“G. D'Annunzio” SCHOOL OF ADVANCED STUDIES

Ph.D. in BUSINESS & BEHAVIORAL SCIENCES

XXXVIII CYCLE

Prosody as an Underlying Mechanism in Human Communication: Perception, Development, Atypicality and Persuasion.

Department of Neuroscience, Imaging and Clinical Sciences
Scientific-Disciplinary Sector: M-PSI/04/L-LIN/01

PhD Candidate
Valeria D’Aloia

Coordinator
Prof. Riccardo Palumbo

Supervisor
Prof. Mirco Fasolo

Academic Years 2022–2025

Abstract

This doctoral thesis investigates prosody as a central mechanism underlying human communication across the lifespan, populations (typical and atypical) and communicative contexts. Rather than treating prosody as a peripheral suprasegmental component, the thesis argues that it plays a foundational role in structuring spoken interaction, and that its effects can be empirically observed and quantified, with direct consequences for communicative effectiveness, including persuasion. Methodologically and theoretically, the thesis bridges phonetic-acoustic and phonological approaches by bringing together acoustic analysis (e.g. pitch range, speech rate) and representations grounded in the Autosegmental-Metrical theory of Intonation. It further adopts a cross-linguistic and cross-contextual approach, analyzing data from Italian and Portuguese and examining prosody in significantly different communicative settings. Empirically, the thesis includes four studies. The first examines prenatal auditory sensitivity by investigating fetal physiological responses to instrumental musical stimuli differing in tempo, providing evidence that sensitivity to suprasegmental features is already present in utero. The second study presents a longitudinal analysis of pitch-related prosodic features in Italian infants' pre-lexical vocalizations, highlighting early prosodic organization and substantial individual variability. The third study provides a phonological account of contrastive focus prosody in Portuguese-speaking children with Autism Spectrum Disorder, showing that atypical prosody reflects differences in general phonological structuring rather than purely phonetic deviation. Finally, the fourth study provides the first corpus-based acoustic investigation of persuasive prosody in Italian television advertising, demonstrating how pitch variability and temporal organization differentiate persuasive speech from other broadcast registers. Overall, the thesis reconceptualizes prosody as a transversal communicative mechanism, measurable across development and contexts, with implications for linguistic theory, clinical assessment, and applied domains such advertising, and marketing.

TABLE OF CONTENTS

CHAPTER 1| INTRODUCTION 5

1.1. PROSODY: DEFINITION, FORM AND FUNCTION. 5
1.2. PROSODY AND PHONOLOGICAL REPRESENTATION: THE AUTOSEGMENTAL-METRICAL THEORY OF INTONATION. 9
1.3. EARLY PROSODIC SENSITIVITY AND DEVELOPMENT. 10
1.4. PROSODIC IMPAIRMENTS IN ATYPICAL POPULATIONS. 12
1.5. WHEN PROSODY IS OPTIMIZED: CHARISMA, INFLUENCE, AND PERSUASION. 14
1.6. OVERVIEW OF THE PRESENT THESIS: CONTRIBUTIONS AND STRUCTURE. 15

CHAPTER 2 | EXPLORING FETAL AUDITORY DISCRIMINATION AND CIRCULATORY INDICATORS OF WELL-BEING IN RESPONSE TO INDIE, DISCO-POP AND DRUM & BASS MUSIC: A PILOT STUDY. 18

2.1. INTRODUCTION 19
2.2. METHODS 22
2.2.1. Participants 22
2.2.2. Materials 22
2.2.3. Procedure 23
2.2.4. Plan of Analysis 23
2.3. RESULTS 24
2.3.1. Fetal Heart Rate (FHR) 24
2.3.2. Umbilical Artery Pulsatility Index (UA-PI) 25
2.4. DISCUSSION 27
2.5. CONCLUSIONS 30

CHAPTER 3 | THE DEVELOPMENT OF PROSODY IN PRE-LEXICAL VOCALIZATIONS: A LONGITUDINAL STUDY OVER THE FIRST 16 MONTHS OF INFANT LIFE. 31

3.1. INTRODUCTION 32
3.1.1. Prosodic development over infancy and toddlerhood: a contrasting picture. 34
3.1.2. Suggestions for a better comprehension of the phenomena. 39
3.1.3. The present study 40
3.2. METHODS 41
3.2.1. Participants 41
3.2.2. Procedure 42
3.2.3. Coding: pre-lexical productions 45
3.2.4. Coding: Prosody 46
3.2.5. Reliability 47
3.2.6. Plan of Analyses 47
3.3. RESULTS 48
3.3.1. Between subjects' variability exploration 49
3.3.2. F0 mean trajectory over time 50
3.3.3. F0 range trajectories over time 53
3.3.4. F0 final contours trajectories over time 55
3.4. DISCUSSIONS 61
3.5. CONCLUSIONS 67
3.6. ACKNOWLEDGMENTS 67

CHAPTER 4| THE PROSODY OF CONTRASTIVE FOCUS IN CHILDREN WITH HIGH-FUNCTIONING AUTISM 68

4.1. INTRODUCTION 69

4.1.1. Prosody in Autism Spectrum Disorder 69
4.1.2. Production of Contrastive stress in speakers with Autism Spectrum Disorder. 71
4.1.3. The present study 73
4.2. METHODS 74
4.2.1. Participants 74
4.2.2. Materials 75
4.2.3. Procedure 77
4.2.4. Plan of Analysis 82
4.2.5. Reliability 83
4.3. RESULTS 83
4.3.1. Sample characteristics and preliminary results 83
4.3.2. Target Match: Focus Accuracy Index (HFA vs TD) 84
4.3.3. Target Mismatch: Error types (HFA vs TD) 85
4.3.4. Prosodic Strategies for contrastive focus in Match and Mismatch responses. 87
4.4. DISCUSSION 93
4.5. CONCLUSIONS 100
4.6. DATA AVAILABILITY STATEMENT 100

CHAPTER 5 | PERSUASIVE PROSODY OF ITALIAN TV COMMERCIALS: AN ACOUSTIC BRIEF STUDY. 101

5.1. INTRODUCTION 102
5.2. METHODS 104
5.2.1. Materials 104
5.2.2. Procedure 106
5.2.3. Reliability 107
5.2.4. Plan of Analysis 108
5.3. RESULTS 108
5.4. DISCUSSION 110
5.5. CONCLUSIONS 114

CHAPTER 6 | GENERAL DISCUSSION 115

6.1. OVERVIEW OF THE MAIN FINDINGS 115
6.2. BROADER IMPLICATIONS AND FUTURE DIRECTIONS. 118
6.3. GENERAL CONCLUSION 119

GENERAL BIBLIOGRAPHY 121

ACKNOWLEDGEMENTS 153

CHAPTER 1| Introduction

1.1. Prosody: definition, form and function.

The term *prosody* refers to those non-segmental properties of speech that contribute to the organization and interpretation of spoken language. Prosodic phenomena are not tied to the choice of lexical items, but reflect how linguistic material is structured, highlighted, and interpreted within an utterance (Lehiste, 1970; Wagner & Watson, 2010). Although prosody has traditionally been described as *suprasegmental*, this label can be potentially misleading, as prosodic structure is intrinsic to spoken language and necessarily shapes the realization of all segments. There are no speech sounds in natural languages that are not affected by prosodic organization (Applebaum et al., 2014). At the level of the speech signal, prosody is manifested through systematic variation in acoustic parameters such as fundamental frequency (F0), duration, and intensity, which constitute the physical correlates of pitch, length, and loudness as perceived by listeners. Fundamental frequency reflects the rate of vibration of the vocal folds, duration corresponds to the temporal extent of articulatory gestures, and intensity is related to the amount of articulatory and respiratory effort involved in speech production (Sorianello, 2006). These acoustic parameters are intrinsic to speech production and perception and constitute the phonetic substance through which prosodic structure is realized. Variations in these dimensions give rise to phonological phenomena such as stress, rhythm, and intonation, which are conventionally grouped under the umbrella term *prosody*.

Despite its central role in speech, prosody does not admit a single, universally accepted definition. As noted by Shattuck-Hufnagel and Turk (1996), existing definitions differ in whether they emphasize acoustic variation, abstract phonological organization, or the integration of both. In line with integrative approaches, prosody is here conceived as a component of the grammar that establishes structured relations of prominence and grouping

among linguistic elements, and whose organization is systematically reflected in patterns of acoustic variation.

Beyond its formal properties, prosody fulfills a wide range of communicative functions. The prosody of an utterance is shaped by the words and phrases used, their syntactic and discourse context, and the speaker's intentions, emotional state, and attitude (Halliday, 1967; Cruttenden, 1986; Gussenhoven, 2004; Ladd, 2008). Prosodic functions can be broadly organized along a continuum ranging from more grammatical to more pragmatic domains (Grice & Baumann, 2007). Crystal (1969) instead differentiates prosodic functions into *linguistic* and *paralinguistic* uses. What these distinctions have in common is the assumption that some prosodic functions are governed by language-specific conventions and are rule-based, whereas others are more intuitive, gradient, and sensitive to contextual and interactional factors.

Grammatical functions of prosody involve relatively stable rules governing the placement of prominence and prosodic boundaries. These include, for example, lexical stress contrasts that distinguish grammatical categories in languages such as English (e.g. *IMport* as a noun vs. *imPORT* as a verb; Chomsky & Halle, 1968), lexical contrasts in tonal languages, where pitch patterns are lexically specified (e.g. Mandarin */ma/* "mother" vs. */mǎ/* "horse"; Yip, 2022), as well as the use of prosodic phrasing to disambiguate syntactic structure (Price et al., 1991). Prosody also contributes to the grammatical encoding of sentence type, such as the distinction between statements and questions, although the specific intonational patterns involved are not universal across languages (Pierrehumber, 1980).

Pragmatic functions of prosody relate to how utterances are interpreted in context and to how speakers adapt prosodic cues to communicative goals. A central function of prosody in this domain concerns the marking of information structure, including the distinction between given and new information and the organization of an utterance into background and focus

(Chafe, 1976; Kügler & Calhoun, 2020). Information that is discourse-new or informative is typically made prosodically prominent, whereas given information tends to be reduced (Cruttenden, 1986). Pragmatic uses of prosody also allow speakers to signal correction or opposition to information previously mentioned in the discourse, thereby highlighting the contrasted element, as in contrastive focus (Samek-Lodovici, 2018). In addition, prosody contributes to the management of conversational interaction, for instance by signaling turn-taking intentions and by conveying communicative intentions such as whether an utterance is meant literally or ironically (Dynel, 2019; Hirschberg et al., 2020). Lastly, prosodic modulation also provides cues to the speaker's internal state, conveying affective meaning through acoustic parameters such as pitch, speech rate, and intensity. This dimension of prosody is commonly referred to as *affective prosody*, as it concerns the expression and perception of emotions and affective states in speech; for example, a higher pitch and faster tempo are often associated with happiness, whereas a lower pitch and slower rate are linked to sadness and depression (Mozziconacci, 2002; Bänziger et al., 2015; Belyk & Brown, 2014).

The range of prosodic functions described above can be summarized along a continuum from more formal, rule-based functions to more intuitive, context-dependent ones. A visual representation of this continuum is provided in Figure 1. (adapted from Grice et al., 2017).

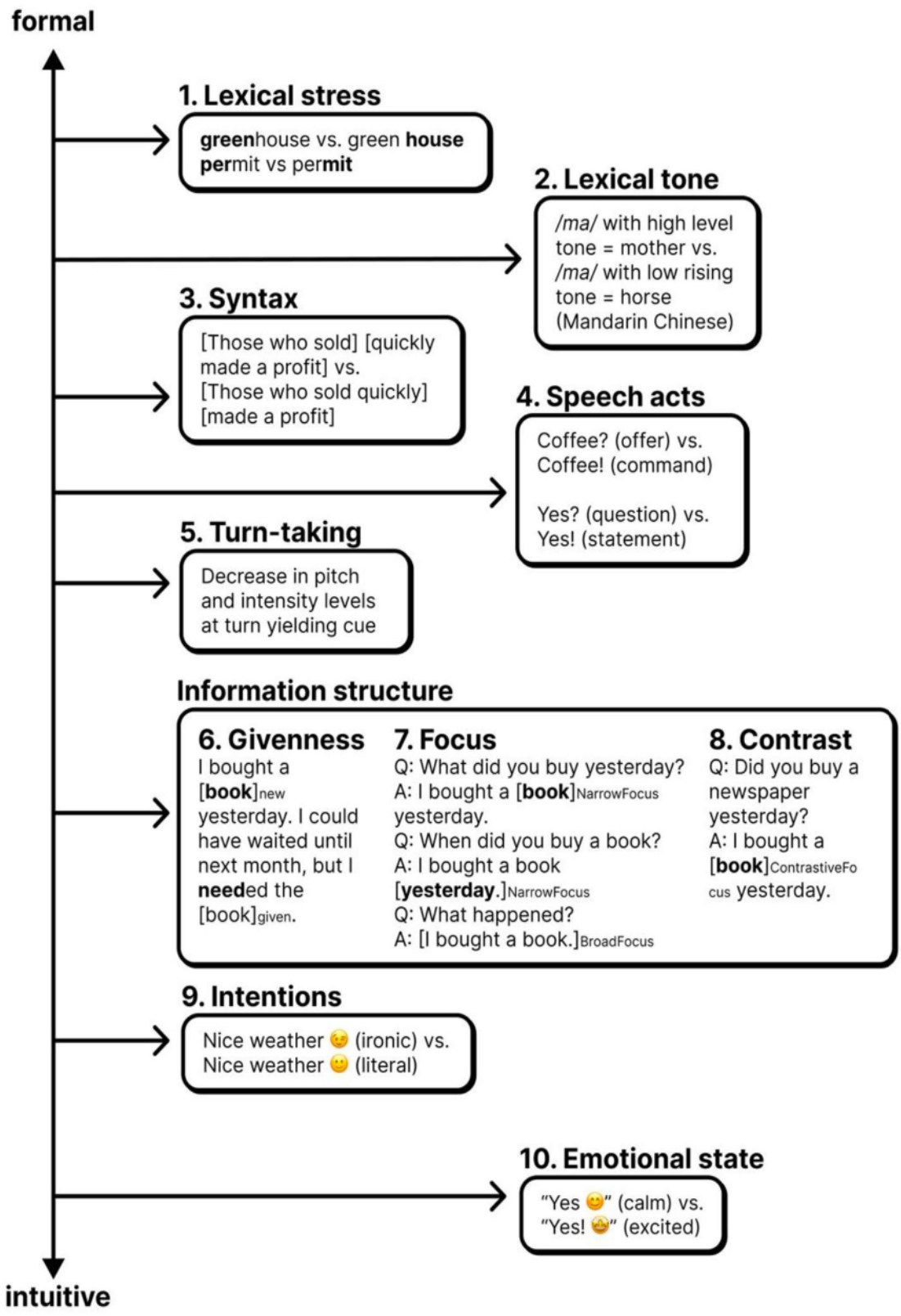


Figure 1. Prosodic functions organized on a scale ranging from more formal and rule-based to more intuitive and context-dependent uses (from Grice et al., 2017)

1.2. Prosody and phonological representation: The Autosegmental-Metrical Theory of Intonation.

Within the broader domain of prosody, intonation represents one of the most extensively studied components, as it exploits pitch variation to structure and interpret spoken utterances. Although intonation is realized through continuous acoustic variation, it has been shown to be systematically organized. Intonational patterns can be categorized into recurrent, language-specific configurations that form part of the grammatical system (Pierrehumbert, 1980; Ladd, 2008).

The Autosegmental-Metrical (AM) theory of intonation was developed to capture this structured nature of intonation by providing a phonological representation of pitch patterns. Within the AM framework, intonation is represented as a sequence of discrete tonal elements, which can be High (H) or Low (L), and function as abstract phonological units, rather than as a continuous phonetic representation of the F0 contour (Pierrehumbert, 1980; Beckman & Pierrehumbert, 1986; Ladd, 2008). These tones associate with prosodic structure in systematic ways. In the AM framework, utterances are organized into hierarchically structured prosodic constituents, ranging from smaller units such as the syllable to larger domains such as the phonological and intonational phrase (Selkirk, 1984; Nespor & Vogel, 1986; Pierrehumbert & Beckman, 1988). Within this structure, tones may associate with metrically prominent syllables, namely the nuclear syllable, defined as the most prominent syllable within a phonological or intonational constituent, giving rise to pitch accents (PAs), or with the edges of prosodic constituents, giving rise to edge tones. itch accents are tonal events associated with stressed syllables and are indicated by a star (*), which marks the tone aligned with the accented syllable *(e.g.* H* or L*). Pitch accents can be monotonal, consisting of a single tone (e.g. H*), or bitonal, combining two tones (e.g. L*+H or H*+L), resulting in rising or falling pitch movements across the accented syllable (Pierrehumbert, 1980; Ladd, 2008). In contrast, tones

associated with the boundaries of prosodic constituents serve a delimitative function. These include phrase accents, typically marked with a hyphen (e.g. H- or L-), and boundary tones, marked with a percent sign (e.g. H% or L%), which signal the right edge of larger prosodic units such as the intonational phrase (Pierrehumbert & Beckman, 1988; Pierrehumbert & Hirschberg, 1990). By distinguishing between tonal events associated with prominence and those marking prosodic boundaries, the AM framework provides a principled account of how intonation encodes both structural and pragmatic information. This representational approach has proven particularly suitable for cross-linguistic comparison, as it allows intonational systems to be analyzed within a shared theoretical framework, accounting for both language-specific variation (Jun, 2005a; Frota & Prieto, 2015; Gili Fivela et al., 2015) and differences across speaker populations, including typical and atypical development.

1.3. Early prosodic sensitivity and development.

Sensitivity to prosodic structure emerges very early in development and precedes the acquisition of segmental and grammatical knowledge. Auditory development begins prenatally, with functional hearing becoming possible during the third trimester of gestation (Moore, 2002; Kisilevsky et al., 2000). Crucially, the intrauterine environment acts as a low-pass filter: maternal tissues and amniotic fluid attenuate higher-frequency spectral information while preserving low-frequency cues related to rhythm, tempo, and intonation (Querleu et al., 1989; Gerhardt et al., 1992; Smith et al., 2003). Importantly, these preserved low-frequency acoustic dimensions (rhythm, tempo, and melodic contour) are not specific to speech, but are also core features of music. Consistent with this overlap, experimental studies using cardiac and motor habituation paradigms have shown that late-stage fetuses can discriminate between low-pitched musical stimuli, such as two piano notes (D4 vs. C5; Shahidullah & Hepper, 1994), as well as between low-frequency sounds in the 250–500 Hz range (Lecanuet et al., 2000).

As a result, the fetus' earliest exposure to language primarily consists of prosodic patterns rather than segmental features (Vihman, 2014). In line with this early sensitivity, fetuses and newborns show responsiveness to temporal and melodic regularities and exhibit preferences for their mother's voice and for familiar auditory stimuli encountered prenatally (DeCasper & Fifer, 1980; Kisilevsky et al., 2003).

Postnatally, prosody continues to play a central role in early language development. Infants show a robust preference for Infant-Directed Speech (IDS), a register characterized by exaggerated prosodic features such as expanded F0 range, higher overall pitch, slower speech rate, longer durations, and simplified syntactic structures (Teixidó et al., 2018). Beyond its linguistic salience, it has been shown that IDS supports affective communication by conveying emotional states and fostering attunement and bonding between caregiver and infant (Spinelli et al., 2017). From a linguistic perspective, studies showed the prosodic properties of IDS facilitate language acquisition through *prosodic bootstrapping*, whereby infants use cues such as pauses, final lengthening, and intonational contours to access higher levels of linguistic structure, including word segmentation and syntactic organization (Nespor & Vogel, 1986; Vihman, 2014; Gervain et al., 2020; Carvalho et al., 2018).

