DOI: 10.17469/O2109AISV000001

CINZIA AVESANI, SERENA BONIFACIO, GIULIA CALIGNANO, VALERIA D'ALOIA, FRANCESCO OLIVUCCI, MARIO VAYRA, CLAUDIO ZMARICH

The emergence of lexical and post-lexical prominence in Italian. A case study

Our study identifies the developmental trajectory of prominence at lexical and post-lexical levels. From very early in life infants are sensitive to lexical stress contrasts, but, due to very limited vocal capabilities, the production of stress contrasts only starts in the second year of age. We address the question of whether, when and how a child learns to differentiate lexical (stress) from post-lexical prominence (accent) by acoustically examining the spontaneous productions of one child from North-East Italy recorded every 3 months from 18 to 36 months of age. Our analysis is cast in the framework of the Autosegmental Metrical Theory of Intonation. Results show that during the child's prosodic development the duration of IP nuclear vowels increases linearly, the duration of unstressed vowels decreases linearly and the duration of stressed, prenuclear and ip nuclear vowels is progressively but non-linearly adjusted, consistent with the adult prosodic hierarchy.

Keywords: development of prosodic prominence, stress, accent, Italian, Autosegmental Metrical Theory of Intonation.

1. *Introduction*

Although studies on the acquisition of prosody are fewer in number than studies on the acquisition of other aspects of language, a large number of experimental studies have shown that children from a very early age are sensitive to prosody. Such sensitivity is rooted in prenatal experience with speech, which consists mainly of prosodic information, and it has been shown to already impact how newborns perceive speech and produce communicative sounds (Gervain, 2015).

Not only do newborns recognize their mother's voice and prefer it to other female voices (De Casper, Fifer, 1980), but they also recognize vowels heard prenatally (Moon, Lagercrantz & Kuhl, 2013) and their native language (Moon, Cooper & Fifer, 1993). Soon after birth, at only two days of age, newborns are sensitive to the alternation of stressed and unstressed syllables: Italian two-day newborns can discriminate between disyllabic and trisyllabic words differing in stress pattern regardless of consonant variations (màma vs. mamà, tàcala vs. tacàla), and seem to be able to categorize words based on their stress pattern (Sansavini, Bertoncini & Giovannelli, 1997).

The preference for the prosody of the maternal language becomes more specific at about 6 months. At this age, infants discriminate and prefer to listen to words of the maternal language rather than to those of a foreign language when the two languages differ in their global prosody, whereas this preference appears only at 9 months when the two languages mostly differ in their phonetic and phonotactic properties (Jusczyk, Friederici, Wessels, Svenkerud & Jusczyk, 1993).

At 4-9 months of age, infants discriminate between patterns of lexical stress, lexical pitch contours, and lexical tones if these patterns are contrastive in their ambient language, and they seem to consider all acoustic parameters (duration, pitch and intensity) to build their preferences (Bahatara, Boll-Avetisyan, Hohle & Nazzi, 2018). Infants of this age are sensitive to lexical stress contrasts and show a preference for the predominant stress pattern of the ambient language, which, in turn, partly determines the capacity of extracting word forms from fluent speech (Bahatara et al., 2018; Bion, Benavides-Varela & Nespor, M., 2011).

Moreover, 7- to 10-month-old infants prefer to listen to clauses of the maternal language in which artificial pauses are inserted between rather than within clauses Hirsh-Pasek, Kemler Nelson, Jusczyk, Cassidy, Druss & Kennedy, 1987) and 9-month-old infants prefer to listen to phrases of the maternal language in which artificial pauses are inserted between rather than within phrases (Jusczyk, Hirsch-Pasek, Kemler Nelson, Kennedy, Woodward & Piwoz, 1992).

It thus appears that, in their first months of life, infants pay attention to and discriminate the global prosody of speech and become attuned to the prosodic patterns of the maternal language.

To summarize, infants are sensitive to a variety of prosodic cues if they are linguistically relevant in the language environment. Several studies have proposed that they can use them to access lexical and morpho-syntactic information of their native language (e.g. the familiarity with the typical stress patterns of the ambient language enables infants to segment the continuous speech stream into words; by relying on the correlation between the position and the acoustic realization of phrase-level prominences and word order, infants can distinguish Head-Complement from Complement-Head languages). In this view, prosody has been interpreted as a flywheel for language acquisition, because the ability to perceive the native language prosodic patterns is considered the trigger to the acquisition of other language domains ("prosodic bootstrapping": for a review, see Gervain, Christophe & Mazuka 2021).

1.1 Stress

In production, young children seem to show more accurate control of intonation earlier than duration (Prieto, Estrella, Thorson & Vanrell 2012). Toward the end of the first year of life, children of intonational languages begin to produce linguo-specific intonational patterns, which develop rapidly during the second year of life as they learn to associate pragmatic meanings and prosodic features (Esteve-Gibèrt, Prieto, 2018).

The production of contrastive stress patterns occurs later. Several studies have shown that English children begin to produce contrasts of stressed and unstressed syllables at or after 18 months of age: at 18 months according to Kehoe, Stoel-

Gammon & Buder (1995), at 22 months according to Schwartz, Petinou, Goffman, Lazawski & Cartusciello (1996), at 24 months according to Pollock, Brammer & Hageman (1993). In these early stages, duration is the parameter that is used by children to produce accentual oppositions, while F0 and intensity will contribute to accent production only later (Pollock et al., 1993).

In those studies, however, no distinction is made as to whether the prominent syllables under consideration are stressed at a word or phrase level. That is, whether within the utterance they are not only lexically stressed but also associated with melodic configurations that give them a higher degree of prominence.

1.2 Italian

Studies on the acquisition of prosody in Italian children are mainly due to the work of D'Odorico and colleagues (D'Odorico, Carubbi, 2003; D'Odorico, Fasolo & Marchione, 2009; D'Odorico, Fasolo & Zanchi 2010), who studied the development of intonation at 24-36 months of age, adopting a global approach and relating it to children's syntactic and narrative ability.

An analysis of intonation cast within the AM theoretical framework is due to Zanchi, D'Imperio, Zampini & Fasolo (2016), who studied 3- to 4-year-old children. Their results indicate that 3-year-olds master nuclear pitch accents as adults, but they do not produce rising boundary tones in the same measure as adults.

Specifically on stress development, Arciuli and Colombo (2016) analyze the productions of 3- to 5-year-old children to delineate developmental trajectories in the ability to produce stressed and unstressed syllables in trisyllabic words with a trochaic (SW) or iambic (WS) beginning. The approach is a traditional phonetic one: stressed and unstressed syllables are analyzed according to their acoustic characteristics of duration, peak intensity, and F0. The analysis is centered on a relatively late age group (3-5 years), in a stage of development (well past that of the early vocabulary) in which possible discrepancies with adult stress targets may have been resolved. The study has the merit to be one of the first to address the production of stress in Italian children but, by running a group analysis, it fails to delineate individual developmental trajectories.

In a different vein, the study by Olivucci, Pasqualetto, Vayra & Zmarich (2016) delineates the developmental trajectories of lexical stress in the production of 5 young children from 21 to 27 months of age. The authors analyzed duration, intensity, spectral emphasis, F1 and F2 formant trajectories of stressed and unstressed vowels (/a, i, o, u/), and compared them to those of an adult control group. Their results indicate that toddlers differentiate unstressed and stressed vowels starting from 21 months of age, and that the acoustic parameters that most significantly cue such difference are duration and spectral emphasis. In two later studies (Olivucci, Vayra, Avesani & Zmarich, 2018, 2019), three different children have been analyzed, extending the age window to comprise the children's production from 18 to 42 months. Stressed and unstressed vowels were distinguished based on their position in one and multi-word utterances. Results showed that lexical stress distinctions appear starting from 18 months of age, and that stressed vowels were significantly more prominent than unstressed ones in single-word utterances and in the final words of multi-word utterances, while they were not distinguished in words occurring within the utterance. Duration confirms to be the most reliable correlate of lexical stress, and the developmental trajectory shows that the difference between unstressed and stressed vowels increases with age. Interestingly, the increasing difference is not only due to an increase in the duration of stressed vowels but also to a decrease in the duration of unstressed ones.

