po). During an obstruent air accumulates in the oral cavity thus increasing Po. When the Po approaches Ps, the airflow falls below that needed for vocal cord vibration and thus voicing is extinguished.

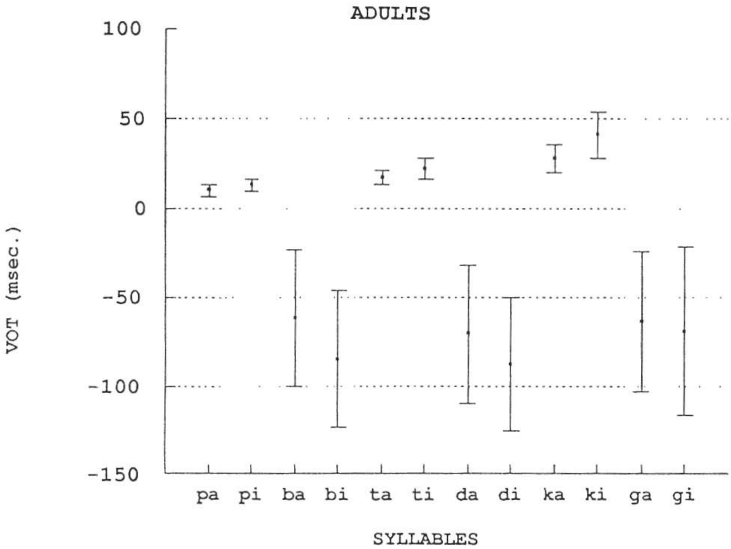
Speakers of languages like Italian where voiced stops require laryngeal vibration may need some articulatory adjustments, used singly or in combination, to limit the increase of oral pressure and thus achieve the pressure differential for voicing initiation. Such articulatory adjustments are: oral leak, nasal leak (i.e., releasing airflow through an incomplete velopharyngeal closure), larynx lowering, tongue root advancement (Rothenberg, 1968; Westbury, 1983). Among these adjustments, only the larynx lowering, and maybe the tongue root advancement, could guarantee the achievement of the perceptual effect of voicing without introducing collateral effects such as nasalisation, frication or vowel epenthesis. The presence of consistent percentages of segment productions that are nasalized or preceded by a *schwa*-like vowel in adult speakers of true voicing languages like French or Spanish (Solé, 2018) is explainable with perceptual considerations that are not investigated here.

Developmental studies on the acquisition of voicing in languages with voiced stops with negative VOT values show that two-year-old children have still not acquired the VOT values for initial voiced stops (for Spanish: Macken, Barton, 1980b; Eilers et al., 1984; for French: Allen, 1985; for Jordanian Arabic: Al-Tamimi et al., 2021). Italian data from Zmarich et al. (2013) confirm this difficulty for some children even at the beginning of the fourth year of age. If the acquisition criteria are the attainment of mean and standard deviation of adult values for consonant place of articulation and vocalic context (see Fig. 1, from Bortolini et al., 1995; see also Esposito, 2002), then most of the children in Zmarich et al. (2013) are still far from target values.

On the basis of both literature results and our previous investigations, the present study aims to test the following working hypotheses:

1. VOT: Since the production of initial voiced stop consonants requires some larynx-external mechanism in order to sustain an adequate transglottal pressure drop during stop closure (as an active lowering of the glottis), children will be more advanced in the acquisition of appropriate VOT values for voiceless than for voiced consonants.

Figure 1 - *VOT values as produced by Italian adults (from Bortolini et al., 1995)*



2. Anticipatory coarticulation: While the coarticulation degree decreases with age, children will not be able to organize consecutive articulatory gestures with a uniform organization scheme (e.g., segmental or syllabic) (Noiray et al., 2018). Instead, coarticulatory organization will be subjected to different articulatory constraints according to Sussman et al. (1999) and sensitive to the underlying articulatory properties of the combined segments (different lingual coarticulatory resistance and aggressiveness for consonants and vowels according to the DAC model). The model predicts that "*the size, temporal extent, and direction of lingual coarticulation are conditioned by the severity of the requirements imposed on the tongue for the production of vowels and consonants*" (Farnetani, Recasens, 2010: 340). Criteria for acquisition are the attainment of mean adult coarticulation values, as indexed by the 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 F2cons = k * F2vow + c. 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 stops (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 productions by children at different ages to those by adults. Petracco, Zmarich (2006) established the following values for k, averaged over the productions of four Italian adults (see Fig. 2, from Petracco, Zmarich, 2006). We also report the magnitude of R^2, which in the regression analysis is used to indicate the fraction of variance of the dependent variable scores explained by the independent variable scores:

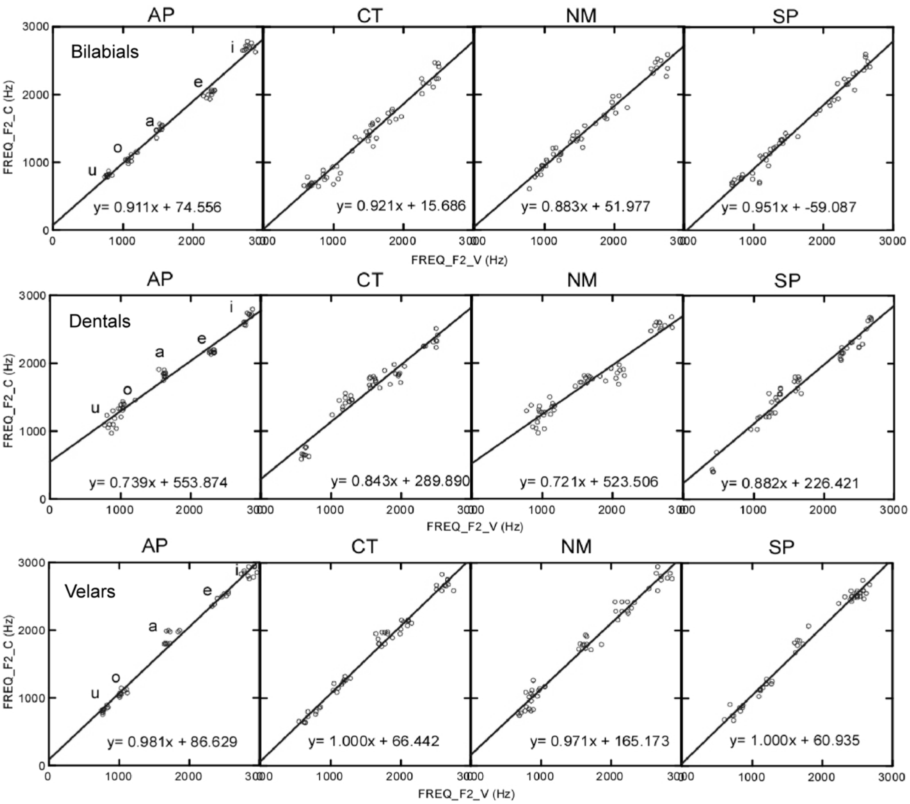
a. Bilabials: k = 0.915, R^2 = 0.973
b. Dentals: k = 0.790, R^2 = 0.952
c. Velars: k = 0.989, R^2 = 0.983