Given its early perceptual availability and central communicative role, it is not surprising that prosody is also among the earliest linguistic components to emerge in production**.** Language-specific prosodic patterns are observable from birth, as shown by cross-linguistic differences in newborn crying melodies that mirror the intonational properties of the ambient language (Mampe et al., 2009). During the first year of life, infants increasingly use prosodic modulation to express communicative intent, and by the end of the first year they begin to control pitch and duration as a function of communicative context (Chen & Fikkert, 2007). Importantly, intonational contours associated with specific pragmatic functions emerge before the stabilization of multi-word utterances: between 12 and 18 months, children produce falling

contours in declarative contexts and rising contours in requests (Aureli et al., 2017), and similar findings have been reported cross-linguistically (Prieto et al., 2012). By the second year of life, children show adult-like melodic patterns and a largely established inventory of nuclear pitch accents and boundary tones, supporting the view that prosodic development precedes and scaffolds grammatical development (Frota & Butler, 2018). Within the Autosegmental-Metrical framework, these findings have led to the description of early intonational grammars for several languages, including English, Dutch, Portuguese, Spanish, and Catalan, highlighting the central role of prosody in both typical language development and its theoretical modeling (Frota & Butler, 2018).

1.4. Prosodic impairments in atypical populations.

Prosodic variation does not only characterize typical language development but also plays a central role in shaping communicative profiles in atypical populations. Across a wide range of developmental, genetic and neuropsychiatric conditions, prosody shows systematic deviations from typical patterns, often co-occurring with broader cognitive, linguistic or pragmatic impairments. These alterations frequently manifest as disruptions in rhythm, intonation and in the appropriate use of prosodic cues during interaction. Importantly, prosodic deficits rarely affect all functions uniformly: across disorders, pragmatic, socially oriented and context-dependent uses of prosody appear to be more severely compromised than rule-based grammatical uses (Crystal, 1981; Sorianello, 2024).

This pattern is consistently reported in Autism Spectrum Disorder (ASD). Several studies have shown that, while grammatical uses of prosody, such as the marking of phrasal boundaries, may be relatively preserved, pragmatic uses, including the signaling of contrast, the correction of prior information and the management of discourse interaction, are often impaired (Fine et al., 1991; Paul et al., 2005; Peppé et al., 2007). At the acoustic level, these

difficulties are frequently associated with reduced pitch variability, atypical distribution of prominence and atypical speech rate, which in turn contribute to impairments in social communication and interaction (for an extensive discussion of prosody in ASD, see Chapter 4). Prosodic impairment has also been widely documented in schizophrenia spectrum disorders, which have long been described as disorders of communication (Ruesch, 1957). Patients typically show severe pragmatic deficits, including difficulties in interpreting non-literal meaning such as irony or humor, alongside marked impairments in emotional prosody (Kuperberg, 2010a; Bambini et al., 2016; Couture et al., 2006). Acoustically, speech is often characterized by flattened intonation, reduced F0 range and diminished modulation of intensity and timing, resulting in monotonous prosodic patterns (De Boer et al., 2023). In genetic conditions such as Down syndrome, prosodic profiles reflect a complex interaction between cognitive delay, motor constraints and phonetic limitations. Speech is typically characterized by slower rate, frequent pauses, unstable intensity and a reduced pitch range, often accompanied by elevated mean F0 (Nadel, 1988; Law & Bishop, 2004; Sorianello, 2012). Interestingly, receptive abilities may be less impaired than expressive ones: individuals with Down syndrome often understand basic intonational contrasts (e.g. rising vs. falling contours in questions vs. statements) better than they can produce them (Wells & Peppé, 2003; Zampini et al., 2016).

Given the close interaction between prosodic form, communicative function and social interaction, the assessment of prosodic deficits has therefore become a central issue in clinical linguistics, leading to the development of dedicated tools such as the Prosody-Voice Profiles (Shriberg, 1993) and the PEPS-C battery (Peppé & McCann, 2003; see Chapter 4), alongside perceptual and acoustic approaches. Overall, the evidence reviewed points to a systematic link between atypical prosodic patterns and sociopragmatic dysfunction, highlighting prosody as a crucial interface between linguistic structure, cognitive processes and social communication.

1.5. When prosody is optimized: charisma, influence, and persuasion.

If atypical prosodic patterns are often associated with communicative breakdowns and sociopragmatic difficulties, strategically enhanced prosody can, conversely, function as a powerful resource for influence and persuasion. A substantial body of research has shown that how a message is delivered vocally can be as crucial as its linguistic content, with prosodic modulation shaping listeners' attention, emotional engagement and judgments of speaker credibility (Tracy et al., 2011; Guyer et al., 2019; Gomes et al., 2023). From an acoustic perspective, several prosodic features have been identified as particularly relevant for persuasive communication, including fundamental frequency (F0) mean, pitch range, intonational contour and speech rate (Blanc & Dominey, 2003; Grichkovtsova et al., 2009).

Pitch-related cues play a central role: lower-pitched voices are consistently associated with dominance, leadership and competence, influencing political outcomes as well as commercial behavior (Klofstad & Anderson, 2018; Montrey, 2005). In sales contexts, speakers using falling pitch contours and sentence-final pitch lowering have been shown to achieve higher sales performance, suggesting that intonational patterns directly impact persuasive effectiveness (Peterson et al., 1995). At the same time, increased pitch variability, reflected in a wider F0 range has been linked to higher perceived charisma and speaker engagement (Berger et al., 2017; Niebuhr et al., 2016). These effects are grounded in more general perceptual mechanisms: high-pitched sounds tend to attract attention more efficiently, a phenomenon known as the *high voice superiority effect*, which is partly explained by the physiological properties of the cochlea, designed to respond rapidly and robustly to high-frequency stimuli (Marie & Trainor, 2013; Hill & Miller, 2010).

Beyond pitch, speech rate also contributes to persuasion, with moderately faster delivery associated with confidence and effectiveness, although excessively rapid speech can place greater demands on processing, ultimately reducing persuasive effectiveness (Apple et al.,

1979; Smith & Shaffer, 1995). Together, these findings indicate that prosody influences persuasion by modulating both attentional and affective responses, in line with Aristotle's notion of *pathos* as the ability of speech to shape the audience's emotional disposition (Aristotle, *Rhetoric*, 1.2.1356a; Kischek, 2018). These mechanisms are routinely exploited in public-sphere communication, most notably in advertising, where vocal delivery must rapidly capture attention, sustain engagement and promote favorable attitudes toward products and brands (Leech, 1966; Kanner, 1999; Mehrabian et al., 1995). Despite this consensus in the literature, acoustic studies of persuasive prosody in broadcast media, and in Italian television advertising in particular, remain limited, motivating part of the present investigation to explore the acoustic correlates of persuasive prosody in Italian TV commercials.

1.6. Overview of the present thesis: contributions and structure.

Considering the evidence reviewed in the preceding sections, this doctoral thesis advances the view that prosody constitutes a structural backbone of spoken communication, rather than a peripheral suprasegmental component. By integrating evidence from prenatal auditory experience, early language development, atypical populations, and persuasive adult speech, the thesis demonstrates that prosody supports communicative processing across the lifespan, across populations (typical and atypical), and across diverse contexts of language use. Crucially, across all these domains, prosody is approached not only as an abstract linguistic construct, but also as a measurable and categorizable behavioural phenomenon, investigated primarily through acoustic indices, interpreted within a phonological framework, and, in one study, explored through physiological responses.

This perspective allows prosody to be investigated as a functional mechanism linking phonetic-acoustic realization, phonological structure, cognitive processing, and communicative behavior across different linguistic systems. Building on this perspective, the

thesis makes several original contributions to the study of prosody.

At a methodological and theoretical level, it bridges phonetic-acoustic and phonological approaches by combining fine-grained acoustic analyses (e.g., F0 mean and range, speech rate) with phonological analyses grounded in the Autosegmental-Metrical (AM) framework, showing how phonetic-acoustic variation and phonological representations jointly contribute to explaining prosodic form and function. In addition, the thesis adopts a cross-linguistic and cross-contextual perspective, examining prosody across different languages, speakers' populations and communicative contexts. Empirically, the thesis provides novel evidence on prosody in multiple domains. It offers a phonological analysis of Autism Spectrum Disorder's speech, showing that atypical prosodic patterns reflect differences also in prosodic structuring rather than only phonetic deviation. The thesis further presents the first corpus-based acoustic investigation of persuasive prosody in Italian television advertising, addressing a clear gap in the literature and highlighting the importance of studying prosody in naturally occurring media speech. Finally, the thesis contributes new evidence on early auditory sensitivity by situating prenatal perception within a broader framework of suprasegmental processing. By focusing on rhythm and tempo, acoustic dimensions shared by speech prosody and music, the study justifies the use of instrumental musical stimuli as a proxy for early prosodic experience and highlights the impact of prenatal acoustic stimulation on measurable physiological responses related to sensory processing and fetal well-being.

Tracing the development of prosodic abilities across the lifespan, the thesis is structured as follows. After the introductory **Chapter 1**, **Chapter 2** investigates prenatal auditory sensitivity by examining fetal responses at 33 weeks of gestation to three instrumental musical stimuli differing in tempo. Using physiological measures like Fetal Heart Rate and Umbilical Artery Pulsatility Index, the chapter addresses the earliest stage of suprasegmental auditory processing before birth.

Chapter 3 focuses on early prosodic development in infancy, presenting a longitudinal study of Italian infants between 4 and 16 months of age. The chapter analyzes F0-related prosodic measures in pre-lexical vocalizations, with particular attention to developmental trajectories and individual variability.

Chapter 4 examines prosodic organization in atypical development, presenting a phonological analysis of contrastive focus prosody in Portuguese-speaking children (age 5-15) with Autism Spectrum Disorder. Using data from the Contrastive Focus Production task of the PEPS-C and adopting AM framework, the chapter analyzes both the placement of prominence, and the prosodic strategies used to signal it.

Chapter 5 shifts to adult communication and investigates prosody in Italian persuasive speech through a corpus-based acoustic analysis of television advertising. Using data from the *CLIPS* corpus, the chapter compares persuasive and non-persuasive broadcast speech, focusing on prosodic measures such as F0 range, speech rate, and the duration of speech units.

Chapter 6 presents a general discussion, integrating findings across chapters and outlining their implications for the study of prosody across development, populations and communicative contexts.

CHAPTER 2 | Exploring Fetal Auditory Discrimination and Circulatory Indicators of Well-being in response to Indie, Disco-Pop and Drum & Bass Music: A Pilot Study.

D'Aloia V., Logrieco M.G., Passaquindici I., Baldassarri C., Spinelli M., Lionetti F., Palumbo R., Buca D., Fasolo M. *(under review)* Exploring Fetal Auditory Discrimination and Circulatory Indicators of Well-being in response to Indie, Disco-Pop and Drum & Bass Music: A Pilot Study. *Developmental Psychobiology.*

Abstract

Fetal auditory perception begins in the second trimester, yet little is known about how different musical genres affect fetal discrimination responses and circulatory indices of well-being. This study investigated the effects of three musical genres (Indie, Disco-Pop, Drum & Bass) on Fetal Heart Rate (FHR) and Umbilical Artery Pulsatility Index (UA-PI) in nine fetuses at 33.4 ± 1.4 weeks gestational age. Music with varying tempos was played while FHR and UA-PI were measured using Doppler ultrasound. Baseline values for both FHR and UA-PI were recorded before each stimulus. Results showed a general decrease in FHR during musical exposure, with a significant reduction for Disco-Pop and Indie conditions. UA-PI increased significantly from baseline during Disco-Pop and Drum & Bass, but remaining withing the normal range. No significant differences emerged between the musical stimuli themselves, with statistical effects observed only between the between baseline and music exposure. The decrease in FHR suggests an orienting response consistent with early auditory discrimination, whereas the UA-PI increase in UA-PI may indicate a physiological adjustment linked to well-being processes, such as blood flow redistribution. These preliminary findings highlight how external auditory stimuli can influence both sensory discrimination and circulatory parameters, warranting larger-scale studies.

Keywords: *Fetal Auditory Perception, Fetal Heart Rate, Umbilical Artery Pulsatility Index, Prenatal Sensory Development, Music Exposure.*

2.1. Introduction

Auditory development in fetuses begins at the 20th gestational week with cochlear maturation, followed by axonal myelination at 26 to 28 weeks and rapid, synchronous conduction in brainstem pathways at 29 weeks (Moore, 2002). It follows that fetuses can hear at least by the last trimester of pregnancy with functional maturation of active cochlear mechanism nearly complete at around 33 weeks of gestational age (GA) (Kisilevsky et al., 2000). In recent years, there has been growing interest in fetal cognition, which has led to an increasing body of documentation on fetal auditory perception in response to acoustic stimuli. Cardiac and motor habituation/disabituation procedures have shown that late-stage fetuses, can discriminate two low-pitched piano notes, D4 and C5 (Shahidullah & Hepper, 1994), and two low-frequency sounds (250 Hz -500 Hz) (Lecanuet et al., 2000). Furthermore, studies that have focused on linguistic stimuli revealed mature fetuses' ability to discriminate vowels, sequences of phonemes with vowels presented in reversed order ([ba][bi] vs [bi][ba]), and temporal variations of a new language compared to the one heard *in utero* (Carvalho et al., 2019).

Studies focusing specifically on music perception have shown varying results. In some cases, a calming effect with a significant reduction in fetal heart rate (FHR) was observed during classical music exposure in Non-Stress Tests (Gebuza et al., 2018), whereas in others, an increase in FHR was found during music-induced relaxation (García Gonzáles et al., 2017). Movalled et al. (2023) highlighted methodological inconsistencies across these studies including different delivery methods (e.g., loudspeakers, headphones, or direct abdominal transmission), which may affect the sound quality reaching the fetus, thereby altering its effects. Additionally, all studies involved mothers listening to the music, potentially skewing fetal behavior measurements due to maternal emotional and hormonal influences. Research studies that have evaluated fetal behavior in response to multiple musical stimuli while masking mothers hearing, are limited.

Granier-Deferre et al. (2011) studied heart rate (HR) responses in near-term fetuses exposed to piano melodies with contrasting contours, a natural Icelandic sentence, and a version of the sentence with its spectral information replaced by broadband noise. All stimuli caused HR deceleration, with no significant differences between categories, indicating sustained attention and enhanced information processing. In contrast, fetal responses to tempo variations in Brahms' Lullaby have also been explored (Kisilevsky et al., 2004), with the faster tempo (118 BPM) leading to increased HR, likely due to heightened arousal from more stimulating input to the auditory system. Similar findings were reported in a study by Tristao et al. (2020), where HR and motor behavior were measured in 41 fetuses (37.4 GA) exposed to a Brazilian lullaby at slow (Adagio) and fast (Allegro) tempos, as well as a version sung by the mother in Adagio. HR increased with the faster tempo but decreased more significantly during the mother's voice, suggesting a preference for that stimulus. These findings collectively demonstrate that fetuses can distinguish between different sound sources and tempo variations based on both physiological and behavioral responses.

Heart rate and body movements were once the primary indicators used to assess fetal well-being (Lecanuet et al., 1992). However, with advancements in medical technology, vascular indices such as fetal circulation measurements, have become the best indicators for fetal health assessment (Esposito et al., 2021). Among these, the Pulsatility Index (PI), particularly when measured in the umbilical artery (UA-PI), is widely used to evaluate fetal and placental circulation and has been associated with fetal well-being and pregnancy outcome (Ferdousi et al., 2013). The UA-PI is an index that evaluate the resistance of blood flow in the umbilical artery (Meise at al., 2001), with abnormal values often indicating conditions like Intrauterine Growth Restriction (IUGR) (Marsàl, 2009) and Small for Gestation Age (SGA) (Bolz et al., 2013). To this day, the studies that have explored variations in fetal-placental circulation in response to acoustic stimulation, exploring indices like UA-PI, are exceedingly limited.

Kawakami et al. (2002) employed Umbilical Artery Velocity Waveforms (UA-VW) and Middle Cerebral Artery Velocity Waveforms (MCA-VW) to assess fetal responses to auditory stimuli. While the Umbilical Artery measures the resistance to blood flow from the placenta, the Middle Cerebral Artery evaluates the blood flow to the fetal brain. Nine healthy fetuses between 37 and 39 GA were exposed at 105 decibels (dB) to two different auditory stimuli, white noise and Japanese drumming, while their mother was listening to music of their choice. From baseline throughout the two stimuli, both UA-VW and MCA-VW showed increased variability; however, this effect was statistically significant only for the UA-VW. The authors concluded that UA-VW resistance indices may be a sensitive index in response to sounds. Another study reporting music-related changes in fetal hemodynamics is that of Gupta and Gupta (2018). However, this was a music therapy study conducted in women with pregnancy-induced hypertension (PIH), a condition associated with elevated Doppler indices. In the experimental group, after one month of classical music listening, UA-PI decreased. The authors argued that music had a positive effect on uteroplacental flow in pregnancy-induced hypertension.

To conclude, the literature has shown that music is perceived by the fetus, with its variations in tempo and tonalities potentially influencing physiological parameters, such as heart rate and blood flow (Kulinsky et al., 2022). However, limited research has examined the relationship between musical stimuli, auditory discrimination, and fetal well-being indicators. This study aims to investigate whether three music genres with tempo variations elicit distinct fetal responses, specifically auditory discrimination reflected by changes in FHR, and potential modulation of fetal-placental circulation, assessed via UA-PI, a physiological measure often considered indicative of fetal well-being.

2.2. Methods

2.2.1. Participants

The study was conducted in a private gynecological clinic in the city of Pescara, Italy. The project involved the collaboration of a gynecologist, assisted by an obstetrician, who selected the sample and presented the project proposed by our research team to the mothers attending the clinic. The sample, composed of mothers who voluntarily accepted to participate in the research, consisted eventually of nine adult pregnant women, with gestational ages ranging from a minimum of 32 weeks to a maximum of 35 weeks (33,4 ± 1,4) and with fetuses showing no evidence of chromosomal, congenital or neurological abnormalities. The study was approved by the ethical committee of the Department of Neuroscience Imaging and Clinical Science of the University of Chieti-Pescara (Ethical approval number: DNISC2962, 06.11.2019) and was conducted according to the American Psychological Association guidelines in accordance with the 1964 Helsinki Declaration. A written informed consent (approved by the ethical committee) for participation in the study has been obtained by the mothers.

2.2.2. Materials

For this study, three distinct music genres were selected, each varying in tempo: Indie, Disco-Pop, and Drum & Bass. The Indie genre was represented by "Pesto", an Italian song by the indie singer Calcutta, with a tempo of 81 beats per minute (BPM). For the Disco-Pop genre we chose "Dynamite," a track by the K-pop group BTS, with a tempo of 114 BPM. Finally for the Drum & Bass genre we selected "Slam," a song by the electronic-rock band Pendulum, with a tempo of 174 BPM. All selected versions were instrumental, and the order of stimuli presentation was randomized to control for any potential order effects, ensuring balanced exposure to each genre throughout the study. The instrumental version of the songs and their

tempo features were extracted using the Audacity software (Version 3.2.5, Audacity Team 2023). Measurements of fetal heart rate (FHR) and the Umbilical Artery Pulsatility Index (UA-PI) were obtained using the Samsung V8 Doppler ultrasound system.

2.2.3. Procedure

The study was conducted during a follow-up routine exam. Before fetal auditory stimulation began, the mothers' hearing was masked using a Marshall (Major IV) headset playing white noise to prevent interference with the fetus' reactions. Sound stimuli, presented in random order, began only after no fetal movements were recorded and the heart rate had stabilized for at least 90 seconds. Each stimulus lasted for 30 seconds, with a 30-second interval between each sound. In total, each session lasted 2 minutes and 30 seconds. The room's noise level was initially measured in silence during the ultrasound procedure, with ambient noise level at 40 dB (± 20) (digital sound level meter, Taledo SL720). Songs were played at 80 dB (± 20), considering that attenuation in the womb is around 20 dB (Granier-Deferre at al., 2011). Stimuli were delivered through a speaker set (Trust 16697-04 MILA 2.0) positioned 100 cm from the mother's abdomen. After taking baseline values for both fetal heart rate (FHR) and Umbilical Artery Pulsatility Index (UA-PI), the gynecologist measured these parameters again at the end of each stimulus presentation.