2. *Aims and predictions*

The present study builds on the works by Olivucci and colleagues, and extends its focus on the acquisition of post-lexical prominence in Italian. We have two aims: by enlarging the sample of analysis, we would like to confirm the developmental trajectory of lexical stress. Our second aim is to go further and address the acquisition of post-lexical prominence. That is, we would like to understand when, in individual development, stressed syllables start acquiring relatively different degrees of prominence once the words to which they belong form part of a structured sentence. Triggered by the observations in Olivucci et al. (2018, 2019) that stress differences emerge in the strongest prosodic position in multi-word utterances (i.e. only in the utterance-final word), we aim at uncovering the emergence of the prosodic structuring of the sentence in the child's speech by looking at the acoustic properties of the lexically stressed syllables in one, two and multi-word utterances.

We adopt the framework of Prosodic Phonology (Selkirk, 1984; Nespor, Vogel, 2007[2]) and the Autosegmental Metrical Theory of Intonation (Beckman, Pierrehumbert, 1986; Ladd, 2008[2]). Prosody constitutes the organizational structure of spoken language, and Prosodic Phonology, integrated into AM theory, considers it to consist of several hierarchically ordered metrical constituents that, above the Syllable and the Foot, scholars agree to be the Prosodic Word (W), the Phonological or Intermediate Phrase (respectively, ϕ or ip, depending on the version of the theory) which consist in one or more Prosodic Words, and the Intonational Phrase (IP), which consists of one or more Phonological Phrases.

Each constituent is endowed with a head or nucleus (the strongest element in the constituent). In Italian, the head of the Word is the syllable designated in the lexicon to carry the lexical stress (**S**tressed), while at the higher levels of hierarchically ordered constituents the head is (the stressed syllable of) the last word in each constituent. These postlexically metrically strong syllables are designated to be the exponents of phrasal prominence and are associated with linguo-specific tonal configurations that give rise to pitch accents. We code the head syllable of an IP, or *nuclear* in the IP, as **N** and the head syllable of ip, or nuclear in ip, as **Nphp**. Metrically weaker syllables in IPs and ips can be associated with a pitch accent even if they are not the head of their constituent, and by virtue of this association, they acquire greater prominence than a lexically stressed syllable that does not have a pitch accent (e.g. Beckman, 1996). These syllables, stronger than the lexically stressed ones and weaker than the nuclearly accented head ones, are named prenuclear (**P**).

In Fig 1, the head syllables of each constituent (W, ip and IP) are indicated by a star.

Figure1 - *Schematic representation of the higher levels of prosodic structure: Prosodic Word (W), Phonological/Intermediate Phrase (ip) and Intonational Phrase (IP). Stars represent the head elements in each constituent*

Based on their position in the hierarchy of prosodic constituents, metrically strong syllables carry a progressively higher degree of structural prominence, according to the following progression:

IP Nuclear > ip Nuclear > Prenuclear > Stressed (> Unstressed)

In adult speech, the hierarchy of word- and phrase-level prominence is substantiated by significant differences in a set of acoustic and articulatory parameters. In Italian, at the lexical level, stressed vowels show longer acoustic durations, more intensity, and more spectral emphasis than unstressed vowels. Moreover, they show less centralization in F1-F2 formant space: low stressed vowels show a higher degree of opening, with a higher F1, and high vowels a higher degree of fronting (with a higher F2) than unstressed vowels, and less C-V coarticulation (Farnetani, Kori, 1982, Vayra, Fowler, 1987; Vayra, 1991; Savy, Cutugno, 1996; Vayra, Avesani & Fowler, 1999; Tamburini, 2009).

Articulatorily, palatographic data show that stressed vowels are more open than unstressed ones and that tongue contacts decrease for unstressed high vowels and increases for unstressed low vowels (Farnetani, Faber, 1992). Kinematic data show that the jaw is lower in stressed low vowels (Vayra, Fowler, 1992; Magno Caldognetto, Vagges & Zmarich, 1995) and their opening gesture has a longer duration, higher peak velocity and more displacement as compared to unstressed vowels (Avesani, Vayra & Zmarich, 2009; Zmarich, Avesani, 2015).

At the postlexical level, sentence-level prominence manifests acoustically as enhanced acoustic parameters. Accented low vowels which are IP nuclear are longer, with a more extreme F1 trajectory, and with more energy in the high-frequency bands (lower values of spectral balance, leading to more prominence) than lexically stressed vowels (Avesani, Vayra, 2013). IP nuclear vowels are also significantly longer and with lower values of spectral tilt (more energy in the high frequencies of the spectrum, leading to more prominence) than prenuclearly accented vowels. However, the latter ones are not significantly longer than stressed vowels while there is evidence of their having more spectral emphasis in the high energy bands (less spectral balance) than stressed ones (Bocci, Avesani, 2011). Articulatorily, IP nuclear vowels show opening gestures that are longer, with higher peak velocity and more displacement than stressed vowels.

Summarizing, the literature on Italian lexical and postlexical prominence shows that postlexical strong vowels are phonetically realized with acoustic and articulatory parameter values that directly correlate with their position in the prosodic hierarchy: accented IP nuclear vowels are more prominent than prenuclearly accented vowels than lexically stressed ones.

In the present study, we build on the results offered by the literature on the infants' perception of phrasal prosody (Gervain et al., 2021; Chen, Esteve-Gibert, Prieto & Redford, 2021) to investigate when a child begins to produce utterances which are prosodically phrased, and when the internal structure of the prosodic constituents in terms of relatively strong and weak elements will emerge. As reported by Gervain et al. (2021), infants perceive intonational phrase boundaries from 5 months of age (Hirsh-Pasek et al., 1987) and intermediate phrase boundaries from 9 months of age (Gerken, Jusczyk & Mandel 1994; Shukla, Wite & Aslin, 2011). At 20 months of age, French and English toddlers are able to get the correct interpretation of ambiguous sentences by relying on their prosody (de Carvalho, Lidz, Tieu, Bleam & Christophe, 2016; de Carvalho, Dautriche, Lin & Christophe, 2017), and at 18 months of age are able to exploit the syntactic structure accessed through phrasal prosody to guess the meaning of a novel word (de Carvalho, He, Lidz & Christophe, 2015).

We expect that, in production, prosodic phrasing will consistently appear when the child will produce not only utterances formed by the simple juxtaposition of two words but utterances effectively endowed with argument structure, progressively longer and syntactically more complex. At this stage, we expect utterances will be phrased into prosodic constituents in a hierarchical relationship to each other. Along with the emergence of prosodic phrasing, we also expect that the head syllables of each hierarchically ordered prosodic constituent will be marked by a degree of prominence coherent with their position in the prosodic hierarchy and that the preceding words can be optionally endowed with relatively weaker prenuclear accents.

We predict that, in the path of linguistic and prosodic development, the child will attain to produce lexically stressed, prenuclear and nuclear vowels which will be differentiated for duration. Moreover, we want to discover when this progressive differentiation would occur and how. One hypothesis is that the postlexically strong vowels will lengthen linearly as the linguistic and prosodic development procede. But speech development is hardly a linear process, and therefore an alternative hypothesis is that in the acquisition of prosodic prominence the adult target could be attained nonlinearly.

3. *Method*

3.1 Corpus and recordings

Data analyzed in this study are part of a corpus collected by Serena Bonifacio at Trieste, from 2007 to 2009. The corpus includes 10 Italian children, 4 males and 6 females, recorded every three months from 18 to 48 months of age. Parents compiled the MacArthur CDI surveys, in which they reported the most frequent words produced by their child at each developmental stage (Caselli, Casadio, 1995) and filled out a questionnaire aimed at verifying their normal psycho-physical and linguistic development. When they were 18-months-old, the children underwent audiologic screening to exclude the presence of hearing impairments (Ling, 1976).