The work presented here is exploratory in nature, aiming to: (i) review the state of the art on VOT acquisition and coarticulation development; (ii) present a theoretical frame with which to formulate the experimental hypotheses and the appropriate methods for future investigations; and (iii) apply these methods to investigate one subject in a longitudinal design, thus keeping aside, for the moment, potentially complicating questions like individual variability.

3. *Subjects and Methods*

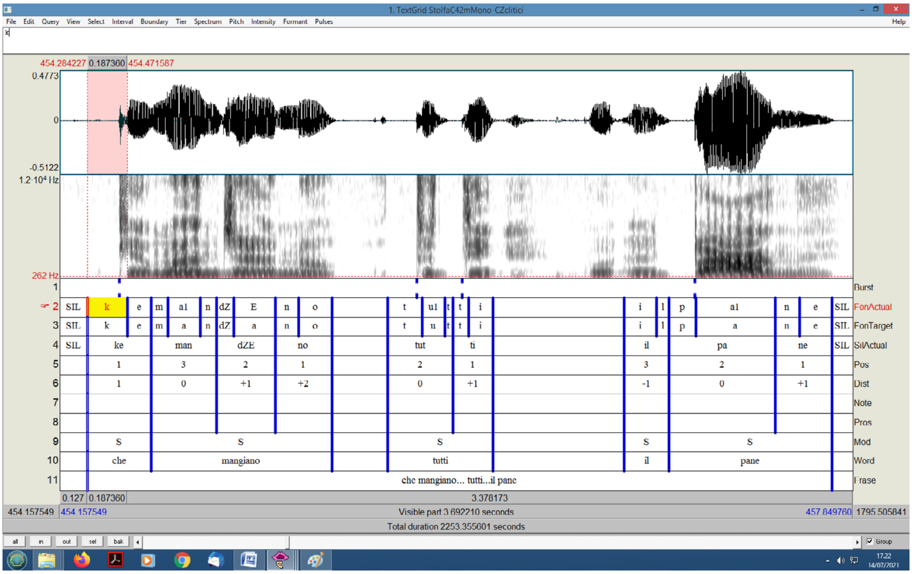
This individual case study is part of a longitudinal corpus of ten children collected with the aim to investigate Italian children's speech development. The study uses data from both the phonetic transcription of segments and the acoustic analysis of speech. The subjects were recruited by one of the authors (S. Bonifacio) in Trieste (Italy) from 2007 to 2009, in two kindergartens. The parents compiled the MacArthur CDI survey for their children's lexical productions ("Primo Vocabolario del Bambino", Caselli, Pasqualetti & Stefanini, 2007) and filled out a questionnaire reporting information on normal psycho-physical development and monolingual (Italian) language development. All the parents signed an informed consent form. When children were 18-months old, they underwent auditory screening (*Ling Six Sound Test*, Ling, 1976) to exclude the presence of hearing impairments.

Figure 2 - *Diagram distribution of pairs of F2onset and F2vowel values (Hz) relative to syllables p/bV (top), t/dV (middle), k/gV (bottom) produced by the four adult subjects in Petracco, Zmarich (2006). The points within each diagram are interpolated by the regression line, whose equation is reported at the bottom*



The children were recorded every month from 18 to 24 months, and then every three months from 24 to 48 months. The session organization was semi-structured (Schmitt, Meline, 1990), with the child interacting 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 items at least three times, immediately after they were uttered by the third author. The test items were the following minimal pair pseudo-words, all stressed on the first syllable, and contrasting labial, dental and velar voiced and voiceless stops: *papa*, *baba*, *pipi*, *bibi*, *tata*, *dada*, *titi*, *didi*, *kaka*, *gaga*, *kiki*, *gigi*. Each recording session lasted on average about one hour. Once the session was over, the sample of language collected at 18 and 21 months was considered valid and representative of the child's linguistic abilities only if the number of lexical forms produced represented at least 50% of the words in the lexical list compiled by the parent. All the recordings are available in digital format (.wav) at a rate of 44.1 kHz and 16 bit (Edirol R-09 by Roland). SC is a female child. When she was 18-months old she was credited with a vocabulary of 120 words by their parents (through CDI), and then she spontaneously produced more than 50% of the words in the lexical list compiled by the parent during the first session. For statistical reasons, we chose to add to stage one also the speech productions from the 19th month of age. From there onwards, the sessions we chose to analyse for the present contribution concern the 24, 30, 36, 42 and 48 months of age, i.e., every 6 months. The acoustic files were annotated using *Praat* (Boersma, Weenink, 2021) with TextGrids (see Fig. 3).

Figure 3 - *Example of a typical Praat display as used in this study*



For each file, all the productions containing CV or CVC syllabic structures where C was a stop consonant were selected. If the words were not isolated, we segmented and labelled also all other words in the utterance. As for segmentation, we followed the conventions proposed by Salza (1990). For voiced consonants, the boundaries marking the consonant beginning and the vowel beginning (where the burst separates the consonant from the following vowel) were used for the measurement of the VOT interval. For voiceless consonants VOT was measured from the burst, representing the beginning of VOT, up to the vowel onset. The voiceless stop beginning a new utterance was credited with a duration interval of 150 ms before the burst, a conventional value based on previous determination of the average duration of intervocalic voiceless stops in Italian children of the same age.