2.2.4. Plan of Analysis

Due to the small sample size, and the potential for non-normality in the data, a non-parametric approach was chosen for the analysis. Specifically, the Wilcoxon signed-rank test was employed to assess differences between paired measures. This test allowed to compare the pairs of measurements obtained from stimuli responses effectively. Statistical analyses were conducted using Jamovi (Version 2.3.28.0; Jamovi, 2023) and the significance level was set at $\alpha = 0.05$. Median values are used to summarize the central tendency of each condition due to

the non-parametric nature of the test.

2.3. Results

2.3.1. Fetal Heart Rate (FHR)

Figure 2.1 shows a general decrease in FHR from baseline in response to music conditions. The Wilcoxon signed-rank test revealed a significant difference between baseline FHR and FHR during the Indie condition (W = 36.5, p = 0.05), with a median FHR of 140 beats per minute (bpm) at baseline compared to 137 bpm during the Indie exposure. A similar difference was found between baseline FHR and Disco-Pop condition FHR (W = 31.5, p = 0.03), with a median FHR of 138 bpm during the Disco-Pop exposure. In contrast, no significant difference was found between baseline FHR and FHR during Drum & Bass condition (W = 34.5, p = 0.08, *Md* Drum & Bass = 138 bpm). Excluding the baseline, comparisons between stimuli revealed no difference. Specifically, there was no statistically significant difference in FHR between the Indie and Disco-Pop conditions (W = 34.5, p = 0.08), nor between Indie and Drum & Bass exposure (W = 18.0, p = 0.52). Similarly, no significant difference in FHR was observed between Disco-Pop and Drum & Bass condition (W =14.5, p = 0.84). Descriptive statistics and Wilcoxon outcomes for FHR are reported in Tables 2.1 and 2.2.

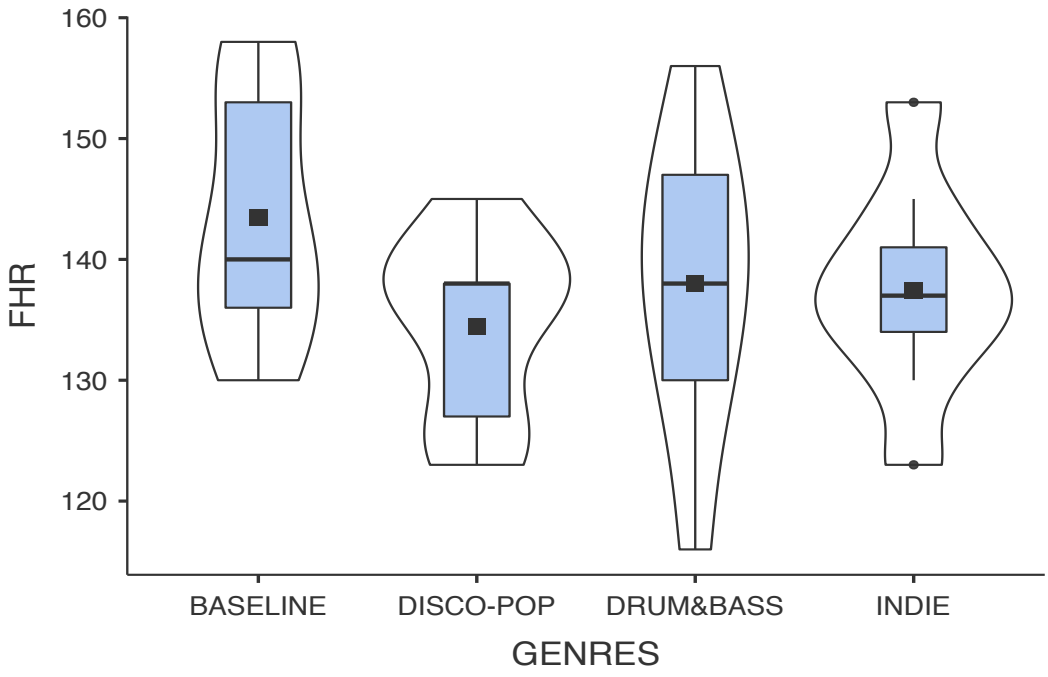


Figure 2.1 Boxplots of Fetal Heart Rate (FHR) across baseline and three music conditions (Indie, Disco-Pop and Drum & Bass). Boxplots show the median (horizontal line), mean (square), interquartile range (box edges) and outliers (circles).

2.3.2. Umbilical Artery Pulsatility Index (UA-PI)

Figure 2.2 illustrates an increase in UA-PI from baseline condition in response to music stimuli. Wilcoxon signed-rank test comparisons between baseline UA-PI and the responses to different musical genres revealed a significant difference for the Disco-Pop condition (W = 7.00, p = 0.03), where the median UA-PI was 0.880, compared to a median of 0.850 at baseline. A significant difference was reported as well in response to the Drum & Bass condition (W = 3.00 and p = 0.03) with the median UA-PI being 0.900. In contrast, the comparison between baseline UA-PI and the Indie condition (*Md* UA-PI = 1.00) showed a non-significant result (W = 12.00, p = 0.12). When examining UA-PI values across different music conditions, the comparison between Indie and Disco-Pop condition revealed no significant difference (W = 20.00, p = 0.63), nor did the comparison with the Drum & Bass exposure (W = 17.50, p = 0.29). Lastly, the analysis of UA-PI values between the Disco-Pop and Drum & Bass conditions also did not yield significant results (W = 13.00, p = 0.26). Descriptive statistics and Wilcoxon outcomes

for UA-PI are reported in Tables 2.1 and 2.2.

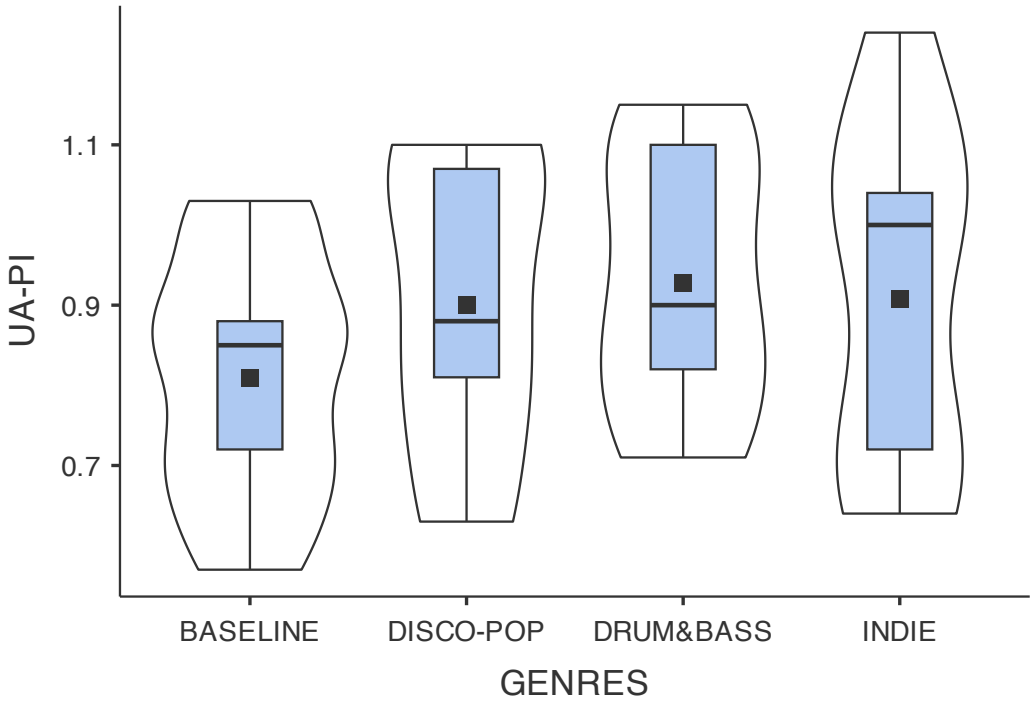


Figure 2.2. Boxplots of Umbilical Artery Pulsatility Index (UA-PI) across baseline and music conditions (Indie, Disco-Pop and Drum & Bass). Boxplots show the median (horizontal line), mean (square), interquartile range (box edges) and outliers (circles).

Table 2.1. Median values for Fetal Heart Rate (FHR) and Umbilical Artery Pulsatility Index (UA-PI) across baseline and music conditions (N=9)

Condition	FHR, Md	UA-PI, Md
Baseline	140	0.850
Indie	137	1.000
Disco-Pop	138	0.880
Drum & Bass	138	0.900

Note. Md = Median; FHR = Fetal Heart Rate; UA-PI = Umbilical Artery Pulsatility Index; N = 9 fetuses.

Table 2.2. Wilcoxon signed-rank test results for Fetal Heart Rate (FHR) and Umbilical Artery Pulsatility Index (UA-PI) comparing baseline and music conditions (N=9).

Comparison	Measure	W	p
Baseline vs Indie	FHR	36.5	0.05
Baseline vs Disco-Pop	FHR	31.5	0.03
Baseline vs Drum & Bass	FHR	34.5	0.08
Indie vs Disco-Pop	FHR	34.5	0.08
Indie vs Drum & Bass	FHR	18.0	0.52
Disco-Pop vs Drum & Bass	FHR	14.5	0.84
Baseline vs Indie	UA-PI	12.0	0.12
Baseline vs Disco-Pop	UA-PI	7.0	0.03
Baseline vs Drum & Bass	UA-PI	3.0	0.03
Indie vs Disco-Pop	UA-PI	20.0	0.63
Indie vs Drum & Bass	UA-PI	17.5	0.29
Disco-Pop vs Drum & Bass	UA-PI	13.0	0.26

Note. FHR = Fetal Heart Rate; UA-PI = Umbilical Artery Pulsatility Index; W = Wilcoxon signed-rank test statistic; $\alpha = .05$; N = 9 fetuses.

2.4. Discussion

While literature indicates that music is perceivable in utero (Querleu et al., 1989), research on fetal responses to various musical stimuli remains limited and contradictory due to methodological inconsistencies. In this study, we exposed nine fetuses (33.4 ± 1.4 GA) to three music genres with varying tempos, to assess whether they could discriminate between genres, as reflected by changes in FHR, and whether these stimuli could influence indices commonly associated with fetal well-being, measured through variations in UA-PI.

Fetal heart rate decreased from baseline across all stimuli, with no significant differences between stimuli. However, a more pronounced reduction was observed during the Disco-Pop (114 BPM tempo) and Indie (81 BPM tempo) conditions. These reductions in FHR suggest that, while the fetus does not discriminate between musical genres, it does differentiate between the absence (baseline) and the presence of auditory stimuli. Our general findings contrast with

those of Kisilevsky et al. (2004) and Tristao et al. (2020), who reported increases in FHR in response to music stimuli. In support of our results, Granier-Deferre et al. (2011) found that various acoustic and musical stimuli led to decreased FHR, interpreted as an orienting response indicative of information processing (Graham, 1992). The decrease in our study may also reflect orientation or attention to the different stimuli. Notably, the studies reporting increased FHR (Kisilevsky et al., 2004, Tristao et al., 2024) employed faster tempos derived from accelerated versions of the same musical pieces. In contrast, our stimuli, being three completely different songs, varied not only in tempo but also in genre, encompassing different instruments, harmonies, and melodies. FHR deceleration can be influenced by changes in acoustic structure, as noted by Lecanuet et al. (1992), suggesting that our results may reflect an orienting reflex in response to varying musical stimuli. Additionally, in line with our findings, Groome et al. (1997) also presented auditory stimuli at 80 dB and similarly observed a deceleration in FHR, further supporting the link between low intensity sounds and FHR reduction.

Our aim was also to examine whether the UA-PI, a parameter traditionally used to assess placental blood flow and commonly associated with fetal well-being, would change in response to external auditory stimuli during music exposure. In this study UA-PI, unlike heart rate, showed an increase from the baseline across the music conditions. Although no differences emerged between the stimuli, there was a statistically significant difference between the baseline and both the Disco-Pop and Drum & Bass conditions. Given the limited literature on this topic, it is difficult to directly compare our results with other studies, however, our findings align with Kawakami et al. (2002). In contrast to our results, Gupta and Gupta (2018) observed a decrease in UA-PI, though this occurred after a month of music therapy and involved women diagnosed with PIH, meaning they already had abnormal UA-PI values.

An increase in the UA-PI usually reflects higher resistance in placental circulation (Shulman et al., 1984). It is important to say that in our study, although fetuses UA-PI

increased, the values remained within the normal range for their gestational week (Drukker et al., 2020) not posing any danger to the fetuses' health. This may suggest that the increased vascular resistance, in response to the complex auditory stimuli, was a regulatory transient physiological response rather than an indication of placental compromise, particularly since the fetuses were selected without any pathological condition. A possible explanation for the increase in UA-PI is that the fetus may be redistributing blood flow in response to the auditory stimulus, as a regulatory mechanism to support overall fetal well-being during periods of heightened external stimulation. There is evidence that a higher UA-PI in some cases correlates with a lower Middle Cerebral Artery Pulsatility Index (MCA-PI), suggesting a change in blood distribution in favor of brain circulation in fetuses, commonly referred to as the brain-sparing effect (Sjöström et al., 1997; Malamitsi-Puchner et al., 2006). Although we did not measure MCA-PI, it is possible that the auditory stimulation caused a redistribution of blood flow, leading to an increase in vascular resistance in the umbilical flow and a decrease in resistance in the middle cerebral artery, prioritizing blood flow towards the brain for sensory processing and overall fetal well-being. This hypothesis is further reinforced by studies that have demonstrated changes in MCA-PI in response to airborne sounds (Jeličić et al., 2023).

In summary, no significant differences were found between the different musical stimuli for both FHR and UA-PI, rather changes were only observed between the absence and presence of stimulation. The observed decrease in heart rate may be indicative of an orienting response, signaling not only attention to the external stimuli but also an early capacity for auditory discrimination, confirming the idea that this ability begins in utero. Concurrently, the increase in UA-PI highlights a potential link to fetal well-being through the modulation of blood flow, likely in support of enhanced sensory processing, indicating that music stimulation can induce measurable physiological changes including circulatory parameters.

The main strength of this study is its contribution to the limited body of research on

physiological responses to contemporary music, especially regarding alterations in blood flow patterns. To our knowledge, few studies have directly examined fetal auditory discrimination together with circulatory indicators of well-being in response to different musical genres. This study has limitations, particularly the small sample size and the absence of a non-music control condition, which limit the generalizability of the findings. Nonetheless, the results provide an initial perspective on fetal physiological and circulatory responses to multiple musical stimuli, emphasizing the need for larger-scale investigations.

Evidence indicates that music interventions in neonatal intensive care units (NICU) can support physiological stability in preterm infants (Yue et al., 2021). Although our study focused on fetuses, the observation that music influences FHR and UA-PI in utero is consistent with this clinical evidence and suggests that prenatal music exposure may contribute to early regulatory processes that are fundamental for development.

2.5. Conclusions

Despite being a preliminary study, our findings provide valuable insights into the fetal response to musical stimuli, highlighting cardiac and vascular reactions. Given its impact on sensory processing and physiological markers of well-being, music as prenatal stimulation may offer new perspectives on early development, potentially fostering optimal fetal growth while paving the way for sensitivity to suprasegmental structure.

CHAPTER 3 | The Development of Prosody in Pre-lexical Vocalizations: A Longitudinal Study Over the First 16 Months of Infant Life.

D'Aloia V., Zanchi P., Logrieco M.G., Passaquindici I., Palumbo R., Lionetti F., Spinelli M., Fasolo M. *(under 2nd review).* The developing of prosody in infants: A longitudinal study over the first 16 months of infant life. F1000 Research. https://doi.org/10.12688/f1000research.154114.1.

Abstract

Early prosodic development plays a crucial role in the emergence of communicative abilities, yet evidence on pitch-related features in pre-lexical vocalizations remains limited and methodologically heterogeneous, particularly for Italian. The present longitudinal study investigated the development of F0-related prosodic features in the pre-lexical vocal productions of Italian infants between 4 and 16 months of age, with particular attention to individual variability and production type. Fifteen typically developing infants were recorded during repeated mother–infant face-to-face interactions, yielding a large corpus of pre-lexical vocalizations classified into distinct production types. Fundamental frequency mean, F0 range, and F0 final contour direction were acoustically analyzed. Developmental trajectories were examined using multilevel modelling to account for the hierarchical structure of the data, with productions nested within infants. Results revealed substantial inter-individual variability across all prosodic measures. In addition, different types of vocal production were associated with distinct prosodic profiles, which were observable from the earliest occurrences of each production type. Analyses of F0 final contours indicated a predominance of productions with flat intonation and differences across production types. Together, the findings highlight the importance of considering both individual variability and vocal production type when examining early prosodic development. By focusing on pre-lexical vocalizations, this study provides new evidence on the organization of pitch-related prosodic features prior to the

emergence of stable lexical forms.

Keywords: *Prosody development, pre-lexical vocalization, fundamental frequency (f0), longitudinal acustic analysis, developmental trajectories.*

3.1. Introduction

Infants' early vocal productions form the basis for language development and are important social signals that stimulate caregiver response, which in turn facilitates the development of phonology and speech (Goldstein et al. 2009; Spinelli et al., 2017). Research on pre-linguistic vocalizations of typically developing infants has focused mostly on the segmental and syllabic properties of spontaneous vocal productions. Fewer scholars have focused on suprasegmental features of infant vocalizations, particularly pitch-related prosodic measures, most commonly assessed using fundamental frequency (F0). These prosodic features of language convey grammatical and pragmatic meaning as well as emotional or affect intent, and highlight particular elements of the utterance (Crystal, 1978; Gleitman et al., 1988; Wells et al, 2004). The development of prosody is of special interest within the larger context of language acquisition , because it is among the earliest aspects of speech to be acquired (Esteve-Gibert & Prieto, 2018) and is strictly linked to the maturity of the vocal tract (Kahane & Kahn, 1984). It is well known that infants are sensitive to prosodic information from a very early age due to their prenatal exposure to speech (Gervain, 2015) and therefore are able to understand the prosodic features of speech since the first months of life (De Carvalho et al., 2019; Fisher & Tokura, 1996; Soderstrom et al., 2008). Moreover, they are able to imitate the prosodic characteristics of caregiver speech, for instance by adapting their F0 patterns to match those of adults (Gratier & Devouche, 2011; Ko et al., 2016; McRoberts & Best, 1997; Papoušek & Papoušek, 1989).

At birth, the infant vocal tract is not fully developed; consequently, infant vocal productions differ from adult speech in both their sound structure and intonational patterns

(Astruc et al., 2013). Some authors reported much more reduced pitch range in children than in adults, and that children acquire falling tones first and final rising tones much later (Snow, 2004; Wells et al., 2004). Differences between infants and adults observed in the early stages of phonetic development have therefore been linked to immature control of the vocal tract. Specifically, Lieberman (1967) assumed that infants do not precisely control the tension of their laryngeal muscles once phonation starts. The child primarily maintains laryngeal muscle tension at or near the level established at phonation onset, with tension gradually increasing from the onset of phonation to a relatively level or slightly falling plateau. For this reason, the majority of breath groups terminate with falling fundamental frequency contour. However, cross-linguistic differences have been reported, with rising contours observed in infancy in languages other than English (Esteve-Gibert & Prieto, 2013; Prieto et al., 2012; Whalen et al., 1991).

To our knowledge, only limited data are currently available on the prosodic characteristics of infants' earliest vocal productions in Italian. One of the first studies in this area was conducted by D'Odorico (1984), who acoustically analyzed cry and non-cry vocalizations produced by four Italian infants. The results of this study showed that cry vocalization produced in different context were acoustically different and that cry and non-cry vocalizations produced within the same context shared similar acoustic properties. These results were later confirmed by D'Odorico and Franco (1991), who investigated the suprasegmental features of the vocalizations produced during social interaction by five Italian children aged 4 to 11 months. Indeed, distinct patterns of suprasegmental features were observed across different interactional contexts

Considering the scarcity of data on Italian and the importance of cross-linguistic evidence in early language development, the main aim of the present study is to investigate the developmental trajectory of pitch-related prosodic properties in Italian infants' pre-lexical

vocalizations during the first 16 months of life. Specifically, we focus on spontaneous non-lexical vocal productions in order to examine early intonational patterning before the emergence of stable segmental structure and lexical contrasts, thus minimizing confounds related to articulatory maturity and linguistic planning. Infants' pitch is analyzed in terms of mean fundamental frequency (F0), F0 range, and F0 final contour types. Unlike previous studies that primarily compare infants' productions to an adult prosodic model, the present study adopts a developmental perspective, aiming to provide a detailed longitudinal description of how infants' intonational repertory evolves across pre-lexical vocalizations from 4 to 16 months of age.