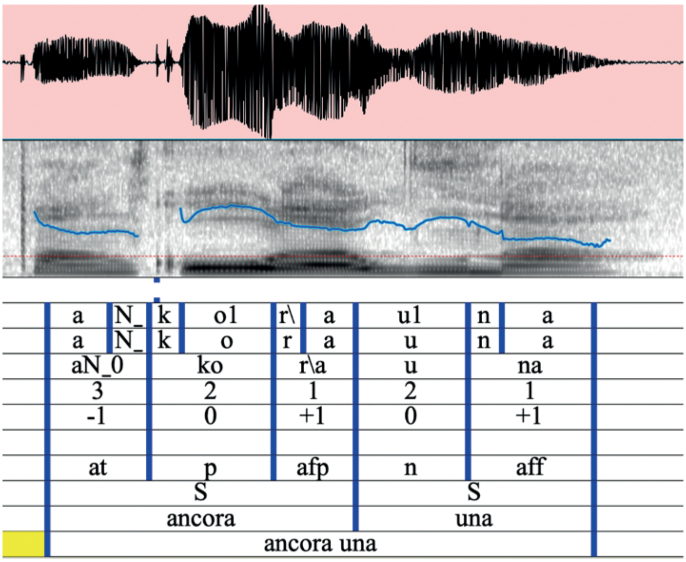
In the semi-structured recording sessions (Schmitt, Meline, 1990), the child interacts with the clinician in front of a set of toys. These objects were chosen based on the list of words compiled by the parents on the MacArthur CDI and were presented on the basis of decreasing frequency of the semantic categories that resulted from CDI. Besides the most common words and for all developmental stages, children were invited to repeat (five times each at least) 12 minimal pseudo-word, initially stressed, contrasting labial, dental and velar voiced and voiceless stops: 'papa', 'baba', 'pipi', 'bibi', 'tata', 'dada', 'titi', 'didi', 'kaka', 'gaga', 'kiki', 'gigi'. In the last sessions, children were engaged in more structured conversations about their interests (holidays, movies, cartoons, family and so on). Each recording session lasted on average about an hour. The speech samples collected at 18 and 21 months were considered valid and representative of the child's linguistic abilities of those developmental stages only if the number of lexical forms produced represented at least 50% of the words in the lexical list compiled by the parent. Speech samples were recorded with an Edirol R-09 digital recorder, at 16-bit sample size and 44.1 kHz sampling frequency.

3.2 Data selection, coding, analysis

For the present study, we selected one of the most talkative children in the corpus (BS) and analyzed his productions at the stages of 18, 21, 24, 27, 30, 33 and 36 months. We considered bi-, tri- and quadrisyllabic words, with stress on the penultimate and ante-penultimate syllable, and uttered spontaneously (we did not consider words produced by repetition). The target words occurred in one-, two-, three- and multi-word utterances. Of the total number of syllabic occurrences (3454), the analysis focuses on open syllables, with C = voiced and voiceless stops, fricatives, nasals, liquids, affricates, and V = all Italian vowels (1157 syllables). Due to the typical distortion processes operated by children at the ages considered here, as well as the connected speech style of several utterances, the quality of some vowels does not completely match the quality of vowels in the adult phonological system.

We phonetically and prosodically annotated in *Praat* (Boersma, Weenink, 2021) the entire utterance in which the target words occur. An example is given in Fig. 1. We created 11 *tiers* in which we indicated, in order: the burst of the stop consonant; the transcription (SAMPA) of the actual phone produced by the child; the adult target; the actual syllable; the distance of the syllable from the word end (1 = end); the distance of the syllable from the lexical stress (0 = stress); possible observations; the accentual status of the syllable: unstressed (**U**), stressed (**S**), prenuclearly accented (**P**), nuclearly accented in intermediate/phonological phrases (**Nphp**), and nuclearly accented in intonational phrases (**N**). Besides, we coded as **postF** the prominent syllable(s) which eventually follow the sentence focus. In Fig. 1, the prosodic labels also indicate the position of the syllable in the word and in the utterance (afp = unstressed and word-final; aff = unstressed and utterance-final). In the final *tiers* we coded, respectively, the modality of production (S = spontaneous), the adult target word and the adult target sentence (orthographic transcription). In order to assess the prominence status of the target syllables, 2 of the authors have annotated the whole corpus. In case of disagreement, the final decision on the prominence status of a syllable was reached by discussing each singular case.

Figure 2 - *An example of phonetic and prosodic annotation with Praat. Waveform, spectrogram and F0 of the utterance "ancora una", transcribed and coded in 11 tiers (see main text for details)*



For each of the 1149 vowels eligible for analysis we automatically measured: 1) duration, 2) F1 and F2 at vowel midpoint; 3a) spectral emphasis, which was calculated as spectral tilt (H1* – A3*) according to Jessen, Marasek (1997), with some adjustments to make calculations suitable for children; and 3b) as spectral balance (difference in dB between four contiguous frequency bands; Sluijter, van Heuven, 1996). Bands values are set based on the analysis of F0 and formant values for each developmental stage); 4) F1 and F2 trajectories (ten equidistant points over the vowel duration).

In the present paper, we analyze only the results of vowel duration, as the previous study by Olivucci et al. (2016) indicated that duration is the most robust cue to prominence among other acoustic cues (formant trajectories, spectral tilt and spectral balance). The data span across 18 months of child production, starting from the 18-month-recording in which the child's speech includes almost exclusively one-word utterances, to the 36-month-recording in which the child produces complex sentences. To factor out differences in speech rate, for each developmental stage we normalized 1) the duration of the vowels on the duration of the syllables in which they occur, and 2) the duration of each vowel on the duration of the nuclearly accented one. The correlation between normalized and row durations across the different developmental stages is very high in both types of normalizations ($R = 0.89$, $p < 0.0001$). Therefore, results will be reported in terms of raw duration values (ms).

4. *Results*

In what follows we offer a brief description of the child's prosodic development and present the results of vowel duration, separately for each recording session. Vowel duration for unstressed (**U**), lexically stressed (**S**), prenuclearly accented (**P**), nuclearly accented in intermediate/phonological phrase (**Nphp**), and nuclearly accented in intonational phrases (**N**), is presented in Fig. 3 in separate box plots per stage.

18 months. At this stage, the child produces all one-word utterances in response to the clinician's solicitations (e.g.: *Come si chiama questo? Cosa si fa con questo?*). Only one two-word utterance with reduplicated words (*baba baba*) is attested in the recording session in which the child interacts with the clinician for about an hour. Adult target words were disyllables with penultimate stress and trisyllables with stress on any syllable (eg: *fòrbice, poltróna, biberòn*).

Trisyllables as uttered by the child undergo a deletion process by which unstressed syllables (Weak) preceding the stressed ones (Strong) are omitted. Accordingly, SWW trisyllables retain the target word structure (*fòr.bi.ce* > *bò.ci.ci*), WSW trisyllables loose the pre-stress weak syllable (*pol.trò.na* > *tòn.na*), WWS oxitons are reduced to the final strong syllable (*bi.be.ròn* > *òn*).

Adult target bisyllables are trochees (SW, eg. *dàdo, càpra, tòpo*), but not all of them are uttered by the child as such: in 22% of the cases (18 out of 82) the stress pattern is reversed, from SW to WS: *tàta* > *tatà*. In 33% of cases (31 out of 82), both syllables are perceived by the transcribers of equal prominence (*dà.do* > *dà.dò).* In the remaining cases (33 out of 82) the child produces the word according to the stress structure of the adult target, but these cases amount to only 40% of the total.

Due to the process of weak syllable deletion and our choice to exclude word- and sentence-final vowels from the computation, we were left with very few cases of unstressed vowels corresponding to the adult target. In order to arrive at a more balanced set, we included the word-initial weak syllables of words uttered with stress inversion (ta.tà). As for stressed vowels, we considered them both as the metrical heads of the lexical word (i.e. stressed) and the metrical heads of the intonational phrase that wraps the one-word utterance (i.e. accented). This operational choice will need to be more thoroughly evaluated in future research.