Children are often expected to misproduce the phonetic shape of target words (with respect to the adult pronunciation). This is due to a number of reasons including: the presence of so-called phonological processes; phonetic phenomena resulting from informal connected speech; voicing errors; other phenomena described later. Thus, individual target and actual C and V segments were labelled in separate tiers using SAMPA symbols. The syllable status as to lexical stress, position in the word and style of production (spontaneous or repeated) were categorized. Clitics were considered a part of prosodic words. Finally, a number of exogenous events (like noise or uncertain transcriptions) or endogenous events, like a number of phonological processes (sometimes related to the production of voicing), altering the syllable target, were also categorized (see later).

All annotations were done by one of the authors (mostly, but not always, M. Gaiotto, as part of her MA thesis). Other annotations were done by B. Colavolpe. Both annotators were Linguistics students at the University of Padova and, at the same time, speech therapists working with children. They had attended an internship at ISTC-CNR[1] under the tutoring of the first author, where they practiced with the methods employed in this study, following a written protocol continually updated and discussing problematic cases with the first author. All of their textgrids were checked by the first author, and, if necessary, corrected and integrated. A number of ambiguous cases were left out. A Praat script developed by one of the authors (F. Olivucci) was used to extract the VOT values (ms). For all the vowels in the annotated textgrid, the script extracts the values for the calculation of coarticulation, i.e., F1 and F2 values in a number of points along the vowel, including the middle. For consonants, the F2 values at the beginning of the formant transition are calculated. The algorithm was designed so as to extract the optimal formants for each vowel as well as the parameters change as a function of the vowel position in the vowel quadrilateral. Vowels are then divided accordingly into 3 macrocategories.

The script produced a .csv file, which allowed to obtain a number of other variables as a result of operations among the columns of the matrix, yielding the duration values of segments and syllables. In this way it was possible to estimate speech rate (not considered here) and, most importantly, to exclude syllables characterized by VOT values greater than +15 ms for the coarticulation analysis, that is, only the syllables characterized by a negative VOT interval or a positive VOT interval lower than 16 ms were selected. Only syllables preceded by silence (equal to or greater than 250 ms) were used for the VOT analysis, as this context is considered the most challenging for speakers of languages, like Italian, that produce voiced stops by means of negative VOT (i.e., with true voicing, see Ohala, 2011; Solé, 2018).

As for anticipatory coarticulation, it was measured by means of Locus Equations (LE, see Lindblom, 1963; Krull, 1989). As described above, 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 F2cons = k * F2vow + c.

Regression analysis is a statistical procedure that describes the relation between two variables: a dependent one, which in this case corresponds to the consonant F2 value; and an independent one, which corresponds to the vowel F2 values, the aim being to analyze how the consonant value changes as the vowel value varies. When

[1] Istituto di Scienze e Tecnologie della Cognizione – CNR, Padova, Italy.

doing regression analysis, the *Systat* software for statistical analysis also calculates, among other things, the values of 4 parameters that are useful for the present study:

1. k: indicates the slope of the regression line, describing the anticipatory coarticulation degree of the vowel on the consonant;
2. c: the intersection of the regression line on the y axis; it thus corresponds to the value assumed by the dependent variable when the independent variable (x) is equal to 0; Iskarous, Fowler & Whalen (2010) suggest that this parameter is directly related to the degree of involvement of the body of the tongue in the realization of the constriction for the consonant production.
3. R^2: indicates the determination coefficient, which measures the variance fraction, that is the amount of variance of the dependent variable as accounted for by the independent variable.
4. SEE: the standard error of estimate is a variability index measuring the distance of the data points from the regression line.

4. *Results*

The number of CV and CVC syllables beginning with a stop consonant produced by SC and constituting the database for the present work is 3255. The presentation of the results will be organized around three main points, following a longitudinal perspective: (1) a qualitative analysis of the relative frequency of the type of endogenous and exogenous processes (see Tab. 1) which altered the potential consonant targets up to their exclusion from further analyses. This investigation becomes particularly interesting when we focus on the utterance-initial syllables as a particularly challenging position for the production of voiced consonants; (2) a statistical analysis of the VOT values for voiced and voiceless targets realized in the absence of processes; (3) a statistical calculation of the degree of anticipatory coarticulation for each of the three main places of articulation, pooling together initial and non-initial syllables (only those with a VOT value lower than 15 ms, see later).

Table 1 - *Types of endogenous and exogenous processes (see text for explanations)*

1	Saturated signal
2	Weak signal
3	Noised signal
4	Unidentified target
5	Uncertain transcription
6	Breathy voice
7	Creaky voice
8	Whispered
9	Ejectivisation (stop realized in an ejective pneumatic modality)
10	Diphthongization
11	Vowel Epenthesis (oral leakage)
12	Consonant Epenthesis

13	Consonant deletion
14	Weak syllable deletion
15	Nasalization (nasal leakage)
16	Frication (oral leakage)
17	Affrication
18	Gliding
19	OK_Fronting
20	OK_Backing
21	OK_Denasalization
22	OK_Metathesis
23	OK_Stopping
24	OK_Consonant harmony
25	OK_Cluster reduction
26	OK_Vowel epenthesis

4.1 Qualitative assessment of the frequency of endogenous and exogenous causes of exclusion from the analysis of potentially eligible targets