3.1.1. Prosodic development over infancy and toddlerhood: a contrasting picture.

Table 3.1 summarizes the main studies on intonational development in the first two years of life, with a specific focus on F0-related measures (mean F0, F0 range, and final F0 contours) in non-distress spontaneous vocalizations. Studies examining distress-related productions, such as crying or whining, are excluded (Mampe et al., 2009; Rothgänger, 2003; Zanchi et al., 2016).

Among the F0-related prosodic variables, the most extensively studied measure is the fundamental frequency mean value (F0 mean), which reflects the overall pitch level of the voice and is determined by the rate of vocal fold vibration within the larynx (Hirst & De Looze, 2021). One of the earliest reviews on the developmental change in mean fundamental frequency (F0) from birth to adulthood was provided by Kent (1976), who reported that F0 is highest at birth and progressively decreases toward adult values. However, the review was based on a very limited number of studies. Some of the studies reported in Table 3.1 showed a similar descending pattern in mean F0 from birth to the second year of life (Amano et al., 2006; Robb & Saxman, 1985; Rothgänger, 2003). Flax, Lahey, Harris, and Boothroyd (1991) found

that 2 out of 3 children showed a reduction in F0 mean between the onset of words and 50-words period. This decrease is explained as reflecting the gradual maturation of the vocal tract and the development of infant ability to control the voice.

Nonetheless, other scholars failed to find significant changes in F0 mean over time (Iyer & Oller, 2008; Robb et al., 1989). Laufer and Horii (1977) measured infants' productions every two weeks during the first six months of life and found that F0 mean values slightly fluctuated. Amano et al. (2006) argued that the main reason why some studies did not find a decrease in the F0 mean values was that speech samples were collected over periods that were too short to catch this effect. Lastly, to our knowledge, only one author reported the opposite pattern. Fairbanks (1942) observed an increase in mean F0 during the first five months of infant life, followed by a period of stabilization.

Even more contrasting findings have been reported with respect to the developmental trajectories of F0 range, defined as the difference between the maximum and minimum pitch values produced within an utterance (Ladd, 2008). F0 range represents the ability to vary the intonation and to make the production more communicative and attractive for the listener (Amano et al. 2006). In infants Amano and colleagues (2006) found an increase of F0 range after the onset of two-word utterances and hypothesized that, as children grow older, they acquire the ability to vary F0 within a production in parallel with the development of their communicative abilities.

Similarly, Snow (2004) found that 4-years-old children showed a wider F0 range than 1-year-old infants. Nonetheless Snow and Ertmer (2012) reported that, in their sample, 10 out of 12 typically developing infants showed a decrease in F0 range between 3 and 9 months. In an initial study, Robb and Saxman (1985) reported a decrease in inter-utterance F0 range between 11 and 25 months, whereas in a subsequent study (Robb et al., 1989) they failed to find changes over time. Furthermore, Laufer and Horii (1977) found a slight decrease in within-utterance F0

range during the first four months of life, followed by a minor increase after the fourth month. Overall, the developmental trajectory of within-utterance F0 range remains unclear.

Another prosodic variable that plays an important role in prosodic development is F0 contour, intended as the shape of pitch (F0) variations within the utterance, which conveys a specific melody to the vocal production (Sorianello, 2006). F0 contour may increase from the beginning to the end of the production (rising F0 contour), decrease (falling F0 contour), or not significantly vary (flat F0 contour). The direction of these contours, particularly at utterance-final position, is considered crucial for conveying the pragmatic meaning of a vocal production. In most languages, falling final contours are typical of declaratives (i.e. statements), while rising final contours are especially used for interrogatives (for Italian see for example Savino, 2012). Many scholars agree that infants begin to use final F0 contours at an early stage to express communicative intentions. For example, Prieto et al. (2012), Prieto and Vanrell (2007), and Esteve-Gibert and Prieto (2013) provided evidence that, from the first year of life, Catalan and Spanish infants, similarly to adults, are able to vary final F0 contours to signal pragmatic meanings, even before they can produce words.

Several studies focusing on infants exposed to the English language, agree that the majority of the productions of infants exhibit falling contours (Fox, 1990), whereas rising contours are rare (see for example Kent & Murray, 1982, and Robb et al., 1989) and become more frequent in adults (Cruttenden, 1997) and preschoolers (Snow, 2004).

According to Lieberman (1967), the falling contour is considered more natural and simpler than the rising contour without implying the infant's intentionality. In line with the author's hypothesis, rising patterns, by being contingent on language experience rather than physiological constraints, stabilize at a later stage of language acquisition. More recently, Snow (2006) found that the development of falling and rising patterns is not linear and follows a U-shaped trajectory. Both falling and rising contours are frequent and well expressed until nine

months, when their robustness decreases. After a period of stabilization, both these contours reach the same configuration observed before the regression period at around 18 months. The author explained this results pointing out that intonation is controlled by physiology in the earliest stage (before nine months of age), whereas later tonal patterns come under linguistic control. Consequently, the observed regression may reflect a process of linguistic reorganization, with the U-shaped trajectory indicating a shift in intonation from a pre-intentional to an intentional stage.

Despite these two main theoretical explanations, empirical studies examining changes over time in the frequency of final F0 contour trajectories have reported inconsistent results. Many of these studies (see, for example, Flax et al., 1991; Murry et al., 1983) failed to find variations in the relative proportions of rising and falling contours across development. Fox (1990) consistently found a predominance of falling final contours (82%) between the 3rd and the 9th month of age. By contrast, Robb et al. (1989) reported that the least frequently occurring final F0 contours were falling–rising and rising patterns, which together accounted for approximately 6% of all vocalizations; this proportion remained stable throughout the first two years of life, independently of vocabulary size.

Table 3.1. Literature review on F0-related (F0 mean, F0 range, F0 final contours) prosodic development.

Study	*Design*	*Type of productions*	*Lang.*	*N*	*Age*	*Frequency*	*Mean F0 values*	*Mean F0 changes*	*F0 Range/ F0 SD changes*	*Final contours changes*
Amano et al., 2006	L	All	J	3	0-5years	Monthly	Mean=385	Decrease	SD increase after the two-word period	
Fairbanks, 1942	L	Hunger wails	AE	1	1-9months	Monthly		Increase till 4 months and then stabilized	Increase till 5 months and then stabilized	
Flax, 1991	L	All	AE	3	before one-word (11-14m), 10-word voc. (15-16m),	3 times		Increase for 2 children		No changes

					50-words voc. (18-22m)					
Fox, 1990	L	All	AE	12	3-9months	Monthly		No changes	No changes	No changes
Iyer et al., 2008	L	Syllables		8 TD, 8 IHL	pre-canonical 4.34m, canonical 7.07m, post-canonical 9.88m			No changes	No changes (SD within)	
Kent et al., 1982	C-S	Non-distress vocalizations	AE	21	3,6,9 months	Every 3 months	3m=445 6m=450 9m=410	No changes		No changes
Laufer et al., 1977	L	Non-distress productions	AE	4	0-6months	Two weeks	335	No changes	SD decrease until 2 months and then slight increase	
McRoberts et al., 1997	L	All	AE	1	3-17 months	Weekly	7m=407 15m=330 17m=326	Decrease		
Murry et al., 1983	L	Cry and Non-distress productions	AE	1	Non-distress productions at 8, 10, 12 weeks	Two weeks	3m=445 6m=450 9m=415	Increase from the 8th to the 10th week and decrease after	Increase from 8 to 12m	
Robb et al., 1985	C-S	All	AE	14	11-25months	Every 3 months	Mean=357 11-13m=400 23-25m=314	Decrease	Decrease	
Robb at al., 1989	L	Mono and bi-syllables	AE	7	8-27m with different starting point	Monthly		No changes	No changes	No changes
Rothgänger, 2003	L	Babbling and cry	G	15	0 to 12 months		1m=389, 12m=336	Decrease		
Snow, 2004	C-S	All	AE	20	1and 4 years old				Increase	
Snow et al., 2012	l	Monosyllable	AE	12 TD, 18 CI	3,6,9 months	Every 3 months			Decrease	

Note: for studies that compared typical and not-typical children only results on typical children are reported, for studies that explored also cry patterns only findings on non-distress

productions are reported; Lang = Language, J = Japanese, AE = American-English, G = German, m = month, L = longitudinal design, C-S = cross-sectional design, TD = typically development children, IHL = hearing loss infants, CI = children with cochlear implantation; mean F0 values are in Hertz.

3.1.2. Suggestions for a better comprehension of the phenomena.

Overall, studies examining the development of F0-related features during the first years of life have failed to provide a clear and homogeneous picture. This lack of convergence is largely due to the substantial heterogeneity across studies, which makes direct comparisons difficult. First, cross-linguistic differences may lead to distinct intonational patterns and developmental trajectories. Second, studies vary widely in terms of the age ranges considered, the number and timing of longitudinal observations, and other key methodological choices. Differences concern the types of infant vocal productions included in the analyses (e.g., whether analyses focus exclusively on syllable-like vocalizations or also include cry-like or squeal-like productions), as well as the overall study design, with longitudinal and cross-sectional approaches potentially producing divergent results due to between-participant variability. Additionally, several studies have highlighted substantial inter-individual variability among infants, affecting both F0 values and age-related changes in vocal production. In particular, Laufer and Horii (1977) reported that mean F0 fluctuated from month to month, with each infant exhibiting a distinct change pattern. Flax et al. (1991) found that all three infants showed different F0 range change patterns, with one child showing no variation at all. Kent and Murray (1982) found a predominance of falling F0 contours across the age range considered (3, 6, and 9 months), but also reported high intra- and inter-individual variability in the production of rising contours, both at a given time point and across development. Other authors also reported the presence of inter-individual differences among infants (Amano et al., 2006); however, the limited number of participants or recording sessions made it difficult to adequately test this hypothesis. This has important implications for study design and data analysis, highlighting the need for larger samples and for explicitly accounting for participant-level variability.

Available studies (Table 3.1) mainly considered all the productions or only one specific type of production, without making comparisons among them. Robb et al. (1989) recorded the utterances of seven infants on 12 successive occasions between 8 and 26 months of age, examining mean F0 in monosyllabic and bisyllabic utterances. They found similar mean F0 values across the two production types but observed a tendency for monosyllables to exhibit a greater F0 range than disyllables in all participants except one child, with this tendency remaining stable across the first two years of life. By contrast, Snow (2004) did not find an effect of the number of syllables on F0 range, suggesting that pitch range is independent of utterance length. However, these investigations focused on syllable count within utterances rather than on developmentally distinct types of vocal productions. Rothgänger (2003) examined the prosodic development of babbling by comparing it with the development of crying rather than with other non-distress vocal productions; as a result, information about potential differences among non-lexical production types remains limited. Other studies have highlighted the importance of considering different types of productions separately, reporting distinct developmental trajectories for prosodic features in vocalizations versus syllabic utterances (Hsu et al., 2000), as well as across different types of multiword combinations (Behrens & Gut, 2005). Overall, these findings underscore the need for a more detailed and systematic description of developmental prosodic phenomena.

3.1.3. The present study

Some authors have hypothesized the existence of linguistic trade-offs during development, so that the increased demands in one component of language, such as syntax, may potentially cause a decreased performance in a second component, such as phonology (Crystal, 1978). Furthermore, the interrelationships among different components of language may vary depending on how recently a given linguistic structure has been acquired (Crystal, 1978; Masterson & Kamhi, 1992; Zanchi et al., 2016). At early stages of development, a similar effect

may also be observed across different types of infants' pre-lexical productions. Accordingly, we hypothesize that pre-lexical productions acquired earlier, and thus more practiced at the oral-motor level (Oller et al., 1976; Stoel-Gammon, 2011), may exhibit prosodic features that differ from those of later-acquired productions, for which infants are not yet fully fluent.

To assess this hypothesis, the first aim of the present study is to explore the development of F0-related prosodic features of different types of pre-lexical productions observed. To our knowledge, this factor has not been systematically investigated so far. As noted above, previous work has focused either on differences in prosodic development between crying and other non-distress productions (Murry et al., 1983) or on differences among productions varying in syllable number (Robb et al., 1989). The second aim of the present study is to investigate individual variability in prosodic features of infant speech. Given previous evidence of substantial variability in F0-related measures and developmental trajectories across infants (e.g., Laufer & Horii, 1977; Flax et al., 1991; Kent & Murray, 1982; Amano et al., 2006), all analyses were conducted using multilevel modelling, with observations nested within children. We expected to find:

- different trajectories of prosodic features of each kind of production considered, with earlier-acquired productions exhibiting prosodic characteristics that differ from those of later-acquired productions.
- significant individual variability among children.

3.2. Methods

3.2.1. Participants

Fifteen infants (3 females) participated in the study. The sample was not gender balanced. Some studies have reported sex-related differences in speech motor development, with boys showing a slower maturational course than girls from early childhood through adolescence

(Smith & Zelaznik, 2004). However, previous research on prosodic aspects of language development has not consistently documented sex-related differences in early developmental stages, with such differences typically reported only from late puberty onward (Bennett, 1983; Fox, 1990; Lee et al., 1999). All infants were full-term and reported to be in good health at the time of recording. Health status was assessed based on parental report and medical history, with no known neurological, sensory (hearing or vision), or developmental disorders.

All families were monolingual Italian-speaking. Infants were exposed to Italian as spoken in Northern Italy, with no regular exposure to other languages or regional dialects. Mothers' mean age was approximately 35 years (range: 28-42). Sixty percent of mothers had completed high school education, and 40% held a university degree.

3.2.2. Procedure

Families were recruited shortly after the infants' birth through pediatric practices in the local area, via flyers distributed in pediatric clinics and through word-of-mouth contacts. Mothers were contacted after delivery and, for those who expressed interest in participating, the first laboratory visit was scheduled at the beginning of the infant's fourth month of life (mean age = 4 months, 2 days; SD = 5 days).

This age was selected because the period between 4 and 6 months is characterized by significant anatomical and physiological changes in the infant vocal tract, which lead to increased control over speech articulation and to the emergence of more speech-like vocal productions (Kent, 1976). The study was approved by the Ethics Committee of the Department of Neuroscience, Imaging and Clinical Sciences of the University of Chieti–Pescara (approval number: DNISC2962, 06.11.2019) and was conducted in accordance with the American Psychological Association guidelines and the principles of the 1964 Helsinki Declaration. Written informed consent for participation in the study was obtained from all mothers prior to data collection.

Infants were followed from the 4th to the 16th month of age. From the 4th to the 14th month a researcher visited the infant and the mother at home every 15 days (twice a month); after the 14th-month visits were monthly. See Table 3.2 for a summary of the sessions recorded for each participant. Not all planned sessions were completed for all infants. Some sessions were missed for personal or practical reasons, and some participants did not continue participation up to 16 months of age. As a result, the dataset includes missing observations, which are indicated as missing data (m) in Table 3.2. During each visit, mother and infant were audio-video recorded during a face-to-face interaction, which included spontaneous free-play and routine activities (e.g., diaper changing). To ensure comparability across sessions and participants, interactions were conducted without toys. This choice was motivated by the fact that infants' play at these ages typically involves sound-producing objects, which could interfere with the acoustic quality of the recordings. As a consequence, mothers were instructed to interact with their infants as they normally would in everyday face-to-face situations, in the absence of objects. In total, 295 recording sessions were collected (mean per infant = 20, SD = 5), with an average duration of approximately 10 minutes per session (M = 10.06 minutes, SD = 1.99).

Table 3.2. Number of pre-lexical productions by child and session. *Note: m = missing data*.

Children																	
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
Month	n																Total
4°	1	44	57	17	31	m	m	46	17	24	14	21	m	m	m	m	271
4b	2	57	28	11	72	64	71	44	4	29	6	54	2	6	m	82	530
5°	3	64	12	21	29	38	72	47	23	21	14	16	32	m	m	19	408
5b	4	81	46	29	54	65	42	23	6	39	12	30	76	21	61	24	609
6°	5	27	125	1	50	31	32	51	23	44	6	28	7	73	m	86	584
6b	6	112	76	6	29	41	60	3	28	43	7	41	27	26	m	22	521
7°	7	48	26	7	40	91	33	20	9	12	9	37	37	6	94	34	503
7b	8	61	87	19	102	51	40	m	m	16	6	19	7	20	52	43	523
8°	9	81	103	10	41	18	43	17	29	19	14	28	30	57	41	58	589
8b	10	77	29	7	86	31	43	m	14	18	6	21	61	20	35	22	470
9°	11	68	53	18	66	35	42	m	6	16	19	8	32	78	41	102	584
9b	12	36	63	25	38	28	36	m	m	24	9	26	52	10	15	100	462
10a	13	49	59	19	31	22	37	m	18	8	11	6	32	24	16	102	434
10b	14	62	58	33	36	27	23	77	m	29	13	23	21	14	14	31	461
11a	15	72	45	7	38	48	49	21	m	19	17	28	50	m	22	54	470
11b	16	33	88	9	37	24	47	47	m	40	5	3	24	m	m	11	368
12a	17	28	57	20	34	53	58	19	m	12	24	14	42	19	m	8	388
12b	18	32	32	5	55	21	63	20	m	20	22	6	16	m	m	m	292
13a	19	17	30	7	39	15	27	m	m	14	m	22	8	48	m	m	227
13b	20	29	11	16	66	19	21	m	m	m	m	9	14	m	m	m	185
14a	21	17	39	14	20	42	14	12	m	m	18	43	m	34	m	m	253
14b	22	28	27	13	55	20	34	m	m	12	m	1	m	21	m	m	211
15	23	26	28	10	40	21	29	m	m	m	27	44	16	34	m	m	275
16	24	m	m	m	17	12	25	20	m	14	11	11	9	m	m	m	119
Total		1149	1179	324	1106	817	941	467	177	473	270	539	595	511	391	798	9737
Total m		1	1	1	0	1	1	9	13	3	3	0	3	7	14	8	65

3.2.3. Coding: pre-lexical productions

The audio tracks from all recorded sessions were extracted using Audacity (Version 3.2.5; Audacity Team 2023). All audible pre-lexical infant vocal productions that did not overlap with other sounds and were free from background noise were selected and coded for analysis. In line with previous studies, vegetative sounds (such as wheezes, sneezes, coughs, hiccups, and clicking sounds), stress vocalizations (such as whimpering, fusses, and cries), laughs, words and onomatopoeias were not considered (Fasolo et al., 2010).

Infants' vocal productions (Stark et al., 1993) were coded as:

- Communicative grunt (g): short consonant-like vocalizations (e.g., [m], [ŋ]) produced in the absence of distress and occurring in an interactional context (e.g., during face-to-face interaction with the caregiver). These vocalizations were distinguished from vegetative sounds and cries based on their communicative use (Camaioni, 1992).
- Vocalic sound (v): vocalization constituted by vowel-like sounds (e.g., [a]);
- Simple babbling (cv): vocalizations containing almost one full vowel-like element and one consonant-like element with rapid transition between consonant and vowel (e.g. [ba]);
- Reduplicated babbling (cvcv): vocalization containing rapid repetition of the same sequence of one full vowel-like element and one consonant-like element (e.g., [baba], [tata]);
- Variegated babbling (c1vc2v or cv1cv2): vocalization containing rapid repetition of different sequences; vocalizations comprising at least one full vowel-like element and at least two different consonant-like elements (e.g., [bata]), or two different full vowel-like elements and one consonant-like element (e.g., [beba]).

A total of 9,737 productions were coded (mean per session = 33, SD = 23.38; mean per infant = 645, SD = 329.40). In all tables and figures of the present study, pre-linguistic productions

are labeled as follows: Grunt = communicative grunts; Voc = vocalizations; Babb1 = simple babbling; Babb2 = reduplicated babbling; Babb3 = variegated babbling. Table 3.2 summarizes the total number of productions for each participant at each age.