The analysis of a total of 39 syllables shows that the average duration of unstressed vowels (n = 8, M = 118.25 ms, SD = 36.79 ms) is shorter than nuclearly accented ones (n = 30, M = 193.37 ms, SD = 47.82 ms). The median difference is 71.5 ms.

21 months. At the developental stage represented in the 21-month recording, two- and multi-word utterances appear. The child produces cases of Det + N, N + Adj, N + N, such as *la spazzola, la tata, la mucca, gallina bella, bibi questo, babbo natale,* as well as Prepositional Phrases, Adverbial Phrases and complex NPs such as the following: *sotto il fungo, come la tata, ghighi e la cova, gaga con la voce, scarpe de mamma*.

There are still some cases in which the stress structure of a word is different from the adult target: disyllabic paroxytone words that become oxytones (1 case), bisyllabic words in which both syllables have the same degree of perceived prominence (4 cases).

Along with nominal structures, the child begins to produce utterances with argument structure: two-word utterances in which the argument is in canonical position (object in a postverbal position such as *chiama la zebra, fa pipì, chiudi gli occhi*), and multi-word utterances with non-canonical order i.e. utterances in which at least one element appears in non-canonical order such as Subject in post-verbal position, and/or Object in pre-verbal position, (D'Odorico, Fasolo, Marchione, 2009). Examples are: *mamma palla damme; biberon fa la mamma; limone ti do questo; cucchiaio mano a Stefano questo*.

With the appearance of the phrasal structure, the structuring of the utterance into prosodic constituents also appears, as well as the modulation of postlexical strong syllables in different degrees of prominence. In (1) the utterance is phrased in two intonational phrases separated by a pause in which each IP-final word is nuclearly accented; in (2) the multi-word utterance is not internally phrased, but each lexical word is endowed with a different level of prominence: stress on the utterance initial word *cucchiaio*, prenuclear accent on *mano* and *Stefano*, nuclear accent on *questo*.

(1) [[mamma **pal**la]$_{IP}$ [**dam**me]$_{IP}$

(2) [cuc**chia**io$_{S}$ **ma**no$_{P}$ a **Ste**fano$_{P}$ **que**sto$_{N}$]$_{IP}$

Acoustically, vowels in metrically strong positions (here nuclearly accented in intonational phrase) are longer than the other ones (untressed, stressed and prenuclearly accented). Average values of the 102 eligible tokens are: U (n = 44, M = 140.68, SD = 49.16), S (n = 10, M = 221.30, SD = 54.03), P (n = 10, M = 184.30, SD = 68.64), N (n = 37, M = 234.27, SD = 47.00), With respect to the

18 month stage, the difference between U (M = 140.68, SD = 49.16) and N (M = 234.27, SD = 47.00) increases (median difference = 99 ms).

24 months. Along with uttering two-word utterances (Det+N), at 24 months the child produces simple and coordinated Prepositional Phrases (*con il cappello, della scatola e della bici*), short questions and exclamatives (*Che si chiama Chicchi? Che bon!*), VO (*racconta una storia*) and SVO sentences (*la tata fa pipì*), and sentences with direct and indirect arguments (*mette il vino nella bottiglia*).

Prosodically, at lexical level words with stress inversion are no longer present, but there remain 3 cases of word-final unstressed vowels that are perceptually as prominent as the word-initial stressed ones (e.g p**à**pp**a**).

At postlexical level, two clear cases of phrase accents appear. In the example (3), the stressed syllabe [gi] is the head of an intermediate/phonological phrase and is associated with an H*(+L) pitch accent:

(3) [[chiama **gh**ighi$_{\text{Nphp}}$]$_{\text{ip}}$ [con le **scar**pe$_{\text{N}}$] $_{\text{ip}}$]$_{\text{IP}}$

For the first time in his development the child produces a sentence with information focus on a left-dislocated object (*pipì)* (4b) in response to the question posed by the clinician in (4a)

(4a) Cosa fa? Ti ricordi? Come si dice?

(4b) [[la pi**pì**]$_{\text{ip}}$ [fa] $_{\text{ip}}$]$_{\text{IP}}$

It could be argued that in (4b) the non-canonical order results from a not yet fully developed competence of argument structure, rather than from an option selected by the child to mark the object pragmatically. In a study on Italian by D'Odorico et al (2009), for example, utterances with non-canonical order are attested, but "the distinction between non-canonical and canonical orders is hardly marked at all from a prosodic point of view" (ibid: 326). On the contrary, (4b) appears to be a true case of focus-background partition: first, the pitch countour conforms to the adult norm, as the nuclear syllable "pì" is marked by an L+H* pitch accent, and the post-focal pitch contour is lowered as in the adult language; second, in the same recording session, the child answers the question in (4) with a canonical VO sentence *fa pipì.* Therefore he shows to be able to pragmatically and prosodically master utterances with focus *in-situ* and *out-of-situ*.

Post-focal vowels have not been included in the acoustical analysis that counts 134 eligible tokens. Vowels in strong metrical positions N (n = 51, M = 205.49, SD = 42.95) and Nphp (n = 2, M = 215.0, SD = 9.90) are longer than vowels in metrically weaker positions U (n = 63, M = 105.59, SD = 36.69), S (n = 11, M = 132.73, SD = 53.60), P (n = 7, M = 142.14, SD = 41.16), while no clear differentiation in the duration of the vowels within each group is observable. The median difference between N and U is increasing with respect to the previous stages: 101 ms.

27 months. At 27 months the child interacts more with the clinician and is very talkative. The total number of eligible CV syllables in this recording session is 249. There are no longer words that do not prosodically conform to the adult target: all words are produced with the expected stress structure. At syntactic level the child produces utterances with progressive verb forms (*sta mettendo in testa l'ombrello; sta stirando la camicia*) and sentences with subordinate clauses (*sta cucinando per andare a casa; bambino per andare in bicicletta*). Utterances with a sentence-initial information focus in response to a question posed by the clinician are more frequent:

(5a) C: questa, cosè questa?
(5b) BS: [**tor**ta]$_{FOC}$ [sè]$_{Backgr}$

(6a) C: e dove vola via?
(6b) BS: [sotto la **mac**china]$_{FOC}$ [va]$_{Background}$

(7) [tutte (le) **bam**bole]$_{FOC}$ [voglio]$_{Background}$

And a case of in-situ contrastive focus occurs, which is marked by a L+H* pitch accent with a large pitch span:

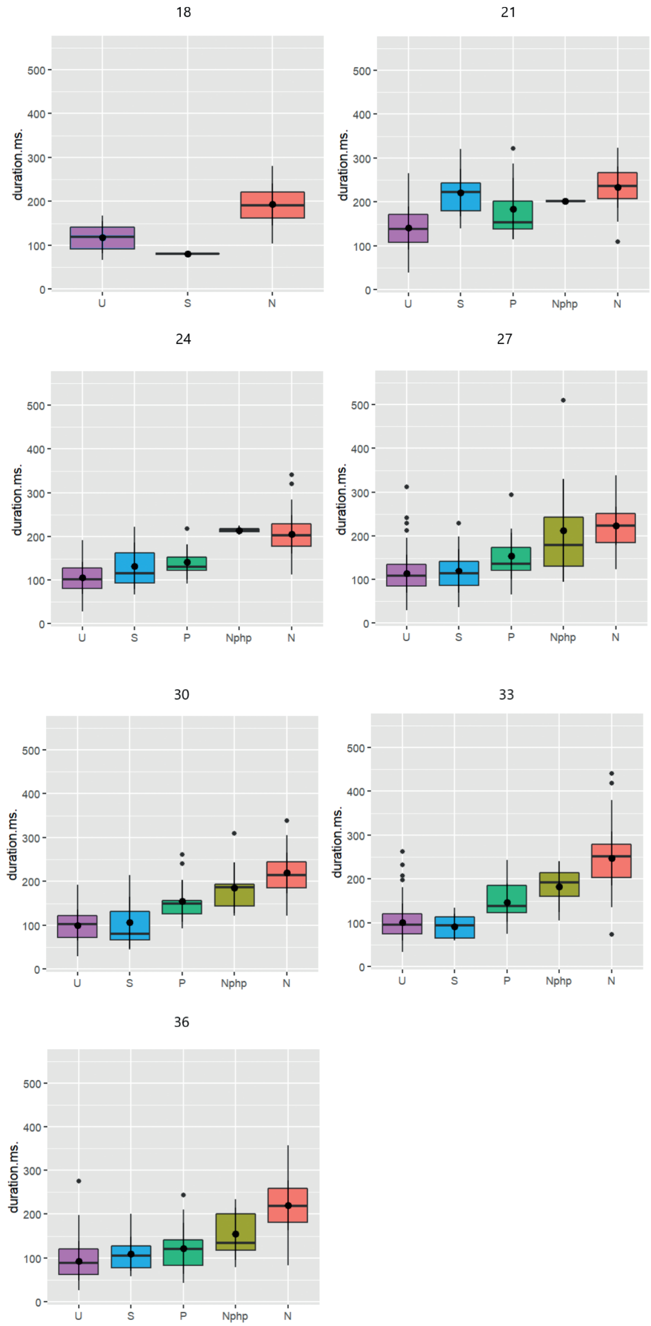
(10) tu sai suonare il pianoforte?
(10b) Son [PIU' GRANDE]$_{ContrFoc}$ [sa]?