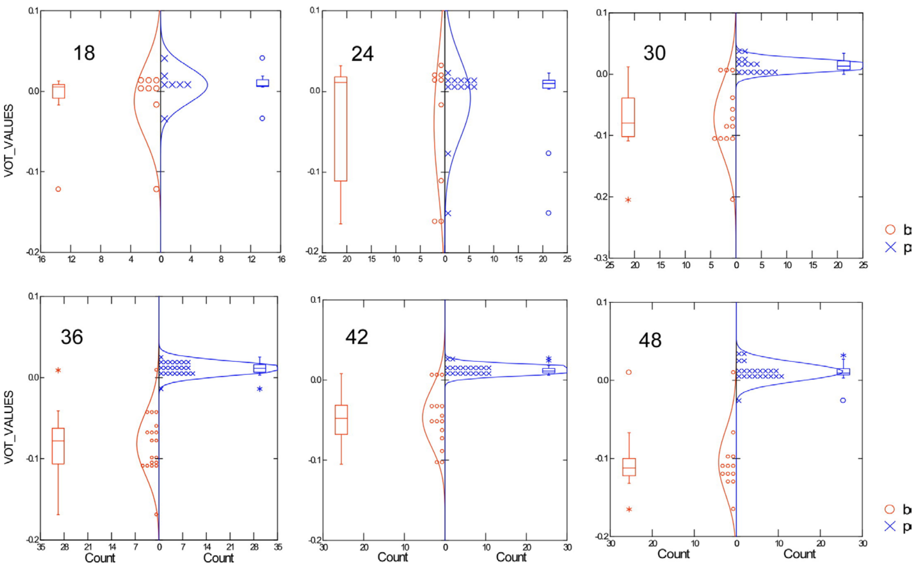
In our definition, exogenous causes of exclusion from the analysis of the CV and CVC syllables beginning with a stop consonant are those external to the child, such as, for example, ambient noise or technical problems with the microphone, the regulation of the level of energy input, failure in the identification or transcription of a target word (causes numbered 1 to 5 in Tab. 1). These causes are scarcely interesting from a scientific point of view, and their low frequency (39 syllables out of 3255, or 1.19%) testifies of the generally good quality of the recordings. The endogenous causes are much more scientifically informative. These led to the exclusion from the analyses of potentially eligible CV and CVC syllables, due to systematic modifications (i.e., phonological processes) and/or non-modal phonation (whisper, breathy voice, falsetto, creaky voice) of the phonetic form of the target words, as compared to the adult pronunciation (causes numbered 6 to 18). Endogenous cases amount to 169 syllables (10.38%). Finally, the third category of endogenous causes still regards systematic modifications made by the child to the phonetic shape of the adult target, but, crucially, they preserve or create the plosive-vowel structure for VOT and degree of coarticulation which is being investigated in the present study (see phonological processes like Fronting, Backing, Stopping, etc., numbered 19 to 26 in Tab. 1). This category amounts to 30 syllables (0.09% out of 3255 syllables).

The usefulness of this kind of analysis is clear when it is applied to utterance initial syllables produced with a negative VOT. Especially during the 24-month and 30-month sessions, the articulatory adjustments identified as extreme (i.e., strongly evident in the acoustic signal), that is, “Vowel epenthesis”, “Frication”, and “Nasalization”, are the most frequent causes of exclusion. All these articulatory adjustments have the effect of reducing the oral pressure increase, and thus achieve the pressure differential for voicing initiation: nasal leak, larynx lowering, tongue root advancement, oral leak (Solé, 2018). In the following sessions their frequency decreases.

4.2 VOT acquisition

Before running the statistical analysis, we eliminated all the child's productions of the utterance-initials CV and CVC stop targets which were affected by any of the exogenous or endogenous processes listed in Tab. 1, with the exception of the processes numbered 18 to 26, which preserved the nature of plosive consonants. For statistical reasons, and for the sake of simplicity, in order to be able to represent the phonological sonority contrast for all the sessions, here we pooled together all the initial CV and CVC syllables regardless of the vowel (all possible vowels, but with a prevalence of [a], and [i]), the "lexical stress" status (stressed or unstressed syllable), or the style of production (spontaneous vs. repeated). A series of separated t-student tests was then performed in order to assess the significance of the voicing contrast at each session.

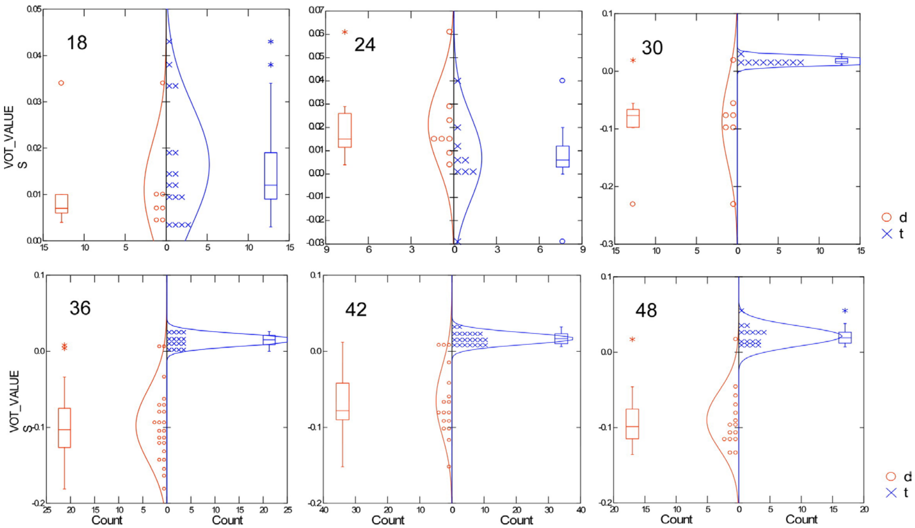
Figure 4 - *Sampling distributions and box plots of the VOT values (s) for each age session, divided according to the voiced vs. voiceless nature of "adult" stop targets, for the bilabial place of articulation*



For each session and for each place of articulation, the child was able to produce a number of occurrences that was sufficient to allow statistical comparisons by means of the t-student test (which however is known to be quite robust to atypical sampling distribution), with the exception of the first session at 18 months of age, in which the child did not produce any phonologically voiced velar stop.

Fig. 4 plots the sampling distribution of the VOT values, divided according to the voiced vs. voiceless nature of "adult" stop targets, as produced by the child in each session, for the bilabial place of articulation. The SC productions at the 18- and 24-month sessions did not differ significantly for voicing, because of the production of voiced targets by means of positive VOT. The first significant difference was achieved at 30 months of age, when voiced and voiceless productions were produced with the mean (and SD) values of -0.072 s (0.060) and 0.014 s (0.009), respectively ($t_{(12.5)} = -5.137$, $p < 0.000$).