3.2.4. Coding: Prosody

Prosodic characteristics of each vocal production were extracted using Praat (Boersma & Weenink, 2009). Production boundaries were identified through visual inspection of the waveform and spectrogram in order to determine the onset and offset of each vocalization (D'Odorico et al., 2009). The following measures were then calculated for each production:

- Fundamental frequency mean (F0 mean): automatically extracted in Hertz (Hz) by PRAAT.
- Maximum and minimum pitch: the highest and lowest F0 values in the vocal production (Cruttenden, 1997), expressed in Hz.
- Fundamental frequency range (F0 range): the extent of F0 variation across the entire pre-lexical production, expressed in semitones. Following Snow and Balog (2002), F0 range was calculated as the logarithmic difference between the maximum and minimum F0 values within a production, according to the formula: [12/log(2)]*[log(maximum F0/minimum F0)].
- F0 final contour: the direction of the last movement in the intonational profile of the production. Each vocal production was classified as having either a rising final contour (F0 final rising contour) or a falling final contour (F0 final falling contour) when the change in F0 between the minimum and maximum values within the production was at least two semitones. Productions with an overall F0 range of less than two semitones were classified as having a level F0 contour (flat contour).

3.2.5. Reliability

Inter-coder reliability was assessed on 20% of the recording sessions randomly selected at each age point. Cohen's kappa coefficient was calculated for the coding of vocal production categories, yielding a value of .93, indicating excellent agreement. The Cohen's K coefficient on the classification of edge F0 final contours was .83.

3.2.6. Plan of Analyses

Statistical analyses were conducted in successive steps. First, descriptive statistics were used to quantify the frequency of each type of pre-lexical vocal production across the observation period. Then analyses were performed to examine developmental changes in pitch-related prosodic features across the first 16 months of life, as well as differences among types of pre-lexical vocal productions and individual variability among infants. Given the longitudinal and hierarchical structure of the data, with repeated vocal productions nested within infants, multilevel modelling was adopted throughout the analyses. Vocal productions constituted Level 1 observations, while infants constituted Level 2 units and were treated as a random factor. For continuous prosodic measures (F0 mean and F0 range), linear multilevel models were fitted. Analyses followed a sequential model-building strategy: an unconditional model with no predictors was first estimated to assess between-infant variability (M0); age was then included as a fixed predictor, both as a linear and a squared term, to model potential non-linear developmental trajectories (M1); subsequently, type of vocal production was added as a fixed effect with *vocalization* as reference category (M2); finally, interactions between age (linear term) and type of production were tested to evaluate whether different production types followed distinct developmental trajectories (M3). For categorical prosodic measures (F0 final contours), separate multilevel logistic (binomial) models were estimated for level, rising, and falling contours, using the presence versus absence of each contour type as the dependent

variable. The same modelling strategy was applied. Model comparisons were performed using likelihood ratio tests based on changes in -2 log-likelihood (-2LL), and more complex models were retained only when they provided a significant improvement in fit. Statistical significance of fixed effects was evaluated using the ratio between parameter estimates and their standard errors, with effects considered significant ($p < .05$) when the absolute value of the estimate was approximately twice its standard error ($|\text{estimate}| \geq 2 \times \text{SE}$). In the random part of the models, *u* represents between-subject variance and *e* within-subject (residual) variance. All multilevel analyses were conducted using MLwiN version 2.33 (Rasbash et al., 2005).

3.3. Results

Table 3.3 reports the total frequency of each type of vocal production and its percentage relative to the total number of productions.

Table 3.3. Frequencies and percentages of each vocal production.

Type of production	Frequency	% of total
Grunt	2523	26.1
Vocalization	4955	51.2
Simple Babbling	1451	15.0
Duplicate Babbling	592	6.1
Variate Babbling	155	1.6
Total	9676	100

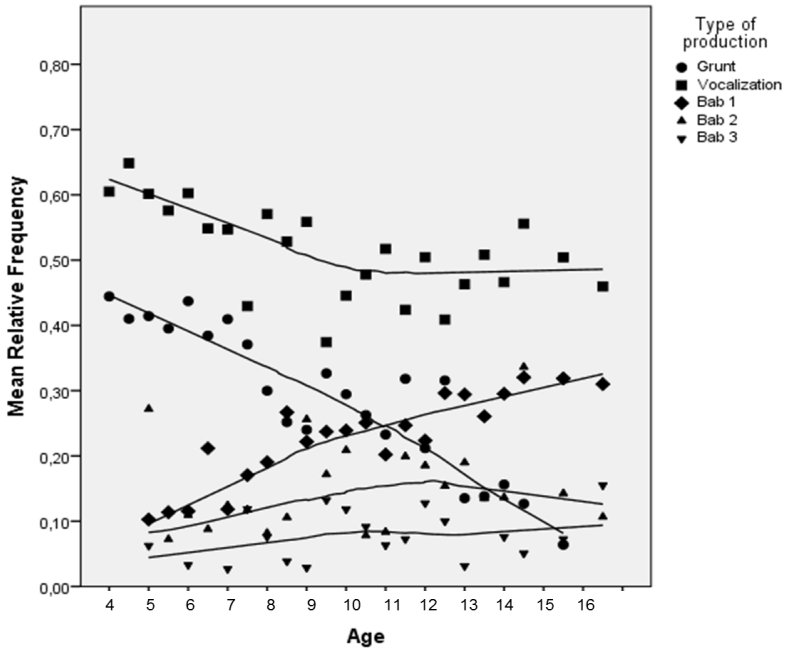


Figure 3.1. Mean relative frequency of the type of productions from 4 to 16 months of age. *Note: fit lines are computer with Kernel method on 50% point fit.*

Figure 3.1 reports the mean relative frequencies of each production type, aggregated within participants at different ages. The graph shows that vocalizations were the most frequent productions over time, with a consistent decrease during the first ten months of life. Grunts were very common during the first seven months, but their frequency decreased over time, becoming increasingly sporadic. All three types of babbling began to appear between the 5th and 6th months, and their frequency increased over time, with a predominance of simple babbling across the entire age period considered.

3.3.1. Between subjects' variability exploration

To explore the presence of between-subjects variability, a linear regression was conducted with age, age squared, and infants (indicators) as predictors on the dependent variable F0 mean. The results were statistically significant, R^2 = .102, indicating that infants and age influence F0 mean. Coefficients reported in Table 3.4 show that infants had significantly different coefficients. This confirms the presence of differences among infants and that these differences

should be taken into consideration.

Table 3.4. Linear regression on F0 mean by children (dummies) and age.

	b	*S.E*	*t*	*p*
Age	1.3	.48	2.6	<.01
Age^2	-0.7	.52	-1.4	.15
Intercept	345.7	3.2	104.1	<.01
Child 2	46.9	3.2	14.6	<.01
Child 3	-.1	4.9	.0	.98
Child 4	2.1	3.2	.6	.52
Child 5	3.6	3.5	1.0	.30
Child 6	23.3	3.4	6.8	<.01
Child 7	-1.4	4.2	-.3	.75
Child 8	1.4	6.3	.2	.82
Child 9	46.2	4.2	10.8	<.01
Child 10	37.8	5.2	7.2	<.01
Child 11	-36.7	4.0	-9.0	<.01
Child 12	41.4	3.9	10.4	<.01
Child 13	5.7	4.1	1.4	.17
Child 14	1.9	4.6	.3	.67
Child 15	-35.1	3.6	-10.1	<.01

Note. Child 1 as a reference category, all the other coefficients should be calculated as differences from the intercept.

3.3.2. F0 mean trajectory over time

The first set of analyses examined changes in F0 mean over time and across different types of vocal productions. A baseline unconditional model with no predictors (M0) was first fitted to assess the presence of inter-individual variability in mean F0. This model revealed significant between-infant variability (Table 5; u = 2081.83, SE = 137.62), therefore, all subsequent models included infant as a random effect. The overall mean F0 across infants and ages was

366 Hz (intercept = 366.32, SE = 0.89), in line with previous findings (e.g., Amano et al., 2006). Age (linear and squared terms) was then added as a fixed predictor (M1). As shown in Table 3.5, only the linear effect of age was significant (β = 1.60, SE = 0.49), indicating that F0 mean values increased over time.

Model 2 (M2) added type of production as a fixed predictor, with vocalizations as the reference category. Results confirmed the significant effect of age (β = 1.40, SE = 0.49) and showed that grunts had the lowest mean F0 values (β = -16.48, SE = 2.04), whereas variegated babbling exhibited the highest values (β = 14.52, SE = 6.70). Vocalizations, simple babbling (β = -5.27, SE = 2.52), and reduplicated babbling (β = -0.77, SE = 3.59) showed similar intermediate F0 mean values. The absence of differences between monosyllabic and disyllabic productions (simple vs. reduplicated babbling) is consistent with previous findings (Robb et al.,1989).

Model 3 (M3) included the interaction between age and type of production. Although the effects of age (β = 1.62, SE = 0.56), age squared (β = -1.13, SE = 0.58), and the interaction between reduplicated babbling and age were significant (β = 1.31, SE = 0.65), the decrease in -2LL was not statistically significant (-2LL = -6.77, df = 10). Therefore, Model 3 did not provide a better fit than Model 2, which was retained as the best representation of F0 mean development and is illustrated in Figure 3.2. Figure 3.2 highlights the differences in intercepts across production types, indicating stable differences in mean F0 levels over time, while the overall shape of the developmental trajectory is shared across productions and reflects the inclusion of linear and quadratic age terms.

Table 3.5. Fixed and random effects from multilevel models predicting F0 mean across age and types of pre-lexical vocal productions.

	M0		M1		M2		M3	
	b	s.e.	b	s.e.	b	s.e.	b	s.e.
Fixed Effects								
Constant	366.32	0.89	354.59	2.63	361.90	2.76	360.58	3.13
Age	--	--	1.60	0.49	1.40	0.49	1.62	0.56
Age^2	--	--	-0,93	0,51	-0,96	0,51	-1,13	0,58
Productions; base vocalization								
Grunt	--	--	--	--	-16,48	2,04	-13,21	3,85
Babb-1	--	--	--	--	-5.27	2.52	-0.36	6.83
Babb-2	--	--	--	--	-0.77	3.59	-19.41	9.71
Babb-3	--	--	--	--	14.52	6.70	24.84	23.20
Age*Grunt	--	--	--	--	--	--	-0.38	0.37
Age*Babb-1	--	--	--	--	--	--	-0.35	0.45
Age*Babb-2	--	--	--	--	--	--	1.31	0.65
Age*Babb-3	--	--	--	--	--	--	-0.69	1.43
Random effects (level 2)								
u	2081.83	137.62	2069.24	137.62	2007.62	136.07	2005.86	136.07
Random effects (level 1)								
e	4573.41	130.36	4561.20	130.15	4564.56	129.50	4561.65	129.55
Model Fit statistics								
-2LL	--	--	-34.8	2 df	-74.19	6 df	-6.77	10 df

Note. Roughly if $|2 \times SE| \leq |parameter|$ then $p < .05$; u = between variance, e = within variance.

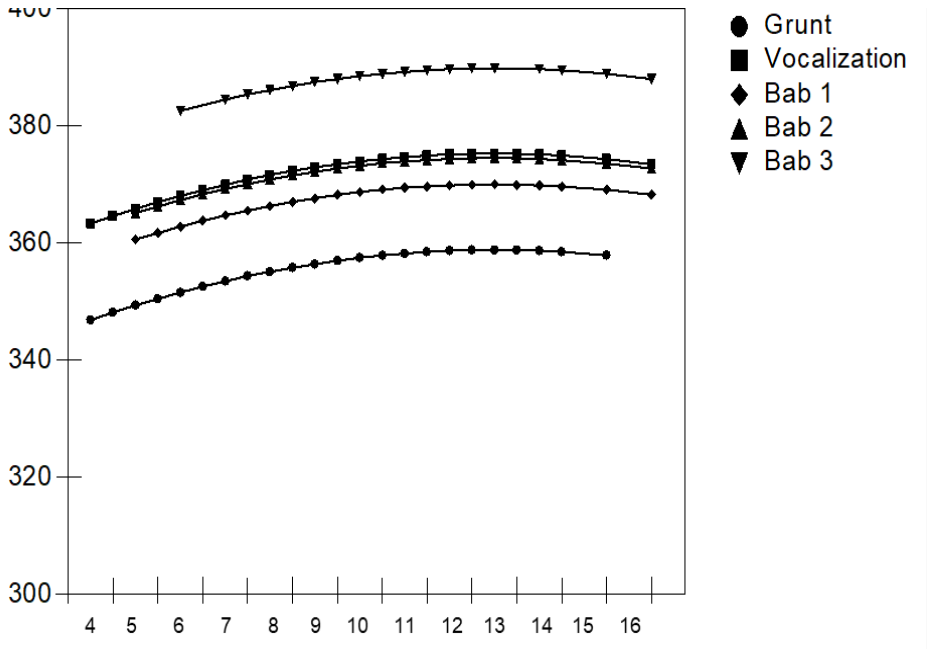


Figure 3.2. F0 mean values (in Hz) of each type of production from 4 to 16 months of age predicted by Model 2 (M2 in Table 3.5).

3.3.3. F0 range trajectories over time

The second set of analyses examined changes in F0 range, measured in semitones, over time and across types of vocal productions (see Table 3.6). The baseline model (M0) yielded a mean F0 range of 4.47 semitones (*intercept* = 4.47, SE = 0.04) and revealed significant variability among infants (u = 12.96, SE = 0.23). In Model 1 (M1), a significant effect of age indicated that F0 range increased over time (β = 0.01, SE = 0.01). The inclusion of type of production as a fixed predictor, with vocalizations as the reference category, in Model 2 (M2) showed significant effects of age (β = 0.00, SE = 0.01), age squared (β = -0.06, SE = 0.02), and of all production types except reduplicated babbling. The effect of age squared indicated that F0 range increased slightly during the first months and decreased at later ages. Overall, more complex productions, such as reduplicated and variegated babbling, tended to be produced with a wider F0 range than vocalizations (Babb-2: β = 0.77, SE = 0.12; Babb-3: β = 0.89, SE = 0.22). Grunts showed the smallest F0 range (β = -0.45, SE = 0.07). Simple babbling also exhibited a narrower F0 range than vocalizations (β = -0.19, SE = 0.08). The difference

between simple and reduplicated babbling contrasts with Robb et al. (1989), who reported slightly higher F0 range values for monosyllabic than for disyllabic productions. The inclusion of the interaction (age x production type) in Model 3 (M3) revealed a significant interaction between age and variegated babbling (β = 0.12, SE = 0.05); however, the decrease in -2 log-likelihood was minimal and not statistically significant (-2LL = 11.75, df = 10). Therefore, Model 2 was retained as the best representation of F0 range development and is illustrated in Figure 3.3.

Table 3.6. Fixed and random effects from multilevel models predicting F0 range across age and types of pre-lexical vocal productions.

	M0		M1		M2		M3	
	b	s.e.	b	s.e.	b	s.e.	b	s.e.
Fixed Effects								
Constant	4.47	0.04	4.70	0.08	4.88	0.09	4.89	0.10
Age	--	--	0.01	0.01	0.00	0.01	0.00	0.02
Age^2	--	--	-0.06	0.02	-0.06	0.02	-0.05	0.02
Productions; base vocalization								
Grunt	--	--	--	--	-0.45	0.07	-0.47	0.12
Babb-1	--	--	--	--	-0.19	0.08	0.29	0.23
Babb-2	--	--	--	--	0.77	0.12	1.01	0.33
Babb-3	--	--	--	--	0.89	0.22	-0.94	0.83
Age*Grunt	--	--	--	--	--	--	0.00	0.01
Age*Babb-1	--	--	--	--	--	--	-0.03	0.02
Age*Babb-2	--	--	--	--			-0.02	0.02
Age*Babb-3	--	--	--	--	--	--	0.12	0.05
Random effects (level 2)								
u	12.96	0.23	12.72	0.23	12.32	0.22	12.31	0.22
Random effects (level 1)								
e	1.36	0.04	1.37	0.04	1.42	0.04	1.42	0.04
Model Fit statistics								
-2LL	--	--	-96.6	2 df	-117.73	6 df	-11.75	10 df

Note. Roughly if |2 x SE | ≤ |parameter| than p < .05; u = between variance, e = within variance

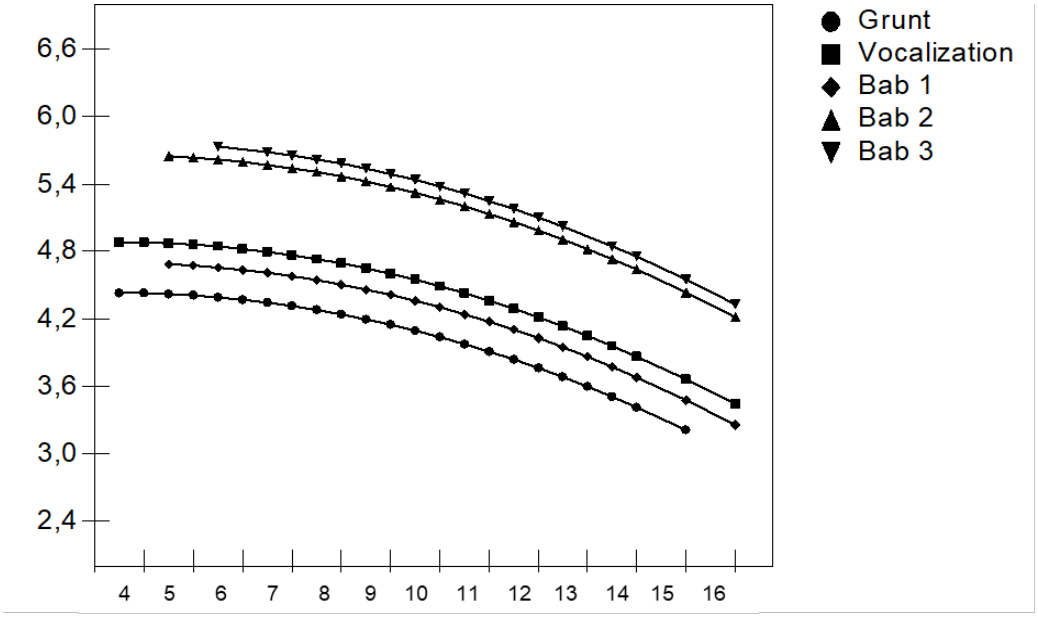


Figure 3.3. F0 range (in semitones) values of each type of productions from 4 to 16 months of age predicted by Model 2 (M2 in Table 3.6).

3.3.4. F0 final contours trajectories over time

To investigate developmental trajectories in infants' ability to assign directional patterns to F0 changes, multilevel (binomial) logistic analyses were performed, using the presence or absence of level, rising, and falling F0 final contours as dependent variables**.** A total of 9,597 productions were analyzed after excluding cases in which the classification of F0 final contours was unreliable due to excessive noise, interference, or spectrographic artifacts. Predicted values of presence of contours in Figures 3.4-3.6 are expressed on the log-odds (logit) scale, as estimated by the multilevel logistic models. Positive values indicate a higher probability of producing a given F0 final contour, whereas negative values indicate a lower probability.

3.3.4.1. Level F0 contours

Model 0 showed significant between-subject variability in the production of level contours, with an average probability of 79% of producing level F0 final contours (intercept = 1.36, SE = 0.05; u = 3.94, SE = 0.17; see Table 3.7). The addition of age and age squared as fixed effects in Model 1 did not yield significant effects (Age: $\beta = 0.00$, SE = 0.03; Age²: $\beta = -0.03$, SE =

0.03). Model 2 revealed significant effects of reduplicated and variegated babbling, which were more frequently produced with level F0 final contours (Babb-2: β = 1.23, SE = 0.21; Babb-3: β = 2.23, SE = 0.51). The probability of producing level contours did not vary over time (Age: β = -0.03, SE = 0.03; Age²: β = -0.02, SE = 0.03). The interaction between age and type of production included in Model 3 was not significant (Age*Babb-2: β = 0.02, SE = 0.04; Age*Babb-3: β = -0.07, SE = 0.11). Therefore, Model 2 was selected as the best representation of the occurrence of level F0 final contours and is illustrated in Figure 3.4.