A higher number of nuclear accents in intermediate phrases is observed (10 cases), related to a more frequent use of prosodic phrasing in the utterance. Utterances with coordinated NPs foster the placement of a prosodic boundary after the first NP. Moreover, the child's new syntactic achievement, i.e. the production of sentences composed of a main and a subordinate clause, fosters the presence of a prosodic boundary at the major syntactic boundary. In (11), for example, a boundary tone is placed between the two clauses "sta cucinando" and "per andare a casa". Each clause is consequently wrapped in an intermediate phrase: "nan" is the head of the first ip and constitues a case of ip nuclear accent (Nphp); "ca" is the head of the second ip but at the same time it is the head of the IP that wraps the whole utterance and therefore constitues a case of IP nuclear accent (N).

(11) [[sta cuci**nan**do$_{Nphp}$]$_{ip}$ [per andare a **ca**sa$_{N}$]$_{ip}$]$_{IP}$

Vowels' duration begins to be differentiated postlexically: while untressed (M = 113.54, SD = 43.70) and stressed vowels (M = 120.25, SD = 50.44) are of equal duration, a progression in duration is observable from prenuclearly accented (M = 154.65, SD = 53.47) to IP-nuclearly accented (M = 212.60, SD = 118.35). The median distance from U to N vowels rises to 116 ms.

Figure 3 - *Box-plot of the duration of the syllable in milliseconds (split for the five vowels of interest U, S, P, Nphp and N) according to Age (18, 21, 24, 27, 30, 33 and 36 months). Lower and upper box boundaries represent the 25th and 75th percentile, respectively; the line inside box represents the median; lower and upper error lines represent the 10th and 90th percentile, respectively; filled circles represent data falling outside the 10th and 90th percentile*



30 months. The child continues with his production of complex NPs, VO and SVO declarative sentences, sentences with a main and a subordinate clause, questions and exclamatives. Examples of sentence-initial focus attested in this recording session are the following:

(13) [[una **mac**china]$_{Foc}$ [voglio]]

(14a) C: Senti, conosci il Libro della Giungla?
(14b) BS: [[Ba**lù**]$_{Foc}$ [sè]]$_{IP}$ [[Ba**lù**]]$_{IP}$

The eligible number of vowel tokens in this recording session is less than in the previous one: 197. As for their duration, unstressed (n = 91, M = 100.63, SD = 35.32) and stressed vowels (n = 11, M = 106.45, SD = 58.64) are still undifferentiated, while the progressively higher average duration of P (n = 13, M = 155.46, SD = 47.44), Nphp (n = 8, M = 185.62, SD = 57.42) and N (n = 74, M = 219.31, SD = 47.70) does not show any appreciable difference with respect to the durations of the 27-month stage. Also the median difference between U and N vowels shows no appreciable variation (111.5 ms).

Generally, in the period ranging from 27 to 30 months of age, the child does not show any significant variation in his linguistic and prosodic development.

33 months. In the recording session at 33 months of age, the child talks more than in the preceding session, and the number of eligible vowels is 249. The type of utterances he produces is in line with those of the 30 and 27 months: noun phrases, sentences with direct and indirect arguments, exclamatives and sentences with sentence-initial information focus.

The average duration of unstressed (n = 120, M = 101.77, SD = 41.71) and stressed vowels (n = 9, M = 92.22, SD = 28.81) is still undifferentiated, but the duration of vowels in IP nuclear position (n = 95, M = 247.32, SD = 61.26) is longer than in the 30 months recording. The median difference between the duration of U and N vowels increases at 156 ms.

36 months. At 36 months the exchange with the clinician is richer and the number of sentences increases. In (15) and (16) we report two examples, with the indication of the focal structure, the prosodic phrasing and the tonal structure of each child's utterance. Pitch accents (H*, L+H*) refer to the nuclear accent (in bold) and boundary tones (H-, L-, LL%) refer to the following prosodic boundary.

(15) C: cosa hai fatto ieri?
BS: [mi hanno fatto la pun**tu**ra]$_{Broad\ Focus}$ H* H- [tutta tutta]
C: la puntura? Che puntura te ga fatto?
BS: quella [della **me**dica]$_{Narrow\ Information\ Focus}$ L+H* LL%
C: quella...??
BS: quella [della **me**dica] $_{Narrow\ Information\ Focus}$ H*+L LL%
C: medica?? E la puntura per cosa servi?
BS: [la **me**dica] $_{Narrow\ Information\ Focus}$ L+H* L- [quella quella che ho avuto dopo]
C: ahh! La vaccinazione te ga fatto!

(16) C: cosa è questo qua?
BS: [un **to**po] $_{\text{Broad Focus}}$ H* LL%
C: e chi va a mangiare i topi?
BS: [i **gat**ti] $_{\text{Information Focus}}$ L+H* L- [va a mangià i **to**pi] $_{\text{Background}}$ L* LL%

The number of eligible vowels is 187. At this age, it appears that the child has acquired the control of vowels' duration according to their role in the hierarchy of prosodic domains. As in adult's speech, BS shows a steady progression in vowels' duration from weak positions (Unstressed, n = 72, M = 93.69, SD = 44.30) to strong positions in Words (n = 22, M = 109.73, SD = 38.20), Intermediate Phrases (n = 7, M = 122.58, SD = 58.1) and Intonational Phrases (n = 74, M = 154.71, SD = 50.09).

4.1 Statistical modelling

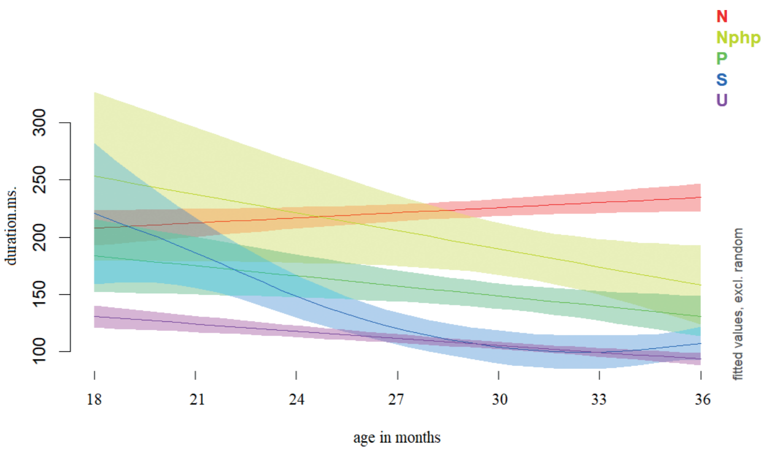
Data were analyzed with the R software (R Core Team, 2020) using generalized additive mixed-effects models (GAMMs) with the *mgcv package* (version 1.8-38, Wood, Wood, 2015). GAMMs approaches allow to model data for random and fixed effects as a function of time. Of note, those methods fit with multiple factors and unbalanced data sets, which are the norm rather than the exception in early child research, and stabilize the estimation of the parameters under investigation (Bates, 2010). In addition, those approaches are indicated to analyze longitudinal n-of-1 or single-subject study (idiographic) in which the target population consists of some larger set of time periods within a person's life (i.e., population-of-one studies (Daza, 2019), as in the case of the present study. Moreover, the mixed effect approach to statistical modelling allows to select a theoretical family distribution that better fits the empirical distribution of residuals. This in turn reduces the violation of the statistical assumptions, usually encountered with traditional approaches e.g. ANOVA, and the likelihood of obtaining false positive results (Boisgontier, Cheval, 2016). Accordingly, in the Statistical Appendix Fig. a and b, respectively, show the good fit of the Gamma (vs Lognormal) theoretical distribution family with the empirical distribution.