Figure 5 - *Sampling distributions and box plots of the VOT values (s) for each age session, divided according to the voiced vs. voiceless nature of "adult" stop targets, for the dental place of articulation*

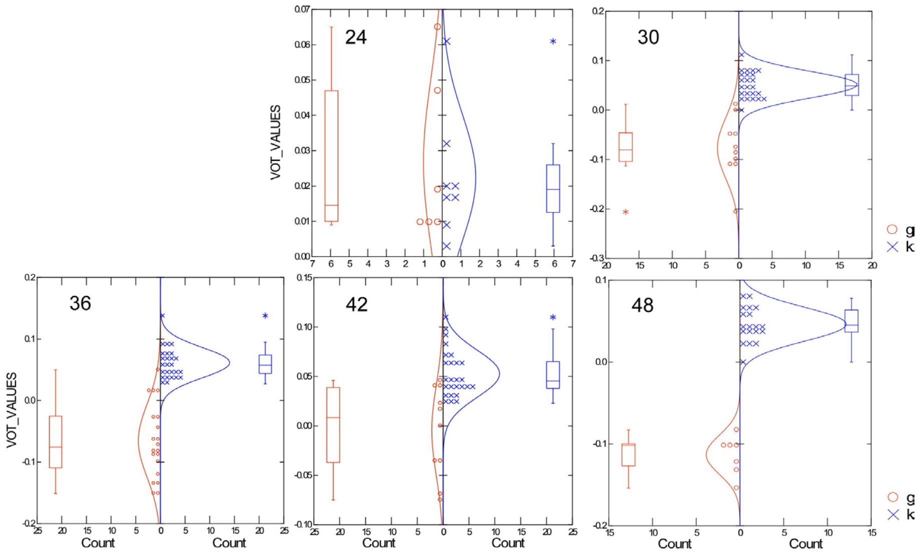


From that time onwards, the child was able to maintain the voicing contrast, with a slight reduction of the negative VOT values at 48 months. The mean (and SD) values at the 48-month final stage were -0.105 s (0.041) for /b/, and 0.011 s (0.011) for /p/.

As for dental stops (Fig. 5), SC's productions at the 18- and 24-month sessions did not differ significantly for voicing, mainly because of the production of voiced targets by means of positive VOT. The first significant difference was achieved at 30 months of age, when voiced and voiceless stops were produced with the mean values (and SD) of -0.098 s (0.075) and 0.015 s (0.006), respectively ($t_{(23.5)}=-11.113$, $p < 0.000$). From that time onward, the child was able to maintain the voicing contrast, with a slight increase of the negative VOT values at 48 months. The mean (and SD) values at the 48-months final stage were -0.090 s (0.039) for /d/, and 0.021 s (0.013) for /t/.

As for velar stops (Fig. 6), at the initial session the child did not produce the voiced targets, and at 24 months of age her targets did not differ significantly for voicing, but at 30 months of age she appeared to be able to differentiate significantly voiced from voiceless targets, with the mean (and SD) values of -0.077 s (0.062) and 0.052 s (0.027) respectively ($t_{(10.4)}= -6.302$, $p < 0.000$). From that time onward, she was able to maintain the voicing contrast, with a sharp increase in negative VOT values at 48 months and an important reduction in variability as compared to the previous session. The mean (and SD) values at the 48 months final stage were -0.113 s (0.021) for /g/, and 0.046 (0.024) for /t/.

Figure 6 - *Sampling distributions and box plots of the VOT values (s) for each age session, divided according to the voiced vs. voiceless nature of "adult" stop targets, for dental place of articulation*



4.3 Coarticulation development

Before running the statistical analysis, we eliminated all the child's CV and CVC stops which were affected by any of the exogenous or endogenous processes listed in Tab. 1, with the exception of the processes numbered 18 to 26, which preserved the nature of plosive consonants.

As explained in section 2 above, we excluded all syllables with VOT values greater than +15 ms. The rationale behind this decision was that, if one wants to use the F2 value measured at the beginning of the formant transition as an acoustic index of the articulatory position of the tongue dorsum, one cannot allow the tongue to have the time to move away from the articulatory place of the occlusion (at least for dentals and velars). Ideally, one has to rely only on voiced consonants, but in this case one is at risk of statistical bias for the small amount of occurrences, especially during the first sessions.

Finally, we performed a series of linear regression analyses, separated for articulation place within any session. The frequency values (Hz) of the F2 transition sampled at its onset was the dependent variable, while the frequency of F2 when measured in the vowel nucleus was the independent variable.

We obtained a scatterplot fit with a linear regression line, the 'locus equation,' of the form F2cons = k *F2vow + c where k and c are the constants, slope and y-intercept, respectively (see also § 2.3).

The statistical values represented by number of occurrences, slope (k), Intercept (c), R^2 and SEE (for their meaning, see § 2) characterizing the linear regressions are shown separately for place of articulation. Tab. 2 reports the values for bilabials. In this case, the mean value of k for adults is 0.915, with R^2 well over 0.950 (Petracco, Zmarich, 2006). With the exception of the value of k for the first session, which is quite low, the other values range around 0.700, with the value of the last session (0.680) still scoring far away from the adults' value. R^2 is quite low and SEE too high, indicating respectively a low explicative power of F2vow over F2cons variance, and a high variability.

Table 2 - *Parameters' scores from regression analysis for bilabials, see text for explanation*

Month	Slope (k)	Intercept (c)	R^2	SEE	Occurrences
18	0.597	1037.9	0.311	601.3	20
24	0.718	558.9	0.559	439.1	48
30	0.750	811.2	0.441	654.8	74
36	0.696	1015.2	0.466	620.0	116
42	0.769	649.7	0.444	619.5	102
48	0.680	936.7	0.452	584.7	92

Tab. 3 reports the values for dentals. In this case, the mean value of k for adults is 0.790, with R^2 scoring 0.952 (Petracco, Zmarich, 2006). With the exception of the value of k for the first session, which is particularly low, the values for the last three sessions range around 0.800, with the value of the last session (0.725) not far from the value for adults. R^2 is quite low and SEE still too high, respectively indicating a low explicative power of F2vow over F2cons variance, and a high variability.