Table 3.7. Predictors of presence of level F0 contour productions (binomial level vs no level). Logistic Binomial Multilevel results.

	M0		M1		M2		M3	
	b	s.e.	b	s.e.	b	s.e.	b	s.e.
Fixed Effects								
Constant	1.36	0.05	1.58	0.16	1.72	0.17	1.74	0.19
Age	--	--	0.00	0.03	-0.03	0.03	-0.04	0.03
Age^2	--	--	-0.03	0.03	-0.02	0.03	0.00	0.03
Productions; base vocalization								
Grunt	--	--	--	--	-0.21	0.12	-0.23	0.23
Babb-1	--	--	--	--	0.03	0.13	0.41	0.35
Babb-2	--	--	--	--	1.23	0.21	0.93	0.58
Babb-3	--	--	--	--	2.23	0.51	3.34	1.90
Age*Grunt	--	--	--	--	--	--	0.00	0.02
Age*Babb-1	--	--	--	--	--	--	-0.03	0.02
Age*Babb-2	--	--	--	--	--	--	0.02	0.04
Age*Babb-3	--	--	--	--	--	--	-0.07	0.11
Random effects (level 2)								
u	3.94	0.17	4.02	0.17	3.88	0.17	3.88	0.17
Random effects (level 1)								
e	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00

Note. Roughly if |2 x SE | ≤ |parameter| than p < .05; u = between variance, e = within variance.

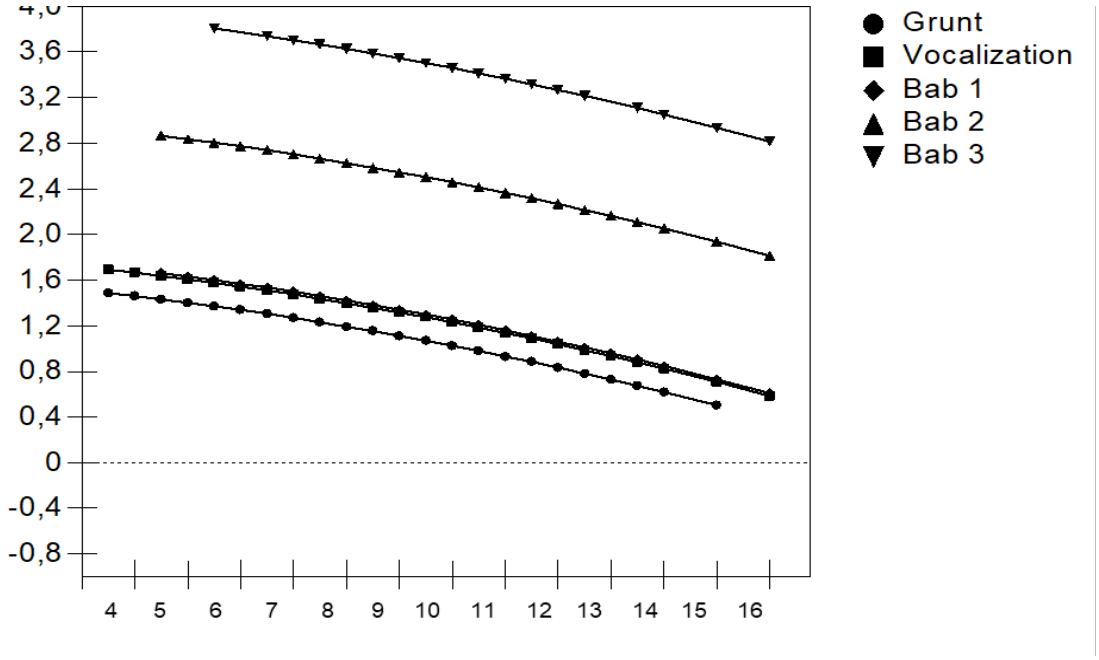


Figure 3.4. Predicted log-odds of producing level F0 final contours across age and type of production, as estimated by Model 2 (Table 3.7).

3.3.4.2. Rising F0 contours

Concerning rising F0 final contours, Model 0 revealed significant between-subject variability in the production of rising contours (intercept = 0.35, SE = 0.04; u = 2.43, SE = 0.11; see Table 3.8). The inclusion of age and age squared as fixed effects in Model 1 did not reveal significant developmental effects (Age: β = 0.01, SE = 0.02; Age²: β = -0.04, SE = 0.02). Model 2 showed a significant effect of type of production: rising final contours were less frequently produced in grunts (β = -0.28, SE = 0.09) and more frequently produced in more complex productions, such as reduplicated and variegated babbling (Babb-2: β = 0.77, SE = 0.15; Babb-3: β = 1.17, SE = 0.27). Overall, the probability of producing rising F0 final contours did not change over time (Age: β = -0.01, SE = 0.02; Age²: β = -0.04, SE = 0.02).

The inclusion of interaction terms between age and type of production in M3 did not produce significant effects (Age*Grunt: β = 0.00, SE = 0.02; Age*Babb-1: β = -0.03, SE = 0.02; Age*Babb-2: β = -0.02, SE = 0.03; Age*Babb-3: β = 0.05, SE = 0.06). Therefore, M2 was selected as the best representation of the presence of rising F0 final contours and is

illustrated in Figure 3.5.

Table 3.8. Predictors of the presence of rising F0 final contours in the production (binomial rising vs no rising). Logistic Binomial Multilevel results.

	M0		M1		M2		M3	
	b	s.e.	b	s.e.	b	s.e.	b	s.e.
Fixed Effects								
Constant	0.35	0.04	0.53	0.13	0.68	0.13	0.68	0.15
Age	--	--	0.01	0.02	-0.01	0.02	-0.02	0.03
Age^2	--	--	-0.04	0.02	-0.04	0.02	-0.02	0.03
Productions; base vocalization								
Grunt	--	--	--	--	-0.28	0.09	-0.28	0.19
Babb-1	--	--	--	--	-0.06	0.11	0.32	0.28
Babb-2	--	--	--	--	0.77	0.15	1.01	0.41
Babb-3	--	--	--	--	1.17	0.27	0.34	0.94
Age*Grunt	--	--	--	--	--	--	0.00	0.02
Age*Babb-1	--	--	--	--	--	--	-0.03	0.02
Age*Babb-2	--	--	--	--	--	--	-0.02	0.03
Age*Babb-3	--	--	--	--	--	--	0.05	0.06
Random effects (level 2)								
u	2.43	0.11	2.47	0.11	2.47	0.11	2.47	0.11
Random effects (level 1)								
e	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00

Note. Roughly if |2 x SE | ≤ |parameter| than p < .05; u = between variance, e = within variance.

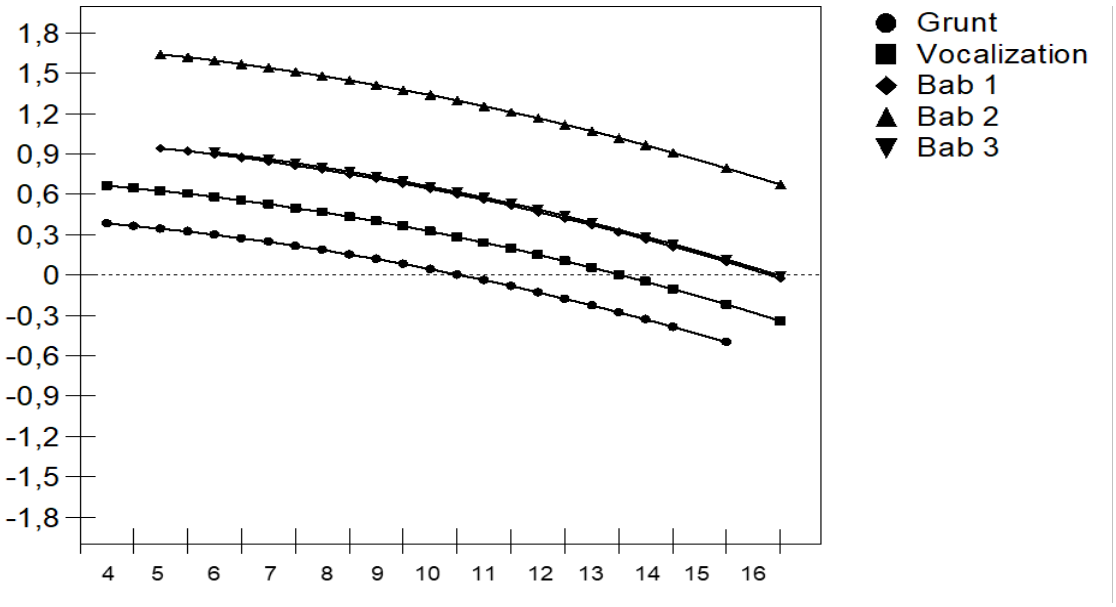


Figure 3.5. Predicted log-odds of producing rising F0 final contours across age and type of production, as estimated by Model 2 (Table 3.8).

3.3.4.3. Falling F0 contours

The same analyses were conducted for the presence or absence of falling F0 final contours. The basic model revealed significant between-subject variability and an overall probability of 62% of producing falling final contours within vocal productions (intercept = 0.51, SE = 0.04; u = 2.53, SE = 0.11; see Table 3.9). M1 showed a significant effect of age squared, indicating that the presence of falling final contours remained relatively stable across the age range considered (Age^2: β = -0.06, SE = 0.02). In M2, the effects of grunts, reduplicated babbling, and variegated babbling were significant (Grunt: β = -0.13, SE = 0.10; Babb-2: β = 1.04, SE = 0.16; Babb-3: β = 1.80, SE = 0.32), with reduplicated and variegated babbling being produced with a higher probability of falling F0 final contours (see Figure 3.6).

When adding interactions between age and type of production in Model 3, no significant effects on the presence of falling F0 final contours were observed (Age*Grunt: β = 0.01, SE = 0.02; Age*Babb-1: β = 0.00, SE = 0.02; Age*Babb-2: β = 0.01, SE = 0.03; Age*Babb-3: β = -0.02, SE = 0.07).

Table 3.9. Predictors of the presence of falling F0 final contours in the production (binomial falling vs no falling). Logistic Binomial Multilevel results.

	M0		M1		M2		M3	
	b	s.e.	b	s.e.	b	s.e.	b	s.e.
Fixed Effects								
Constant	0.51	0.04	0.67	0.13	0.78	0.13	0.80	0.15
Age	--	--	0.02	0.02	-0.01	0.02	-0.01	0.03
Age^2	--	--	-0.06	0.02	-0.04	0.02	-0.04	0.03
Productions; base vocalization								
Grunt	--	--	--	--	-0.13	0.10	-0.18	0.19
Babb-1	--	--	--	--	0.11	0.11	0.09	0.28
Babb-2	--	--	--	--	1.04	0.16	0.94	0.43
Babb-3	--	--	--	--	1.8	0.32	2.19	1.13
Age*Grunt	--	--	--	--	--	--	0.01	0.02
Age*Babb-1	--	--	--	--	--	--	0.00	0.02
Age*Babb-2	--	--	--	--	--	--	0.01	0.03
Age*Babb-3	--	--	--	--	--	--	-0.02	0.07
Random effects (level 2)								
u	2.53	0.11	2.58	0.11	2.57	0.12	2.57	0.12
Random effects (level 1)								
e	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00

Note. Roughly if $|2 \times SE| \leq |parameter|$ than $p < .05$; u = between variance, e = within variance.

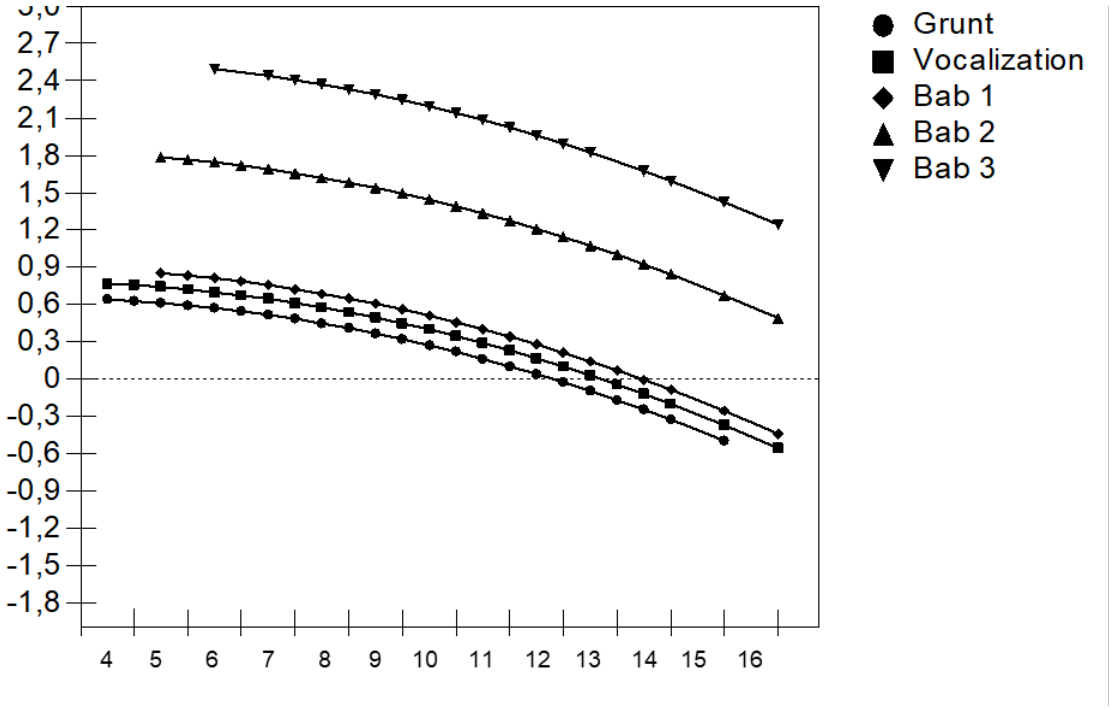


Figure 3.6. Predicted log-odds of producing falling F0 final contours across age and type of production, as estimated by Model 2 (Table 3.9).

3.4. Discussions

To contribute to the mixed findings in the existing literature and to address several limitations of previous studies, the present study aimed to describe the development of F0-related prosodic features in the pre-lexical productions of Italian infants between 4 and 16 months of age. Consistent with previous research, the present results confirm the presence of substantial inter-individual variability in infants' prosodic development. Such variability has been widely documented in early vocal development (e.g., Laufer & Horii, 1977; Kent & Murray, 1982; Vihman, 1985), and the current findings extend this evidence to Italian infants' pre-lexical vocal productions. From the first months of life, infants show differences in voice use in all the prosodic variables examined and these differences can be partly attributed to physiological variability in the shape and maturation of the larynx.

This degree of variability may help explain the heterogeneous values reported in previous studies, many of which were conducted on small samples. Moreover, the frequent lack of

control for inter-individual variability may account for the inconsistent or non-significant results reported in the literature. In this respect, the degree of variability observed in the present sample appears larger than that reported by Laufer and Horii (1977), who examined four infants, and contrasts with the findings of Flax et al. (1991), who did not observe significant inter-individual differences in a sample of three children. We may suppose that a larger sample would provide a clearer picture of individual differences among infants, especially in how prosodic features change over age and across different types of production. In the present study, infants did not all show the same patterns of change over time, suggesting that developmental trajectories may differ across individuals. Further longitudinal investigations at older ages would help clarify whether these inter-individual differences are mainly related to early physiological organization of the vocal tract or whether they persist, or even increase, over development, eventually contributing to the well-known differences in adult voice characteristics.

By controlling for this variability, we were able to describe the developmental trajectories of all the examined variables with greater accuracy. F0 mean showed a slight increase over time, of approximately 3.2 Hz per month. This trend contrasts with several previous studies but is consistent with the slight increase reported in studies based on smaller samples, such as Laufer and Horii (1977), who observed a similar, though not statistically significant, pattern. Other authors have also reported increases in F0 mean values, but only from later stages of development, such as from the fifth month (Fairbanks, 1942) or the tenth month (Murry et al., 1983).

In the first two years of life, infants improve their interactive abilities, as shown by longer gaze durations and an increased use of gestures directed toward the mother (Papaeliou & Trevarthen, 2006; Trevarthen, 1977). Accordingly, we may hypothesize that the voice is increasingly used to establish contact with the caregiver and to respond to the stimulation

received. It is possible that this greater participation in interaction corresponds to a higher state of activation and, therefore, to an increase in F0 mean values. Moreover, previous studies have shown that communicative, explorative or emotional pre-lexical productions display different prosodic features. Specifically, a reduced F0 mean has been observed in non-communicative productions, such as those produced while exploring an object. In contrast, productions directed to the adult during face-to-face interactions, as well as productions associated with imperative gestures, are produced with higher F0 mean values (Aureli et al., 2017; Papaeliou et al., 2002; Papaeliou & Trevarthen, 2006). Since in the present study mother-infant interactions were recorded during face-to-face interactions without objects, we may suppose that the infants' vocal productions were mostly produced with the aim of communicating with the mother and, therefore, with a higher F0 mean.

A different pattern emerged for the developmental trajectories of F0 range. Infants tended to slightly increase the amplitude of F0 variation within their productions during the first months of life, whereas after the seventh month the F0 range of all production types decreased by approximately one semitone every two months. This pattern may indicate that very young infants' productions are less controlled and characterized by large pitch excursions, while during the second half of the first year of life infants progressively gain better control over their voice and begin to use prosodic variations that are closer to those observed in adult speech.

Previous studies have shown that infants tend to imitate the pitch variations of caregivers' vocal productions (Gratier & Devouche, 2011; Ko et al., 2016), and it is plausible that this imitative ability becomes more efficient after the seventh month of life. The wider F0 range observed during the first months is consistent with the findings of Amano et al. (2006) on the prosodic development of Japanese infants; however, those authors reported a decrease in F0 range only after the onset of two-word combinations. By contrast, Snow (2004) reported an increase in F0 range from the first to the fourth year of life. These discrepancies are likely

related to the different age ranges and methodological approaches adopted across studies, which make direct comparisons difficult. In this respect, the present analyses, focusing on a narrower age range and including a higher number of longitudinal observations, may have been better suited to capture subtle developmental variations that were not reported in previous studies.

It should be noted however that F0 range is a global measure that reflects both how much speakers vary their pitch and which types of intonational patterns they use. Because F0 range was analyzed independently of intonational contour type, changes in this measure may partly reflect differences in the pitch patterns produced across time, rather than changes in vocal control alone. Previous work has shown that children gradually increase the height of pitch movements within specific intonational patterns during development (Frota et al., 2016). As a result, developmental changes in F0 range may reflect increasing control over pitch variation as well as likely shifts in the types of intonational patterns children produce at different ages.

Concerning F0 final contours, most infant vocal productions were classified as flat. This result differs from previous studies and may be linked to the interactional context considered. During face-to-face interactions without objects, infants are less likely to need to produce vocalizations with a wider F0 range in order, for example, to attract the mother's attention or to request an object (see Aureli et al., 2017; D'Odorico & Franco, 1991; Esteve-Gibert & Prieto, 2013). A more detailed coding of the communicative function of the productions would have provided further information in this respect.

Similarly, the percentages of rising and falling final contours did not vary across age. Our findings are not consistent with those reported by Snow (2006), nor with the hypothesis proposed by Lieberman (1967). The significant effect of type of production at all ages may contrast the idea that, during the first months of life, prosody is predominantly physiologically controlled. Rather, our results suggest that from the beginning of pre-lexical speech infants are

able to modulate prosodic patterns depending on the type of vocal production, as also supported by previous studies showing that infants can differentiate intonational patterns to signal communicative intent (Prieto et al., 2012). Both the probability of producing rising and falling final contours slightly decreased over time in our data, although we did not observe the U-shaped pattern reported by Snow (2006). If this difference is due to the age range explored in the present study, it is possible that the probability of producing directional final contours increases again after the 16th month of life. However, a further explanation for this finding may lie in the exclusive focus on pre-lexical vocal productions. Previous studies reporting a higher proportion of directional contours analyzed meaningful or communicative utterances, following the criteria proposed by Snow (2006). For example, Prieto et al. (2012) and Frota et al. (2016) found that falling nuclear contours predominate early on, with an increasing diversity of contour types emerging over time, a pattern already visible between 12 and 16 months and more clearly after 16 months of age. Differences in the type of vocal material analyzed may therefore partly account for the discrepancies between the present findings and previous literature.