To find the best approximation to the true model, we followed a model comparison approach with AIC (Akaike Information Criterion) and AIC weight as indexes of goodness of fit. The AIC and AIC weight compare all the models at once and give information on a model's relative evidence (i.e., likelihood and parsimony), so that the model with the lowest AIC and the highest AIC weight is to be preferred (Wagenmakers, Farrell, 2004). We started from the simplest model with only random factors and proceeded by adding predictors. Specifically, we tested two models: i) the null model that uniquely includes smoothing components throughout Age in months and a random slope for Vowel type (U, S, P, Nphp and N); ii) the model with smoothing components throughout Age in months and a random slope for vowel type and with the interaction between Age in months and the type of Vowels i.e. U, S, P, Nphp and N.

The model comparison indicates that the last model with the interaction Age x Vowels better approximates the observed duration in milliseconds (dAIC = 0, AIC weight = 1, R-sq.(adj) = -0.123, Deviance explained = 45.2%) compared with the null model (dAIC = 169.6, AIC weight = 0, R-sq.(adj) = 0.562, Deviance explained = 51.7%). Following the best practice in order to correctly interprete GAMMs results (Van Rij et al., 2019), we visually inspected the model estimates between vowels. As a sanity check, Fig. c in the Statistical Appendix shows the autocorrelation plot of the selected model meeting the statistical assumption of a close to zero autocorrelation between the regressed variables. Fig. 4 shows how Age in months substantially interacts with the selected vowel tokens, i.e. U, S, P, Nphp and N by predicting statistically significant differences in duration.

On the one hand, the estimated smoothed effect indicates that U vowels predicted shorter duration compared to all the others since 18 months of age. On the other hand, N vowels predicted a significant increase in duration from 18 to 36 months of age compared to all the other vowels. The Nphp, P, and S vowels respectively, show a significant difference from 24 months and stay statistically different till 36 months of age. In addition, the Nphp, P and S vowels predicted an overall decrease in duration across time. Finally, the substantial effect indicating shorter duration for U vs S syllables disappeared around 27 months of age.

Figure 4 - *Partial effects (fixed effects only) of the initial GAMM showing the nonlinear regression lines for each of the five syllables (U, S, P Nphp, and N) with pointwise 95% confidence intervals*



5. *Discussion and conluding remarks*

Our aims at the start of this study were rather exploratory, being this the first study on the production of lexical and postlexical prominence in Italian children. Based on the results of our previous studies, we expected that the child we analyzed would distinguish between unstressed and stressed vowels since the earliest recording stage, i.e. 18 months, and that along his linguistic and prosodic development vowels with different prominence degrees would emerge according to the acquisition of the utterance prosodic and syntactic organization.

At the 18-month stage, the child has not yet mastered the production of lexical stress: cases of stress inversion and equal prominence in disyllables, and weak syllable deletion in trisyllables occur. Weak syllable deletion in multisyllabic words is a well-known phonological process in acquisition, attested in many languages. Within a cognitive view of language acquisition, it has been formalized by proposing the existence of a shape constraint such that children's word productions conform to a consistent size and rhythmic pattern. Gerken (1994) formalized it as the output of the application of a metrical constraint known as "trochaic bias": children would align a trochaic (SW) metrical constraint at the beginning of an intended word and weak syllables that do not fit the template are omitted. In a Natural Phonology approach, weak syllable deletion is one of the natural processes which are systematically applied in speech production until children learn to suppress them. Within an alternative usage-based approach, cast in an emergentist framework which does not assume universal constraints on production (and perception), Vihman proposes that the acquisition of prosodic structure "appeals to learning based on both initial perceptual biases of possible evolutionary origin (...) and infants' experience, along with neurophysiological maturation, of vocal production practice (...)" (Vihman, 2018:185). Weak syllable deletion appears to be suppressed by 21 months of age.

At 18 months the child is still in the stage of one-word utterances. In a phrasal perspective, the only possible rythmic difference within one-word utterances is between unstressed and nuclearly accented vowels. Such opposition is clearly characterized by a durational difference: nuclear vowels are significantly longer than unstressed ones.

With the emergence of argument structure and multi-word utterances at 21 months of age, prosodic phrasing appears, and syllables endowed with stress, prenuclear accents, and IP-nuclear accents occur. However, the duration of the prominent vowels, although in the expected direction, is not significantly different between S, P and N. The only significant difference is still that between unstressed and nuclearly accented vowels.

A pivotal stage for this child prosodic development is represented by the 24-month recording: the model predicts that U, S, P, Nphp and N vowels are significantly different, but nuclear vowels in intonational phrases are shorter than in intermediate phrases.

At 27-months of age the prominence hierarchy appears to be in place, as in the adult speech, with increasing duration from U to N vowels.

But the system is not stable yet: at 30- and 33-month recordings, while the higher levels of prominence remain statistically different and IP-nuclear vowels increase in duration, unstressed and stressed vowels loose their distinction and become durationally indifferentiated. This in in line with our expectations that the process of stress/accent acquistion is nonlinear.

Finally, in the 36-month recording the child's production is adult-like: the distribution of prominences within the prosodic constituents is what is expected in adult speech, with a linear increase in the duration from unstressed, to stressed, prenuclear, ip-nuclear to IP-nuclear vowels.

As for the developmental trajectory of postlexical prominence, we had no predictions as to whether it would proceed linearly or not. That is, whether the duration of the strong vowels would lengthen in a linear fashion, progressively differentiating from unstressed vowels and from each other. Our results show that only the phonetic realization of the highest level of prominence proceeds linearly: the duration of IP-nuclear vowels steadily increases across months. At lexical level, also the phonetic realization of unstressed vowels evolves linearly, but in the opposite direction: their duration steadily decreases across months. Along with the progressive divergence in the duration of the lowest and highest degrees of prominence, in his prosodic development the child progressively adjusts the phonetic content (i.e. vowel duration) of the intervening levels of prominence. Two stages appear to be important: the one represented in the 24-month recording, where for the first time all the prominence levels are significantly different, but in which IP-nuclear vowels are shorter than ip-nuclear ones; and the 27-month recording, where the phonetic content is consistent with the place of the vowels in the prosodic hierarchy. At 3 years of age the process of prosodic prominence acquisition appears to be completed, after a temporary "regression" at 30- and 33-months, which is in line with the nature of speech acquisition.

In our future research we will examine the other acoustic correlates that cue prominence which have not been anayzed in the present paper, and will expand the present study to include the data up to 48 months of age. At the same time, we aim to analyse the other Italian children who have been recorded in the Bonifacio's database.

References

Arciuli, J., Colombo, L. (2016). An acoustic investigation of the developmental trajectory of lexical stress contrastivity in Italian. In *Speech Communication*, 80, 22-33.

Avesani C., Vayra M. (2013). Prosodic prominence and its articulatory bases. Poster presented at the International Conference "pS-prominenceS. Prominences in Linguistics", University of Tuscia (Viterbo, Italy), December 12-13, 2013.