Table 3 - *Parameters' scores from regression analysis for dentals, see text for explanation*

Month	Slope (k)	Intercept (c)	R^2	SEE	Occurrences
18	0.258	1797.3	0.223	173.7	35
24	0.682	1003.2	0.632	347.6	52
30	0.515	1526.7	0.408	472.8	84
36	0.816	797.4	0488	546.5	126
42	0.893	542.5	0.569	506.5	107
48	0.725	972.1	0.523	461.6	108

Tab. 4 reports the values for velars. For these stops, the mean value of k for adults is the highest (0.989), and the same happens for R^2 (0.983) (Petracco, Zmarich, 2006). With the exception of the values of k for the first two sessions, which are not reliable due to the low number of occurrences, the results for the last three sessions are highly variable, although the value of the last session (0.787) is the highest across all recording sessions and places of articulation. As for R^2, if one disregards the value at 42 months (accompanied by a high SEE value), the value at 36 and 48 months (accompanied by relatively low values of SEE) may indicate a gradual approximation to the adult norm.

Table 4 - *Parameters' scores from regression analysis for velars, see text for explanation*

Month	Slope (k)	Intercept (c)	R^2	SEE	Occurrences
18	0.807	894.2	0.663	523.8	11
24	0.919	231.6	0.990	56.9	5
30	0.451	1636.2	0.224	682.2	27
36	0.863	792.8	0.858	295.6	33
42	0.595	1650.3	0.430	620.5	12
48	0.787	929.9	0.652	417.0	23

5. *Discussion and conclusion*

This study aims to bring new data to two understudied topics in Italian child language development: VOT and anticipatory C-V coarticulation.

As for VOT, the results show that the difficulties related to the production of negative VOT values for voiced plosives that are still present at 24 months are overcome for all articulatory places in the 30-month session. The reduction in variability in the VOT values within each articulatory place at the final stage (vs. the 30-month stage) could be an important index of the reorganization of the articulatory system (as underlined by all the authors cited in the introduction). Although not supported by a direct statistical analysis, the mean and SD values at the 48 months final stage are ostensibly similar to those of the adult's values in Petracco, Zmarich (2006); in particular, they parallel the difference in magnitude relative to the articulation place (bilabials<dentals<velars). Further, during the second and third sessions, SC made use of vowel epenthesization, nasalization and frication processes in an attempt to reproduce voiced targets (like French and Spanish children, respectively). However, it is worth considering that we did not analyse the months between the 24 months and the 30 months of age: it is entirely possible that the child achieved the voicing contrasts after the 24 months but before the 30 months, possibly with a difference related to the articulation place. In the next future we will look into this possibility by investigating the 27 months session.

In discussing coarticulation development, we should mention that, for adults, several studies on various languages have established the following hierarchy between articulatory places as a function of k values: B(ilabials)> V(elars)> A(lveolars), while their hierarchy as a function of c is the exact reverse: A> V> B (Iskarous, Fowler & Whalen, 2010). In the Italian adult subjects of Petracco, Zmarich (2006), the following orderings were observed: k: V> B> A; c: A> V> B.

Coarticulation follows different development profiles depending on the consonant place considered and it is possible to explain the differences based on the strength of anatomo-physiological constraints, as predicted by Sussman et al.

(1999), and confirmed recently by many authors using the UTI technology (see for instance, Noiray et al., 2018, 2019a, 2019b, 2020).

The bilabial occlusion, not being anatomically binding for the tongue, would allow the largest temporal overlap of the two gestures, as occurs in adults, where the coarticulatory influence is greatest: at the time of release the tongue is already in the position for the vowel (coarticulation as an articulatory co-production). In the 18 months session, SC hardly shows coarticulation: the tongue is still in motion when the occlusion is released. From the subsequent session, independent movements of the tongue for the production of the vowel begin to be loosely timed with respect to the release. Once a first level of coordination is reached, it is maintained.

As for dentals, the child must learn to differentiate almost independently and coordinate the tip and the back of the tongue (anatomical constraints). In adults, coarticulation has little acoustic effects, since the tip of the tongue can theoretically remain stationary on the consonant place regardless of the position of the dorsum which articulates the vowel. In line with this hypothesis, at the 18-month session coarticulation is very low, then it increases, gradually approaching adults' values.

In the case of velars, the biomechanical constraints are the largest (the consonant and the vowel use the same articulator, the tongue dorsum), and in the adult the acoustic effect shows a very high degree of coarticulation (as a reciprocal articulatory adaptation). In SC, after an initial period of variability, productions stabilize at relatively high values, which show a greater amount of coarticulation than the other place categories examined.

Trying to interpret the results with reference to the three hypotheses on coarticulation development put forward by Noiray et al. (2019a) and discussed in § 2.2 above, the present results are more in agreement with the Gestural Hypothesis, which predicts that coarticulatory organization takes place in either a holistic or segmental way depending on whether the articulatory organs necessary to carry out adjacent articulatory gestures compete with each other or not.

The limitations of the present study are many, and most of them are due to the single case design. Thus, this study cannot address basic questions in any developmental investigation, such as individual variability and the many factors behind this (gender, education, socio-economic conditions, etc...). In addition, the effect of important variables such as vocalic context and place of articulation for stops in VOT could not be statistically determined, due to the low statistical representativity, and we did not normalize for speech rate. Another direction for future investigations will be to submit to a perceptual judgement the productions affected by vowel epenthesization, nasalization and frication processes. In fact, we know after Solé (2018) that many adults' voiced productions are affected by some process of that type, and nevertheless they are tolerated and considered acceptable instances of voiced productions. We hope that in future developments of this project, all these problems will be addressed as we will extend the investigation to all of the ten subiects in the "Trieste" children corpus.

Bibliography

ABAKAROVA, D., ISKAROUS, K. & NOIRAY, A. (2018). Quantifying lingual coarticulation in German using mutual information: An ultrasound study. In *The Journal of the Acoustical Society of America*, 144(2), 897-907.

ABRAMSON, A.S., WHALEN, D.H. (2017). Voice Onset Time (VOT) at 50: theorical and practical issues in measuring voicing distinctions. In *Journal of Phonetics*, 63, 77-86.