A consistent result emerging from the present study is the effect of the type of vocal production on all the prosodic variables examined. The different production types are characterized by distinct prosodic patterns, and these differences remain stable across the age range considered.

Grunts were characterized by the lowest F0 mean and F0 range values compared to the other production types. Moreover, it is in grunts and vocalizations that infants appeared to make less use of rising and falling final contours. These production types were among the most frequent across the age range considered but were associated with less marked prosodic modulation. Higher F0 mean values, wider F0 ranges, and the presence of rising and falling contours may index higher levels of activation and prosodic elaboration. In the present data,

these prosodic characteristics were more frequently observed in production types typically associated with more specific communicative functions, such as simple and variegated babbling (Fasolo et al., 2008).

Some limitations of the present study should be acknowledged. First, F0 range was measured in semitones, whereas many previous studies have relied on the standard deviation (SD) of F0 mean, which limits direct comparability across studies. Moreover, we did not take into account the different communicative intentions underlying infants' vocal productions (e.g., whether they were produced in response to maternal speech, to attract maternal attention, or to express a need), which represents an important dimension for the interpretation of prosodic variation. In addition, the present analyses relied on general acoustic measures of pitch variation and final contour direction, without addressing the internal structural properties of the melodic patterns produced, such as the type of nuclear contour (e.g., pitch accents and boundary tones). Future studies combining communicative-function coding with contour-based and phonologically informed analyses would provide a more specific account of early prosodic development.

The present study highlights the importance of considering individual variability when assessing infant and child prosody. The findings suggest that, during the second half of the first year of life, infants show differentiated patterns in the use of pitch-related prosodic features that may contribute to the regulation of interaction with their caregiver, a process considered one of the major prerequisites for language acquisition (Papaeliou et al., 2002). The use of multilevel models proved particularly useful in this respect. Applying statistical approaches that can explore and control individual variability becomes essential when studying developmental processes in childhood. Further longitudinal studies may investigate whether this variability among children remains stable over time or changes with development, helping to identify the developmental stages at which controlling for individual variability is most

crucial. Another implication concerns the importance of exploring linguistic and prosodic development already at the pre-linguistic stage. The present findings indicate that prosodic features should be examined in relation to the type of vocal production analyzed. Combining this approach with a systematic coding of communicative intentions and acknowledging nuclear contours would provide a more comprehensive account of early prosodic development.

3.5. Conclusions

In conclusion, in this Italian sample pitch-related prosodic features in pre-lexical vocal productions vary systematically depending on types of production and individual developmental trajectories, already from the earliest stages of vocal development. By focusing on pre-lexical productions and accounting for individual variability, these findings provide new insights into early prosodic organization, highlighting the importance of considering production type and developmental context when studying the emergence of prosody in infancy.

3.6. Acknowledgments

The authors express their deepest gratitude to the late Professor Hans Schadee for his invaluable contributions to this research. His expertise and dedication in performing the analyses were instrumental to the success of this study. Professor Schadee will be greatly missed, and we are profoundly grateful for his support and collaboration.

CHAPTER 4| The Prosody of Contrastive Focus in Children with High-Functioning Autism

D'Aloia V., Filipe M. G., Fasolo M. and Frota S. (*in preparation*).
The Prosody of Contrastive Focus in Children with High-Functioning Autism. Clinical Linguistics & Phonetics.

Abstract

The expression of focus is crucial for effective communication. Previous work on focus marking in children with Autism Spectrum Disorder (ASD) has mostly adopted a phonetic approach and little is known about the phonology of focus prosody in ASD. This study addresses this gap by investigating the location of prominence and strategies to signal it in children with High-Functioning Autism (HFA). Speech data from 18 Portuguese-speaking children with HFA (5–15 years of age, M = 8.06, SD = 2.31, IQ ≥ 80) and typically developing (TD) peers, matched for age, gender, and non-verbal IQ, from the Contrastive Focus (CF) Production Task of the PEPS-C were analysed. Productions were annotated in Praat following the Portuguese ToBi framework and analyses assessed accuracy of prominence placement, error patterns and prosodic strategies (pitch accents and boundary tones). Participants' scores on the Peabody Picture Vocabulary Test (PPVT) were included as a covariate in analyses of focus production accuracy. Children with HFA showed lower overall accuracy in CF production than TD peers; however, this difference disappeared after controlling for PPVT. Moreover, groups differed qualitatively in their error patterns: HFA showed more distributed prominence across multiple constituents, whereas TD children produced simpler non-target realizations. In successful CF realizations, both groups predominantly used the expected H*+L pitch accent. When prominence was misplaced, TD children tended to preserve H*+L, while children with HFA showed greater variability in pitch accent choice. Across groups, prosodic patterns were more stable when prominence occurred closer to the right edge of the utterance, indicating difficulty with sentence-initial prominence, particularly in HFA. Finally, children

with HFA were more likely than TD peers to produce non-target boundary tones when prominence was misplaced. Overall, these findings suggest that the phonology of CF prosody in children with HFA is impaired in prominence location and intonation categories.

Keywords: *Autism Spectrum Disorder, Prosody, Contrastive Focus, Autosegmental Metric Theory of Intonation, PEPS-C*

4.1. Introduction

Autism Spectrum Disorder (ASD) is a common neurodevelopmental disorder marked by pervasive difficulties in social communication and interaction together with restricted and repetitive behavior or interest (American Psychiatric Association, 2013). Language abilities in ASD may vary, ranging from non-verbal or minimally verbal individuals to speakers with fluent speech (Kjelgaard & Tager-Flusberg, 2001; Tager-Flusberg et al., 2005). Despite this heterogeneity, prosody abnormalities have been recognized as central features of speech in ASD and were already documented in the first clinical description of autism (Kanner, 1943). Prosody has therefore gained a lot of attention in the literature as a salient aspect of language in ASD, also because while many other aspects of language improve over time, prosodic difficulties remain (Paul et al., 2005) . Given its persistent nature, speech prosody is also assessed in gold standard diagnostic tools for ASD, such as the Autism Diagnostic Observation Schedule (ADOS; Lord et al., 2000).

4.1.1. Prosody in Autism Spectrum Disorder

Often summarized as the rhythm or melody of speech, prosody refers to suprasegmental features of spoken language which may include pitch, duration and intensity (Lehiste, 1970; Ladd, 2008). Through the modulation of these acoustic cues, prosody plays a crucial role in social communication since it enhances or changes the meaning of what it's being said. A substantial body of research has documented atypical prosodic patterns in individuals with

ASD. Across studies, receptive prosody has been shown to be less affected than expressive prosody (McCann et al., 2007; Paul et al., 2005), whereas production is often characterized by impairments of stress placement (Peppé et al., 2006) or irregularities such as longer pauses before turn-taking (Ochi et al., 2019).

Prosody in ASD has been described as "bizarre", "singsongy" or overly melodic but also as "monotone", "robotic" and "stilted" (Fay & Schuler; 1980; Frith, 1991; Baron-Cohen & Staunton; 1994). Taken together, these descriptions suggest that prosodic abnormalities manifest differently across individuals with ASD (Green & Tobin, 2009). From an acoustic perspective, several studies reported increased mean pitch, wider pitch range and greater pitch variability compared to typically developing (TD) peers (Diehl et al. 2009; Filipe et al., 2014; Hubbard et al 2017; Drimalla et al., 2020). Individuals with ASD showed longer overall utterance duration (Diehl & Paul, 2013) but also less durational contrast between stressed and unstressed syllables compared to TD speakers (Paul et al., 2008). Findings regarding speech rate are mixed: while some studies documented a slower rate of articulation in ASD (Patel et al., 2020), others failed to observe reliable differences between ASD and TD speakers (McAlpine et al., 2014). Evidence regarding voice intensity is less conclusive, with meta-analyses often indicating no significant group differences (see Asghari et al., 2021; Grice et al., 2023, for comprehensive overviews of the acoustic characteristics of prosody in ASD). However, results from the existing literature remain mixed and this variability was suggested to be driven by factors like age, language level, cognitive abilities and methodological factors (Fusaroli et al., 2017; Bonanno M. et al., 2025).

Besides its acoustic correlates prosody can also be categorized by its communicative functions which span grammatical, pragmatic, and affective domains, ranging from the marking of speech acts (e.g., questions vs. statements), through the expression of speaker intentions and the regulation of turn-taking, to the conveyance of one's emotional states

(Halliday, 1977; Nespor & Vogel, 1986; Gussenhoven, 2004; Pfanner et al., 2008). It has been noted that in speakers with ASD grammatical uses of prosody (e.g. phrasal boundaries; Fine et al., 1991), were less impacted than affective and pragmatic uses (e.g. to signal contrast between referents or to correct information introduced earlier in the discourse; Paul et al., 2005; Peppé et al., 2007).

In the present study we focus on the pragmatic use of prosody, more specifically the production of contrastive focus, a highly context-dependent phenomenon, that relies on perspective-taking and therefore remains particularly challenging for individuals with ASD (Brosnan et al., 2017: Landa, 2000; Grice et al., 2023).

4.1.2. Production of Contrastive stress in speakers with Autism Spectrum Disorder.

Contrastive stress refers to the acoustic emphasis that is employed when distinguishing explicit alternatives or correcting an item already presented in discourse (Rooth, 1992; Samek-Lodovici; 2018). Bates and MacWhinney (1979) noted more specifically that contrastive stress is the device speakers use most often to replace some aspect of the listener's beliefs. An illustrative example is provided by Baltaxe and colleagues (1984):

Example (1)

Experimenter: Is Mike sitting in the chair?

(Scene enacted with dolls: Pat seated in a chair.)

Contrastive response: No, **PAT** is sitting in the chair.

In this case the word "*PAT"* carries the strongest prosodic prominence, which is acoustically realized through higher fundamental frequency (F0), increased duration, and greater intensity relative to the surrounding syllables (Watson, 2010). As a result, the listener is guided to

interpret *PAT* as the intended referent, distinguishing it from alternative potential referents (e.g., *someone else*). In such contexts, the placement of prosodic prominence on "*PAT*" serves to mark *contrastive focus,* a type of narrow focus (Zimmermann, 2007) which highlights a single constituent by contrasting it with a set of contextually relevant alternatives, rather than spreading prominence across all the utterance. Producing contrastive intonation means relying on the discourse context at hand, which is crucially determined by what is said before by both interlocutors. This process means considering the other's person perspective, an ability that has been reported to be impaired in ASD (Baron-Cohen, 1995). Compared to the extensive literature on autism and prosody, relatively few studies have examined the marking of contrastive focus (CF) or information structure in individuals with ASD. General findings suggest that individuals with ASD experience difficulties both in perception and production of CF (Grice et al., 2023).

Early studies on the production of CF in ASD relied primarily on perceptual analysis of stress placement and reported reduced use of contrastive stress and misplacement of prominence. Using an elicitation task in which contrastive focus could fall on the subject, verb, or object, Baltaxe (1984) reported that English ASD children (aged 4-12 years) produced more errors when CF was realized on the verb, and fewer errors when it fell on the subject. Among verb-related errors, the most frequent pattern involved simultaneous stress assignment, which accounted for 36.84% of all verb errors. More recent work on children has made extensive use of standardized elicitation tasks, most notably the *Profiling Elements of Prosodic Systems in Children* (PEPS-C; Peppé & McCann, 2003), which is the most used tool to assess prosody in developmental populations. In another study on English-speaking children using the PEPS-C focus expression task, Peppé et al. (2007) reported lower accuracy in CF production in children with ASD (aged 6-13 years) compared to TD peers, with errors often involving incorrect stress on the adjective when it should have been on the noun. However, other studies using the same

task have found no group differences in stress realization at the expected location in CF (Järvinen-Pasley et al., 2008; Diehl & Paul, 2013). Beyond perceptual measures, a limited number of studies have examined the acoustic correlates of contrastive focus in ASD, with mixed findings. Diehl and Paul (2013) reported greater pitch range and variability, and overall duration, but measures were taken over the entire utterance produced. Using an interactive task and samples matched for language abilities, Nadig and Shaw (2015) found that children with ASD marked contrastive focus in the expected location, with both groups relying on duration and intensity cues. Other work integrating receptive vocabulary measures has shown that pitch range patterns in ASD vary as a function of language abilities, with expanded pitch range in individuals with high language skills and reduced pitch range in those with more moderate skills, although these effects were not specific to contrastive focus per se, but to broader distinctions in information structure (DePape et al., 2012).

Overall, studies on contrastive focus in ASD are limited and primarily restricted to cases in which CF is successfully realized leaving unresolved how prosodic cues are distributed when prominence is misplaced. Moreover, most of the current literature relies on subjective ratings rather than objective acoustic or phonological measures. Phonological accounts of the intonational marking of contrastive focus in ASD are almost entirely absent. To date, only Krüger et al. (2018) explicitly examined intonational categories, showing that German speakers with ASD predominantly used H* accents to mark new referents, whereas typically developing speakers favored rising pitch accent types. Taken together these limitations point to a lack of clarity regarding the location and the phonological nature of prominence when contrastive focus is atypically produced in ASD.

4.1.3. The present study

Considering the mixed findings and literature limitations, the present study aims to investigate how contrastive focus prosody is impaired in Portuguese children with High-Functioning

Autism (HFA). Specifically, we analyzed both the placement of prominence (on the intended target word or elsewhere) and the prosodic strategies used to convey it within the framework of Autosegmental-Metrical Theory of Intonation (Pierrehumbert, 1980; Beckman and Pierrehumbert 1986; Ladd, 2008; Frota, 2014), using the focus expression task of the Profiling Elements of Prosodic Systems in Children (PEPS-C), while also controlling for language abilities. The research questions (RQ) we sought to answer were:

(1) Do children with HFA and TD peers differ in the accuracy of prominence placement on the intended target word when producing CF?

(2) When CF is not successfully realized, do HFA and TD children differ in how prosodic prominence is misplaced across non-target elements of the utterance (i.e. error patterns)?

(3) Do children with HFA and TD peers differ in the prosodic strategies used to mark contrastive focus both in cases of successful and unsuccessful prominence placement?

(4) After controlling for language abilities, do children with HFA and TD peers still differ in the accuracy of prominence placement when producing CF?

We hypothesized that children with High-Functioning Autism (HFA) would differ from typically developing (TD) peers showing lower accuracy of contrastive focus production. Analyses of prominence misplacement patterns and prosodic strategies were exploratory, given the limited available literature. Finally, when examining whether group differences in accuracy persisted after controlling for language ability, no directional hypothesis was formulated.

4.2. Methods

4.2.1. Participants

Eighteen children (6 girls,12 boys) with High Functioning Autism (HFA) (5-15 years; M = 8.06, SD = 2.31) who met the DSM-5 criteria for Autism (American Association Psychiatric, 2013) participated in this study. ASD diagnosis was made by a team of child psychiatrists and

psychologists, based on the Autism Diagnostic Observation Schedule - Second Edition (ADOS-2; Lord et al., 2012) and the Autism Diagnostic Interview-Revised (ADI-R; Lord et al., 1994). All HFA participants were required to have an IQ of 80 or higher, as assessed by the Wechsler Intelligence Scale for Children–III (Wechsler, 1991), to ensure that performance on the prosodic task was not confounded by general cognitive impairment. Exclusion criteria included obsessive-compulsive disorder, attention-deficit/hyperactivity disorder, and leaning disorder as defined by DSM-5. HFA children were matched to a group of Typically Developing (TD) peers by age (M= 8.06, SD = 2.31), gender and non-verbal intelligence (HFA: M = 25, SD = 5.81; TD: M = 27.1, SD = 5.55) assessed with Raven's Colored Progressive Matrices, (RCPM; Raven, 1995; Portuguese version: Simões, 2000). All participants were native speakers of European Portuguese (EP) born and raised in monolingual homes in the North of Portugal with no visual or hearing problems.

4.2.2. Materials

For the purposes of this study, we used recordings collected from the application of the European Portuguese version of the Profiling Elements of Prosody in Speech-Communication (PEPS-C; original version: Peppé & McCann, 2003; Portuguese adaptation: Filipe et al., 2017) obtained using a Marantz PMD661 digital recorder and a wireless headset microphone system (Sennheiser EW 152 G3-A, model 503102). The PEPS-C is a standardized assessment battery designed to evaluate prosodic abilities across multiple subtests that examine both formal and functional aspects of prosody, including perception and production tasks in communicative domains such as turn-taking, affective prosody, chunking, and focus marking. The full PEPS-C dataset, comprising all subtests, was originally collected as part of a PhD project funded by the Fundação para a Ciência e Tecnologia (FCT; SFRH/BD/64166/2009). The present analyses focused exclusively on the *Focus Expression* subtest, which assesses the ability to produce contrastive focus. This subtest was subsequently expanded with additional participants and is

currently publicly available on the Open Science Framework (OSF; https://doi.org/10.17605/OSF.IO/QR74X).

In this task, children viewed a picture of animals playing soccer (e.g., a red cow holding a ball), heard an incorrect description of it (1a), and were instructed to correct the statement (1b), thereby eliciting contrastive narrow focus productions. This elicitation procedure was designed to prompt a contrastive focus on a specific target word, with prosodic prominence typically realized either on the subject noun (e.g., *vaca* "cow*"*) or on the adjective (e.g., *vermelha* "red*"*), thereby restricting the expected accentual placement to these two lexical categories.

(1)

a. *A vaca preta tem a bola*.

'The black cow has the ball.'

b. *A vaca VERMELHA tem a bola.*

'The RED cow has the ball.'

The task consisted of a total of 20 items per child, including two example items, two training items and 16 experimental items to elicit contrastive focus. At the time of the original data collection, responses were audio-recorded in a quiet room at the children's school during a single individual session of approximately 45 minutes. The children were recorded using a headset microphone.

To control for language abilities participants were also evaluated on receptive vocabulary, which was assessed with Peabody Picture Vocabulary Test (PPVT; Dunn & Dunn, 2007). PPVT data were originally collected as part of the same dataset and were later expanded through inclusion of additional participants. In this task, participants were required to select, by pointing, one of four pictures that best represented the meaning of a verbally presented stimulus.

4.2.3. Procedure

4.2.3.1. Annotation of Focus Placement

All recordings were originally collected with informed consent, and approval for their re-analysis was granted by the Research Ethics Committee (CEI; reference 27_CEI2025). All utterances were annotated in Praat (Boersma & Weenink, 2009) to identify the location of prosodic prominence. Each production was coded according to where the main pitch accent fell. In cases where prominence was realized on the expected target word (i.e., a *match* response), only two positions were possible: prominence on the subject noun (N1: the animal) or on the adjective (Adj: the colour of the animal), as by design of this task the CF accent could only fall on one of these two constituents. When prominence occurred elsewhere or was distributed across multiple constituents (i.e., *mismatch* responses), a broader set of coding configurations was applied to capture the specific pattern of prominence. The possible mismatch types were as follows:

a) *Wrong single word*

   Prominence was realized on a single-non target element. For example, when the target focus word was the noun (N1) but the child emphasized the adjective (Adj), or vice versa. Labels used: *N1, Adj.*

b) *Two words with equal prominence*

   Prominence was realized on two words simultaneously, with both elements perceived as equally prominent based on auditory perception and visual inspection of the pitch contour. In these cases, prominence could also extend to the object of the sentence (the second noun; N2), such as *bola* "ball*"* in *a vaca vermelha tem a bola* "the red cow has a ball".

   Labels used:

   - *2(N1, Adj)*: equal prominence on subject noun and adjective;

- *2(Adj, N2)*: equal prominence on adjective and object noun;

- *2(N1, N2)*: equal prominence on subject noun and object noun

c) *Two words with unequal prominence*

Prominence was realized on two words simultaneously, but one was clearly more prominent than the other. This difference was both perceptible to the ear and observable in the spectrogram.