Avesani C., Vayra M. & Zmarich C. (2009). Coordinazione vocale-consonante e prominenza accentuale in italiano. La sfida della Articulatory Phonology, in G. Ferrari, G., Benatti, R. & Mosca. M. (Eds.) (2006), *Linguistica e modelli tecnologici di ricerca.* Atti del XL Congresso Internazionale della SLI (Società di Linguistica Italiana) (Vercelli, 21-23 settembre 2006), Roma, Bulzoni, 365-399.

Bahatara, A., Boll-Avetisyan, B., Höhle, B. & Nazzi, T. (2018). Early sensitivity and acquisition of prosodic patterns at the lexical level. In Prieto, P., Esteve-Gibert, N. (Eds.), *The development of Prosody in First Language Acquisition,* Amsterdam/Philadelphia: John Benjamins, 38-57.

Bates, D.M. (2010). lme4: Mixed-effects modeling with R.

Beckman. M.E. (1996). The parsing of prosody. In *Language and Cognitive Processes*, 11, 17-68.

Beckman, M., Pierrehumbert, J. (1986). Intonational structure in English and Japanese. In *Phonology Yearbook*, 3, 255–310.

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

Bocci, G, Avesani, C. (2011). Phrasal prominences do not need pitch movements postfocal phrasal heads in Italian. In Cosi, P., De Mori, R., Di Fabbrizio, G. & Pieraccini, R. (Eds.), *Proceedings of Interspeech 2011*, Firenze, 27-31 August 2011, International Speech Communication Association, 1357-1360.

Boersma, P., weenink, D. (2021). PRAAT: doing phonetics by computer. [Computer program] Version 6.1.34. https://www.praat.org

Boisgontier, M.P. & Cheval, B. (2016). The anova to mixed model transition. In *Neuroscience & Biobehavioral Reviews*, 68, 1004-1005.

Caselli, M.C., Casadio, P. (1995). *Il Primo Vocabolario del Bambino*. Milano: Franco Angeli.

Chen, A., Esteve-Gibert, N., Prieto, P & Redford, M. (2021). Development of phrase-level prosody from infancy to late childhood. In Gussenhoven, C, Chen, A. (Eds.). *The Oxford handbook of language prosody*. Oxford: Oxford University Press, 553-562.

D'Odorico, L., Carubbi, S. (2003). Prosodic characteristics of early multi-word utterances in Italian children. In *First Language*, 23(1), 97-116.

D'Odorico, L., Fasolo, M. & Marchione, D. (2009). The prosody of early multi-word speech: word order and its intonational realization in the speech of Italian children. In *Enfance*, 3, 319-327.

D'Odorico, L., Fasolo, M. & Zanchi, P. (2010). Prosodic characteristics of multi-argument utterances in Italian children. In *Child Language Seminar*, London, UK.

Daza, E.J. (2019). Person as Population: A Longitudinal View of Single-Subject Causal Inference for Analyzing Self-Tracked Health Data. *arXiv preprint arXiv:1901.03423*.

DeCasper, A.J., Fifer, W.P. (1980). Of human bonding: Newborns prefer their mothers' voices. In *Science,* 208(4448), 1174-1176.

de Carvalho, A., He, A.X., Lidz, J. & Christophe, A. (2015). 18-month-olds use phrasal prosody as a cue to constrain the acquisition of novel word meanings. Paper presented at the *Boston University Conference on Language Development*, Boston.

de Carvalho, A., Lidz, J., Tieu, L., Bleam,T. & Christophe, A. (2016). English-speaking preschoolers can use phrasal prosody for syntactic parsing. In *Journal of the Acoustical Society of America,* 139(6), EL 216-EL 222.

de Carvalho, A., Dautriche, I., Lin, J. & Christophe, A. (2017). Phrasal prosody constrains syntactic analysis in toddlers. In *Cognition* 163, 63-79

Esteve-Gibert, N. Prieto, P. (2018), Early development of the prosody-meaning Interface. In Prieto, P. and Esteve-Gibert, N. (a cura di), *The development of Prosody in First Language Acquisition,* Amsterdam/Philadelphia: John Benjamins, 227-246.

Farnetani, E., Kori, S. (1982). Lexical stress in spoken sentences: a study on duration and vowel formant pattern. In *Quaderni del Centro di Studio per le Ricerche di Fonetica del CNR*, 1,106-133.

Farnetani, E., Faber, A. (1992). Tongue-jaw coordination in vowel production: isolated words versus connected speech. In *Speech Communication*, 11, 401-410

Frota, S., M. Cruz, N. Matos & M. Vigário. (2016). Early Prosodic Development: Emerging intonation and phrasing in European Portuguese. In Armstrong, M.E., N. Henriksen, N. & Vanrell, M.M. (Eds.), *Intonational grammar in Ibero-Romance: Approaches across linguistic subfields*. Amsterdam: Benjamins, 295-324

Gerken, L. (1994). A metrical template account of children's weak syllable omissions from multisyllabic words. In *Journal of Child Language*, 21, 565–584

Gerken, L., P., Jusczyk, W. & Mandel, D.R. (1994). When prosody fails to cue syntactic structure: 9-month olds' sensitivity to phonological vs syntactic phrases. In *Cognition*, 51, 237-265.

Gervain, J. (2015) Plasticity in early language acquisition: the effects of prenatal and early childhood experience. In *Current Opinion in Neurobiology*, 35, 13-20.

Gervain, J., Cristophe, A & Mazuka R. (2021). Prosodic bootstrapping. In Gussenhoven, C, Chen, A. (Eds.), *The Oxford handbook of language prosody*. Oxford: Oxford University Press, 563-573.

Hirsh-Pasek, K., Kemler Nelson, D.G., Jusczyk, P.W., Cassidy, K.W., Druss, B. & Kennedy, L. (1987). Clauses are perceptual units for young infants. In *Cognition*, 26, 269–286.

Jessen, M., Marasek,K. (1997). Voice quality correlates of word stress and tense versulìs lax vowels in German. In *Larynx*, Marseille, France, June 16-18, 127-130.

Jusczyk, P.W., Friederici, A.D., Wessels, J.M., Svenkerud, V.Y. & Jusczyk, A.M. (1993). Infants' sensitivity to the sound patterns of native language words. *Journal of Memory and Language, 32*(3), 402–420.

Jusczyk, P., Hirsch-Pasek, K., Kemler Nelson, D., Kennedy, L., Woodward, A. & Piwoz, J. (1992). Perception of acoustic correlates of major phrasal units by young infants, *Cognitive Psychology*, 24, 252-293.

Kehoe, M., Stoel-Gammon, C. & Buder, E.H. (1995). Acoustic correlates of stress in young children's speech. In *Journal of Speech, Language and Hearing Research*, 38(2), 338-350.

Ladd, R. (2008[2]). *Intonational Phonology*. Cambridge: Cambridge University Press.

Ling, D. (1976). *Speech and the hearing-impaired child: Theory and practice. Washington,* DC: Alexander Graham Bell Association for the Deaf.

Magno Caldognetto, E., Vagges, K. & Zmarich, C. (1995). Visible articulatory characteristics of the Italian stressed and unstressed vowels. *In Proceedings of the XIIIth International Congress of Phonetic Science*s, 1, 366-369.

Moon, C., Cooper, R.P. & Fifer, W. (1993). Two-day-olds prefer their native language. In *Infant Behavior and Development*, 16, 495-500.

Moon, C., Lagercrantz H. & Kuhl P. (2013). Language experience *in utero* affects vowel perception after birth: a two-country study. In *Acta Pediatrica*, 101(2), 156-160.

Nespor, M., Vogel, I. (1986). *Prosodic phonology*. Dordrecht: Foris. Berlin: Mouton de Gruyter.

Olivucci, F., Pasqualetto, F., Vayra, M. & Zmarich C. (2016). Lo sviluppo dell'accento lessicale nel bambino in età prescolare: una prospettiva fonetico-acustica. In Savy, R. and Alfano, I. (Eds.). *La fonetica nell'appendimento delle lingue / Phonetics and Language Learning*, *Studi AISV 6,* Milano: AISV, 219-228.