AL-TAMIMI F., TARAWNEH, R., & HOWELL, P. (2021). Development of voice onset time in Arabic. In *Lingua*, 262, 103-117.

ALLEN, G.D. (1985). How the young French child avoids pre-voicing problem for word-initial voiced stops. In *Journal of Child Language*, 12(1), 37-46.

BARBIER, G., PERRIER, P., PAVAN, Y., TIEDE, M.K., GERBER, S., PERKELL, J.S. & MENARD, L. (2020). What anticipatory coarticulation in children tells us about speech motor control maturity. In *Plos one*, 15(4), e0231484.

BEST, C.T., GOLDSTEIN, L.M., NAM, H. & TYLER, M.D. (2016). Articulating what infants attune to in native speech. In *Ecological Psychology*, 28(4), 216–261.

BOERSMA, P., WEENINK, D. (2021). PRAAT: doing phonetics by computer. [Computer program], Version 6.1.40, retrieved 27 February 2021 from http://www.praat.org/.

BORTOLINI, U., ZMARICH, C., FIOR, R. & BONIFACIO, S. (1995). Word-initial voicing in the productions of stops in normal and preterm Italian infants. In *International Journal of Pediatric Otorhinolaryngology*, 31, 191-206.

BROWMAN, C.P., GOLDSTEIN, L. (1992). Articulatory Phonology: An Overview. In *Phonetica*, 49(3-4), 155-180.

CALLAN, D.E., KENT, R.D., GUENTHER, F.H. & VORPERIAN H.K. (2000). An auditory-feedback-based neural network model of speech production that is robust to developmental changes in the size and shape of the articulatory system. In *Journal of Speech, Language and Hearing Research*, 43, 721-736.

CASELLI, M.C., PASQUALETTI, P. & STEFANINI, S. (2007). *Parole e frasi nel Primo Vocabolario del Bambino*. Milano: FrancoAngeli.

CHO, T., WHALEN, D.H. & DOCHERTY G. (2019). Voice onset time and beyond: Exploring laryngeal contrast in 19 languages. In *Journal of Phonetics*, 72, 52-65.

COLAVOLPE, B. (2020). Analisi acustica della coarticolazione anticipatoria CV: uno studio sperimentale longitudinale su bambini di età compresa tra i 18 e 48 mesi. Tesi di Laurea Magistrale, Università degli Studi di Padova.

CYCHOSZ, M., MUNSON, B. & EDWARDS, J.R. (2021). Practice and experience predict coarticulation in child speech. In *Language Learning and Development*, 17(4), 1-31.

EILERS, E.R., OLLER, D.K. & BENITO-GARCIA. C.R. (1984). The acquisition of the voicing contrast in Spanish and English learning infants and children: a longitudinal study. In *Child Lang*, 11(2), 313-336.

ESPOSITO, A. (2002). On vowel height and consonantal voicing effects: data from Italian. In *Phonetica*, 59, 197-231.

Farnetani, E., Recasens, D. (2010). Coarticulation and connected speech processes. In Hardcastle, W.J., Laver J. & Gibbon F.E (Eds.), *The Handbook of Phonetic Sciences: Second Edition*. Hoboken (New Jersey): Blackwell Publishing, 316-352.

Fodor, J.A. (1983). *The Modularity of Mind*. Cambridge (Massachussets): MIT Press.

Gaiotto, M. (2020). L'acquisizione dell'opposizione di sonorità in bambini italiani: studio sperimentale sulle caratteristiche acustiche del Voice Onset Time. Tesi di Laurea Magistrale, Università degli Studi di Padova.

Gibson, T., Ohde, R.N. (2007). F2 Locus Equations: Phonetic descriptors of coarticulation in 17-to 22-months-old children. In *Journal of Speech, Language and Hearing Research*, 50, 97-108.

Goffman, L. (2015). Effects of language on motor processes in development. In Redford, M.A. (Ed.), *The Handbook of Speech Production*. Hoboken (New Jersey): Wiley Online Library, 555-577.

Grigos, M.I., Saxman, J.H. & Gordon, A.M. (2005). Speech motor development during acquisition of the voicing contrast. In *Journal of Speech in Language, and Hearing Research*, 48(4), 739-752.

Guenther, F.H. (1995). Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. In *Psychological Review*, 102(3), 594.

Guenther, F.H., Vladusich, T. (2012). A neural theory of speech acquisition and production. In *Journal of Neurolinguistics*, 25(5), 408-422.

Hardcastle, W.J., Hewlett, N. (Eds.) (2006). *Coarticulation: Theory, Data and Techniques*. Cambridge (UK): Cambridge University Press.

Harrington, J. (2010). Acoustic phonetics. In Hardcastle, W.J., Laver J. & Gibbon F.E (Eds.), *The Handbook of Phonetic Sciences: Second Edition*. Hoboken (New Jersey): Blackwell Publishing, 81-129.

Iskarous, K., Fowler, C.A. & Whalen, D.H. (2010). Locus equations are an acoustic expression of articulator synergy. In *The Journal of the Acoustical Society of America*, 128(4), 2021-2032.

Kent, R.D., Kim, Y. (2008). Acoustic analysis of speech. In Ball, M.J., Perkins, M.R., Miller, N. & Howard, S. (Eds.). *The Handbook of Clinical Linguistics*. Hoboken (New Jersey): Blackwell Publishing, 360-380.

Kunhert, B., Nolan, F. (1999). The origin of coarticulation. In Hardcastle, W.J., Hewlett, N. (Eds.). *Coarticulation: Theory, Data and Techniques*. Cambridge (UK): Cambridge University Press, 7-30.

Krull, D. (1989). Second formant locus patterns and consonant-vowel coarticulation in spontaneous speech. In *Phonetic Experimental Research at the Institute of Linguistics* (PERILUS), 7, University of Stockholm, Sweden, 66-70.

Laganaro, M. (2019). Phonetic encoding in utterance production: a review of open issues from 1989 to 2018. In *Language, Cognition and Neuroscience*, 34(9), 1193-1201.

Lindblom, B. (1963). On vowel reduction (Report no. 29). In *The Royal Institute of Technology, Speech Transmission Laboratory*. Stockholm, Sweden.

Lindblom, B., Sussman, H.M. (2012). Dissecting coarticulation: How locus equations happen. In *Journal of Phonetics*, 40(1), 1-19.

LING, D. (1976). *Speech and the Hearing-Impaired Child: Theory and Practice*. Washington DC: The Alexander Graham Bell Association for the Deaf.

MACKEN, M.A., BARTON, D. (1980a). The acquisition of the voicing contrast in English: a study of voice onset time in word-initial stop consonants. In *Journal of Child Language*, 7, 41-74.

MACKEN, M.A., BARTON, D. (1980b). The acquisition of the voicing contrast in Spanish: a phonetic and phonological study of word-initial stop consonants. In *Journal of Child Language,* 7(3), 433-458.

MCALLISTER BYUN, T., TESSIER, A.M. (2016). Motor influences on grammar in an emergentist model of phonology. In *Language and Linguistics Compass*, 10(9), 431-452.

MILDNER, V. (2018). Aspects of coarticulation. In GÓSY, M., GRÁCZI, T.E. (Eds.). *Challenges in Analysis and Processing of Spontaneous Speech*. Budapest: MTA Nyelvtudományi Intézet, 27-48.

NITTROUER, S., STUDDERT-KENNEDY, M. & NEELY, S.T. (1996). How children learn to organize their speech gestures: Further evidence from fricative-vowel syllables. In *Journal of Speech, Language, and Hearing Research*, 39(2), 379-389.

NOIRAY, A., ABAKAROVA, D., RUBERTUS, E., KRÜGER, S. & TIEDE, M. (2018). How do children organize their speech in the first years of life? Insight from ultrasound imaging. In *Journal of Speech, Language, and Hearing Research*, 61 (6), 1355-1368.

NOIRAY, A., WIELING, M., ABAKAROVA, D., RUBERTUS, E. & TIEDE, M. (2019a). Back from the future: Nonlinear anticipation in adults' and children's speech. In *Journal of Speech, Language, and Hearing Research*, 62(8S), 3033-3054.

NOIRAY, A., POPESCU, A., KILLMER, H., RUBERTUS, E., KRÜGER, S. & HINTERMEIER, L. (2019b). Spoken language development and the challenge of skill integration. In *Frontiers in Psychology*, 10, 2777.

OHALA, J.J. (2011). Accommodation to the aerodynamic voicing constraint and its phonological relevance. In LEE, W.S. & ZEE, E. (Eds.). *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS XVII): August 17-21, 2011.* City University of Hong Kong, 64-67.

PARRELL, B., HOUDE, J. (2019). Modeling the role of sensory feedback in speech motor control and learning. In *Journal of Speech, Language, and Hearing Research*, 62(8S), 2963-2985.

PETRACCO, A., ZMARICH, C. (2006). La quantificazione della coarticolazione nello sviluppo fonetico. In GIORDANI, V., BRUSEGHINI, V. & COSI, P. (Eds.), *Scienze Vocali e del Linguaggio. Metodologie di Valutazione e Risorse Linguistiche*. Torriana (RN): EDK Editore srl, 135-150.

REDFORD, M.A. (2019). Speech production from a developmental perspective. In *Journal of Speech, Language, and Hearing Research*, 62(8S), 2946-2962.

REDFORD, M.A., OH, G.E. (2017). The representation and execution of articulatory timing in first and second language acquisition. In *Journal of Phonetics*, 63, 127-138.

ROTHENBERG, M. (1968). The breath-stream dynamics of simple released-plosive production. *Bibliotheca Phonetica* 6, 1-117.

Salza, P.L. (1990). Phonetic transcription rules for text-to-speech synthesis of Italian. In *Phonetica*, 47(1-2), 66-83.

Saltzman, E., Munhall, K.G. (1989). A dynamical approach to gestural patterning in speech production. In *Ecol. Psychol.*, 1, 333-382.

Smith, A. (2010). Development of neural control of orofacial movements for speech. In Hardcastle, W.J., Laver J. & Gibbon F.E (Eds.), *The Handbook of Phonetic Sciences: Second Edition*. Hoboken (New Jersey): Blackwell Publishing, 251-296.

Schmitt, J.F., Meline, T.J. (1990). Subject descriptions, control groups, and research designs in published studies of language-impaired children. In *Journal of Communication Disorders*, 23(6), 365-382.

Solé, M.J. (2018). Articulatory adjustments in initial voiced stops in Spanish, French and English. In *Journal of Phonetics*, 66, 217-241.

Stone, M. (2012). Laboratory techniques investigating speech articulation. In Hardcastle, W.J., Laver J. & Gibbon F.E (Eds.), *The Handbook of Phonetic Sciences: Second Edition*. Hoboken (New Jersey): Blackwell Publishing, 9-38.

Sussman, H.M., Hoemeke, K.A. & McCaffrey, H.A. (1992). Locus equations as an index of coarticulation for place of articulation distinctions in children. In *Journal of Speech, Language, and Hearing Research*, 35(4), 769-781.

Sussman, H.M., Duder, C., Dalston, E. & Cacciatore, A. (1999). An acoustic analysis of the development of CV coarticulation. A case study. In *Journal of Speech, Language, and Hearing Research*, 42, 1080-1096.

Tabain, M. (2000). Coarticulation in CV syllables: a comparison of locus equation and EPG data. In *Journal of Phonetics*, 28, 137-159.

Vihman, M.M., Velleman, S.L. (1989). Phonological reorganization: A case study. In *Language and Speech*, 32(2), 149-170.

Westbury, J.R. (1983). Enlargement of the supraglottal cavity and its relation to stop consonant voicing. In *The Journal of the Acoustical Society of America*, 73(4), 1322-1336.

Zharkova, N. (2018). An ultrasound study of the development of lingual coarticulation during childhood. In *Phonetica*, 75 3), 245-271.

Zmarich, C., Bortone, E., Vayra, M. & Galatà, V. (2013). La coarticolazione e il VOT nello sviluppo fonetico: studio sperimentale su bambini dai 42 ai 47 mesi d'età. In Galatà, V. (Eds), *Multimodalità e Multilingualità: la Sfida più Avanzata della Comunicazione Orale.* Roma: Bulzoni Editore, 475-493.