Olivucci, F., Vayra, M., Avesani, C. & Zmarich (2018). Acoustic correlates of word stress in young Italian children's productions. Presented at the *40th Annual Conference of the German Linguistics Society*, Stuttgart, March 9 2018.

Olivucci, F, Vayra, M., Avesani, C. & Zmarich (2019). The development of lexical stress in young Italian children. Poster presented at the *2019 Conference on Phonetics and Phonology in Europe (PaPE 2019)*, Lecce, June 17-19 2019.

Pollock, K.E., Brammer, D.M. & Hageman, C.F. (1993). An acoustic analysis of young children's productions of word stress. In *Journal of Phonetics*, 21,183-203.

Prieto, P., Estrella, A., Thorson, J. & Vanrell, M.M. (2012). Is prosodic development correlated with grammatical development? Evidence from emerging intonation in Catalan and Spanish. In *Journal of Child Language*, *39*(2), 221-257.

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Sansavini, A., Bertoncini, J. & Giovannelli, G. (1997). Newborns discriminate the rhythm of multisyllabic stressed words. In *Developmental Psychology*, 33 (1), 3-11.

Savy, R., Cutugno F. (1997). Ipoarticolazione, riduzione vocalica, centralizzazione: come interagiscono nella variazione diafasica. In Cutugno, F. (Ed.) *Fonetica e fonologia degli stili dell'italiano parlato.* Atti VII Giornate di Studio del GFS (Napoli, 14-15 novembre 1996), Roma: Esagrafica, 177-194.

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.

Schwartz, R.G., Petinou, K., Goffman, L., Lazawski, G. & Cartusciello, C. (1996). Young children's production of syllable stress: An acoustic analysis. In *The Journal of the Acoustical Society of America*, 99(5), 3192-3200.

Selkirk, E. (1984). *Prosody and syntax: The relation between sound and structure*. Cambridge, MA: MIT Press.

Shukla. M., Wite, K.S. & Aslin, R.N. (2011). Prosody guides the rapid mapping of auditory word forms onto visual objects in 6-mo-old infants. In *Proceedings of the National Academy of Sciences of the United States of America,* 108 (15), 6038-6043.

Slujiter, A.M., Van Heuven, V.J. (1996). Spectral balance as an acoustic correlate of linguistic stress. In *The Journal of the Acoustical society of America*, 100(4), 2471-2485.

Tamburini, F. (2009). Prominenza frasale e tipologia prosodica: un approccio acustico. In Ferrari G. (Ed.) *Linguistica e modelli tecnologici di ricerca.* Atti del XL Congresso internazionale di studi della Società di linguistica italiana (SLI) (Vercelli, 21-23 settembre 2006), Roma: Bulzoni, 437-455

van Rij, J., Hendriks, P., van Rijn, H., Baayen, R.H. & Wood, S.N. (2019). Analyzing the time course of pupillometric data. In *Trends in Hearing*, 23, 2331216519832483. https://doi.org/10.1177/2331216519832483

Vayra, M. (1991). Appunti su un effetto di 'centralizzazione' nel vocalismo dell'italiano standard. In L. Giannelli, N. Maraschio, T. Poggi Salani & M. Vedovelli (Eds), *Tra Rinascimento e strutture attuali della lingua. Atti del I Convegno Internazionale della S.I.L.F.I.* (Siena, 28-31 March 1989), Torino, Rosenberg & Sellier, 195-212.

Vayra, M., Fowler, C. (1987). The word-level interplay of stress, coarticulation, vowel height and vowel position in Italian. In *Proceedings of the XIth International Congress of Phonetic Sciences*, (Tallinn, 1-7 August 1987), 4, 24-27.

Vayra, M., Fowler, C. (1992). Declination of supralaryngeal gestures in spoken Italian. In *Phonetica*, 49(1), 48-60.

Vayra, M., Avesani, C. & Fowler, C.., C. (1999). On the phonetic bases of vowel-consonant cooordination in Italian: a study of stress and compensatory shortening. In *Proceedings of 14th ICPhS* (San Francisco, USA, 1-7 August 1999), 495-498.

Vihman, M. (2018). The development of prosodic structure. A usage.-based approach. In Prieto, P. and Esteve-Gibert, N. (a cura di), *The development of Prosody in First Language Acquisition,* Amsterdam/Philadelphia: John Benjamins, 185-206.

Wagenmakers, E.-J., Farrell, S. (2004). AIC model selection using Akaike weights. In *Psychonomic Bulletin & Review*, *11*(1), 192–196. https://doi.org/10.3758/BF03206482

Wood, S., Wood, M.S. (2015). Package 'mgcv'. R package version, 1, 29.

Zanchi, P., D'Imperio, M.P., Zampini, L. & Fasolo, M. (2016). L'intonazione delle narrazioni di bambini e adulti italiani. In Savy, R. and Alfano, I. (a cura di). *La fonetica nell'appendimento delle lingue / Phonetics and Language Learning*, *Studi AISV 6,* Milano: AISV, 179-189.

Zmarich C., Avesani C. (2015), L'influenza della durata consonantica sulla coarticolazione della sillaba CV con gradi diversi di prominenza prosodica. In A. Romano, M. Rivoira, I. Meandri (a cura di), *Aspetti prosodici e testuali del raccontare: dalla letteratura orale al parlato dei media*, Alessandria, Edizioni dell'Orso, 305-318.

Appendix

Figure a - *Density distribution plot of the overall duration residual included in the statistical modelling. The vertical line indicates the median value (144) of the whole distribution of vowel duration in milliseconds*

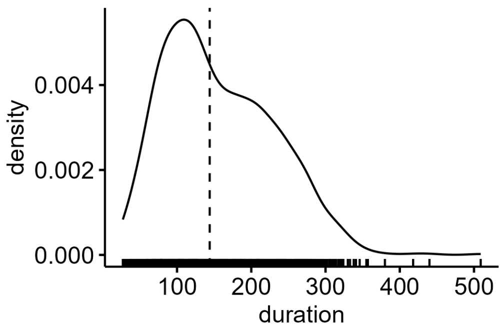


Figure b - *Selection of the best fitting family (Gamma vs Lognormal) distribution for the positive shewed distribution of the duration residuals. The Gamma family emerged to adequately fit the data*

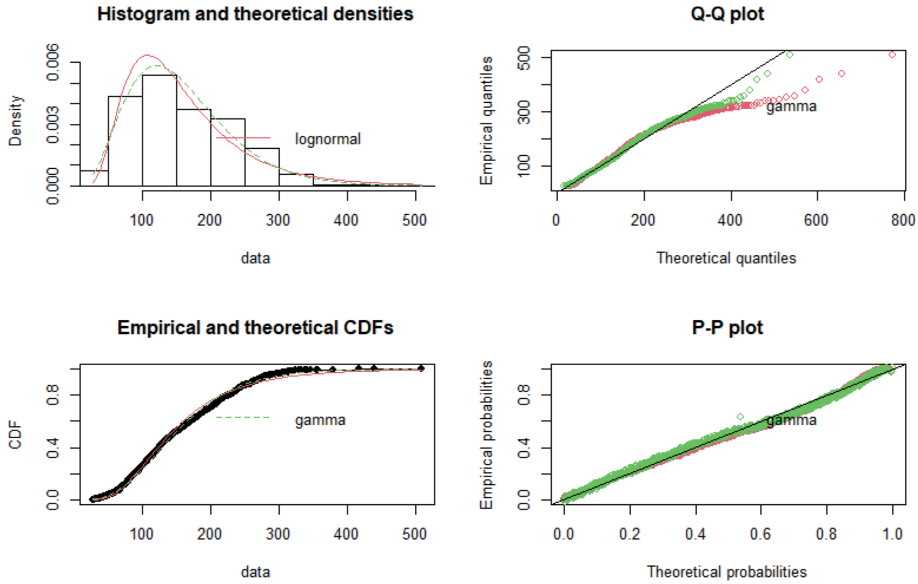


Figure c - *Autocorrelation plot showing an adequate low level of autocorrelation between Age (in months) and vowel duration (in milliseconds) as estimated by the generalized additive model (GAMM)*