Labels used:

- *2(N1>Adj)*: two prominent words but the subject noun is prosodically stronger than the adjective.

- *2(N1<Adj)*: two prominent words but the adjective is prosodically stronger than the subject noun.

d) *Three-word prominence*

Prominence was realized on three words simultaneously, namely the subject noun, the adjective, and the object noun (*N1, Adj,* and *N2*), all perceived as carrying a comparable degree of prominence. Label used: *3*.

e) *Broad Focus*

In this case prominence was realized on the sentence-final accented syllable, typically on the object noun (*N2*). This configuration reflected a broad focus pattern rather than the intended contrastive narrow focus and was realized with a falling pitch accent (H+L*), typical of European Portuguese broad focus statement. (Frota, 2000).

Label used: *BF.*

f) *No prominence*

Prominence was not realized on any specific element. Utterances in this category were characterized by a flat prosodic contour or an excessively fast

speech rate, resulting in the absence of perceptible prominence. Label used: *0*.

g) *Question-like intonation*

Prominence was not realized on any lexical element, but the utterance ended with a high boundary tone (H%), giving it a question-like contour instead of a contrastive statement. Label used: *0(q)*.

h) Unrelated or unintelligible utterances

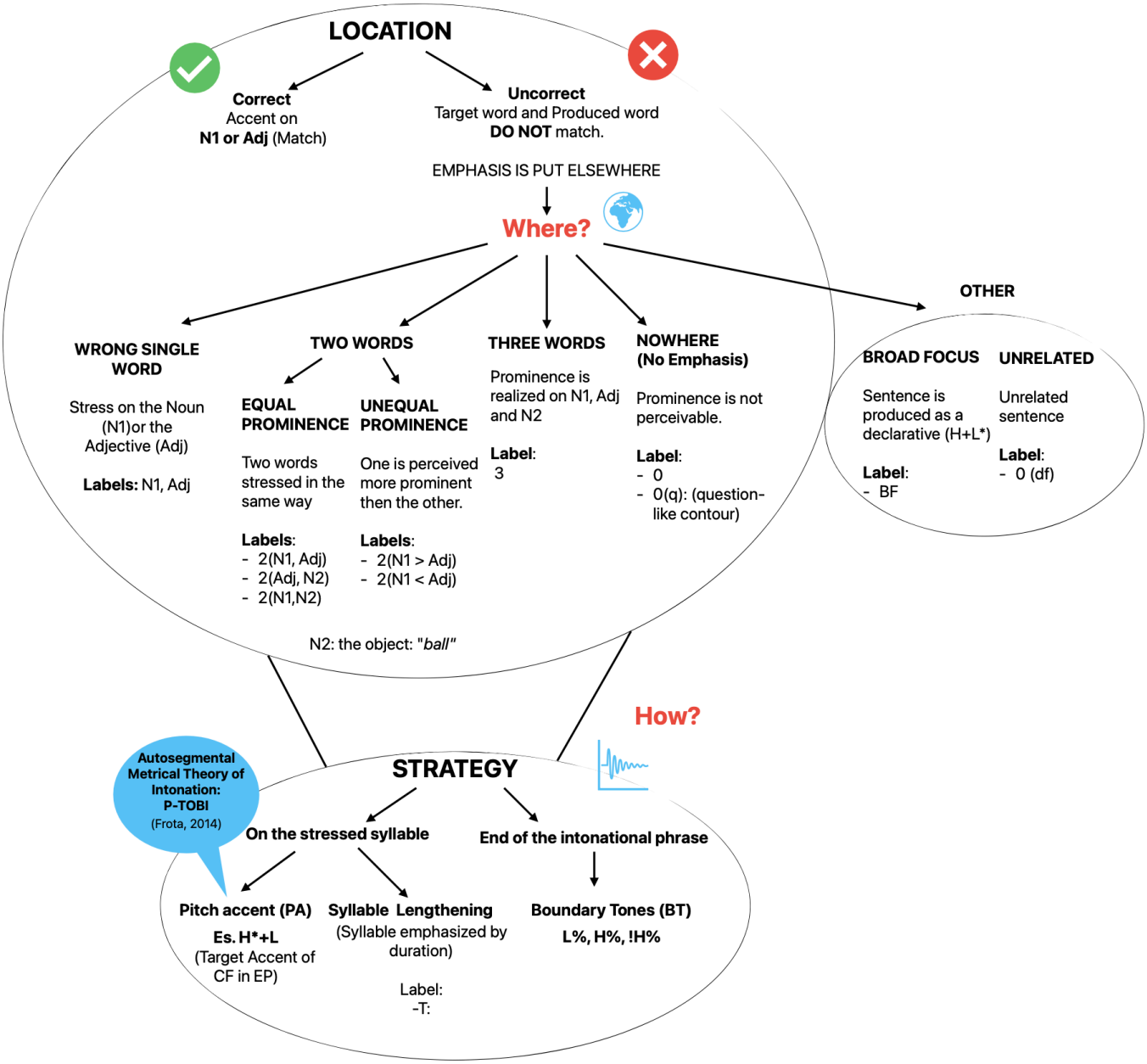
Prominence was not realized on the expected constituents, as the child produced an unrelated or unintelligible utterance that did not correspond to the target sentence.

4.2.3.2. Annotation of Prosodic Strategies for Focus Realization

Beyond identifying where prominence occurred, each utterance was simultaneously annotated according to the Portuguese Tones and Break Indices (P-ToBI) framework (Frota et al., 2015) following the Autosegmental-Metrical (AM) phonological model of intonation (Pierrehumbert, 1980; Beckman and Pierrehumbert 1986; Ladd, 2008; Frota, 2014). The tone labels included an inventory of six pitch accents (PA) (H*, L*, H*+L, H+L*, L*+H, L+H*) marking prominence on the stressed syllables and three boundary tones (BT) (!H%, H%, L%) marking the right edge of the intonational phrase. In addition to PA and BT, some instances of peominence were primarily conveyed through temporal lengthening rather than a specific intonational movement. These lengthened syllables were both perceptually and visually evident in the spectrogram, and they were annotated with the label “T:”.

Annotations were organized across four Praat tiers. The first tier encoded prosodic information, including pitch accents, lengthened syllables and boundary tones. The second tier contained the orthographic transcription of the sentence. The third tier indicated the intended target of CF for each trial (N1 or Adj). The fourth tier captured the element(s) that were actually prosodically prominent in the child’s production, allowing a direct comparison between

expected and realized focus placement. A schematic overview of the prominence coding scheme, capturing both focus configurations and prosodic strategies, is provided in Figure 4.1. Figure 4.2 exemplifies the implementation of this scheme in Praat.



4.1. Schematic overview of the prominence coding scheme used in the Contrastive Focus Production task, showing the possible focus configurations and associated prosodic strategies, along with the annotation labels employed in the analysis

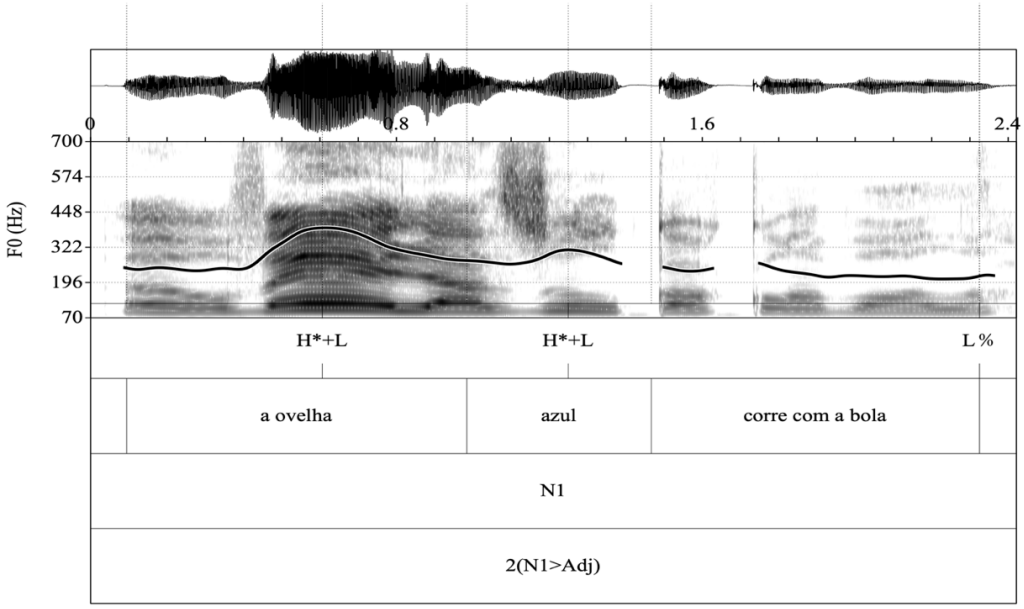


Figure 4.2. Praat annotation of a contrastive focus utterance with target focus on the noun (N1) and realized prominence on both the noun and the adjective 2(N1 > Adj). PA and BT are annotated.

4.2.3.3. Computation of the Focus Accuracy Index (FAI)

After prominence annotation, the coded data were extracted using a Praat script, and each utterance produced received a score reflecting the accuracy of focus realization. The Focus Accuracy Index (FAI) was developed *ad hoc* to quantify each participant's overall performance in the contrastive focus task. A score of 1 was assigned to sentences when prominence was realized on the correct target word (either N1 or Adj, depending on the item of the task). A score of 0.5 was assigned to sentences in which two words were emphasized but the stronger prominence matched the expected target. For example, when the target word was Adj and the production was annotated as *2(N1<Adj)*, or when the target was N1 and the production was coded as *2(N1>Adj)*. A score of 0 was assigned in all other cases, including broad focus, monotone or question-like contours, unrelated utterances, or prominence placed on a single non-target word. Each participant's FAI was then calculated as the mean accuracy across all items (16 sentences per child) providing an individual index of contrastive focus performance. The calculation of the Focus Accuracy Index (FAI) did not take into account whether the target

pitch accent typically associated with contrastive focus in European Portuguese, namely H*+L (Frota, 2002, 2014), was produced, as the analysis prioritized the correct location of prosodic prominence. It was decided that, from a pragmatic perspective, placing prominence on the appropriate lexical item was more relevant to the communicative success of the task than producing the exact pitch accent type. This procedure produced a single FAI value for each participant, allowing contrastive focus accuracy to be examined in relation to standardized individual measures such as receptive vocabulary (PPVT).

4.2.4. Plan of Analysis

In the descriptive overview, full (FAI = 1) and partial (FAI = 0.5) matches were retained to provide a fine-grained characterization of task performance. For the intonational analyses, these categories were collapsed, and utterances were classified as match (FAI = 1 and 0.5) or mismatch (FAI = 0) responses, allowing a comparison between successful and unsuccessful realizations of contrastive focus. Group differences in overall focus accuracy were assessed at the participant level using the Focus Accuracy Index (FAI); given violations of normality, differences between groups were evaluated using a Mann–Whitney U test. Differences in the distribution of prominence misplacement patterns and in prosodic strategies (PA and BT) were examined using chi-square tests of independence on utterance-level data, conducted separately for each constituent (N1, Adj and N2) and for matched and mismatched productions, with low-frequency categories collapsed when necessary to meet test assumptions. Finally, to assess whether group differences in focus accuracy persisted after accounting for language abilities, a linear regression analysis was conducted with FAI as the dependent variable, group as a predictor, and receptive vocabulary (PPVT) included as a covariate. Different units of analysis were adopted depending on the research question: participant-level analyses were used to address overall contrastive focus accuracy (RQ1) and its relation to language abilities (RQ4), whereas utterance-level analyses were employed to capture variation in prominence

misplacement patterns and prosodic strategies (RQ2 and RQ3). All analyses were conducted with statistical software Jamovi (version 2.3.28.0; Jamovi (2023)).

4.2.5. Reliability

To ensure annotation reliability, approximately 80% of the annotated corpus was independently reviewed by a second researcher with expertise in prosodic analysis and the P-ToBI framework for European Portuguese. The review focused on three main annotation dimensions: (i) prominence placement, (i.e., identification of the lexically emphasized constituent(s)); (ii) classification of productions as match, partial match, or mismatch responses, and (iii) labeling of intonational categories, including pitch accents and boundary tones. Particular attention was devoted to cases involving multiple prominence patterns, unequal prominence across constituents, and the identification of intonational categories according to the P-ToBI conventions for European Portuguese. Whenever differences in annotation emerged, these were jointly discussed by the researchers until consensus was reached. The final agreed-upon annotations were subsequently retained for all analyses.

4.3. Results

4.3.1. Sample characteristics and preliminary results

Based on 16 experimental items per child, the initial dataset consisted of 576 utterances. Seventeen utterances were excluded due to recording artifacts or disrupted productions, resulting in a final dataset of 559 utterances included in the analyses. Regarding overall task performance, across the 559 analyzed utterances, children with HFA produced 209 mismatch responses (74.1%), 56 fully correct match responses (19.9%), and 17 partially correct responses (6.0%). In the TD group, 131 utterances were classified as mismatches (47.3%), 124 as fully correct matches (44.8%), and 22 as partially correct matches (7.9%).

Regarding language abilities children with HFA scored lower on the PPVT than their TD

peers. Mean PPVT scores were 120 (SD = 35.4) for the HFA group and 151 (SD = 32.5) for the TD group. An independent-samples *t*-test confirmed that this difference was statistically significant ($t(34) = -2.67$, $p = .012$, Cohen's $d = -0.89$). Given this group difference, receptive vocabulary (PPVT) was included as a covariate in subsequent analyses of focus production accuracy.

4.3.2. Target Match: Focus Accuracy Index (HFA vs TD)

Data distribution of Focus Accuracy Index (FAI) divided by group can be seen in Figure 4.3. Mean FAI scores were lower in the HFA group (M = 0.23, SD = 0.30) than in the TD group (M = 0.48, SD = 0.29). Shapiro-Wilk tests indicated that FAI scores deviated from normality in the HFA group (W = 0.757, $p < .001$), whereas the TD group did not show evidence of non-normality (W = 0.953, $p = .470$). Homogeneity of variances was met (Levene's test: F(1, 34) = 0.024, p = .878). Given the violation of the normality assumption, group differences in FAI were examined using a non-parametric Mann-Whitney U test. Differences in FAI resulted statistically significant (Mann-Whitney U = 81.0, p = .011, rank-biserial correlation = 0.50).

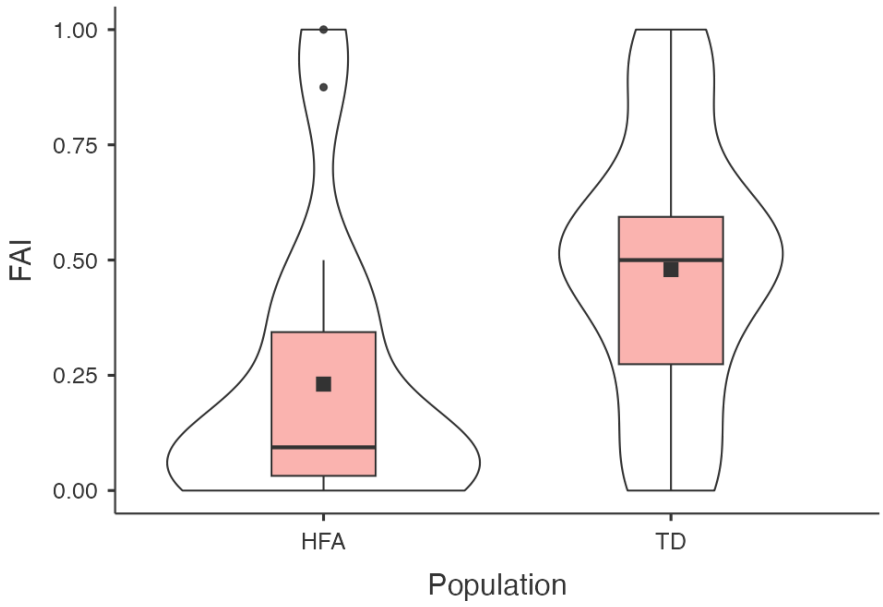


Figure 4.3. Distribution of Focus Accuracy Index (FAI) scores in the HFA and TD groups. Boxes represent the interquartile range, the horizontal line indicates the median, and whiskers represent the range of the data.

4.3.3. Target Mismatch: Error types (HFA vs TD)

Error types observed in mismatch responses are reported in Table 4.1. In the HFA group, the most frequent mismatch pattern was three-word prominence ("*3"*), accounting for 25.8% (n = 54) of mismatch utterances. The second most frequent error type was equal prominence on two words [*2(N1, Adj),* (19.6%; n = 41)] followed by a 13,9% (n = 29) of totally unrelated utterances [(0(df))]. Broad focus realization (BF) also occurred relatively frequently (12.9%, n = 27). Whereas the remaining error types were less frequent, and each accounted for less than 6% of production. In the TD group the most frequent error pattern was a single-word prominence on the adjective *(Adj)* (29.0%, n = 38), followed by equal prominence on N1 and Adj **(**28.2%, n = 37) and unequal prominence with greater emphasis on the adjective than the noun (*2 [N1 < Adj]*) (13.0%, n = 17). To meet the assumptions for a chi-square test of independence, low-frequency error categories were collapsed to ensure that no expected cell count was below 5; specifically, the categories "*0*, *0(df)*, *N2*, *2 (Adj, N2)*, *2 (N1, N2")*, and "*N1"* were merged into a single *Other* category, while the remaining categories were retained

as separate levels. Following this procedure, the distribution of mismatch error types differed significantly between groups ($\chi^2(6, N = 340) = 81.0$, $p < .001$, Cramer's $V = 0.505$).

Table 4.1. Frequencies and percentages of mismatch error types in the HFA and TD groups.

		Population		
Error_Type		**HFA**	**TD**	**Total**
3	Observed	54	14	68
	% within column	25.8%	10.7%	20.0%
2(N1, Adj)	Observed	41	37	78
	% within column	19.6%	28.2%	22.9%
2(Adj, N2)	Observed	3	0	3
	% within column	1.4%	0.0%	0.9%
2(N1, N2)	Observed	7	0	7
	% within column	3.3%	0.0%	2.1%
2(N1< Adj)	Observed	9	17	26
	% within column	4.3%	13.0%	7.6%
N1	Observed	3	3	6
	% within column	1.4%	2.3%	1.8%
Adj	Observed	11	38	49
	% within column	5.3%	29.0%	14.4%
N2	Observed	8	0	8
	% within column	3.8%	0.0%	2.4%
0	Observed	9	0	9
	% within column	4.3%	0.0%	2.6%
0(df)	Observed	29	0	29
	% within column	13.9%	0.0%	8.5%
0(q)	Observed	8	5	13
	% within column	3.8%	3.8%	3.8%
BF	Observed	27	17	44
	% within column	12.9%	13.0%	12.9%
Total	Observed	209	131	340
	% within column	100.0%	100.0%	100.0%

4.3.4. Prosodic Strategies for contrastive focus in Match and Mismatch responses.

4.3.4.1. Match responses: Pitch Accents in N1 and Adj

Observed frequencies of pitch accent (PA) types produced on N1 in matched productions are reported in Table 4.2. The table shows in both groups, H*+L emerged as the most frequent pitch accent on N1 in matched items, accounting for 48.6% of productions in the HFA group (n = 17) and 64.5% in the TD group (n = 40) The second most frequent category was H*, which was used slightly more often by the HFA group (n = 10) than by the TD group (n = 8). All remaining accent types occurred only rarely (each ≤ 8 productions). To meet the assumptions for a chi-square test of independence low-frequency categories (H+L*, L*+H, L+H*, T:) were collapsed into a single *Other* category for the subsequent analysis. The test revealed no statistical difference between TD and HFA children in the distribution of N1 pitch accents in matched items ($\chi^2(2, N= 97) = 3.93$, $p = .140$, Cramer's $V = 0.327$).

Pitch accent distributions on the adjective (Adj) in match productions are summarized in Table 4.3. On Adj, H*+L was again the most frequent pitch accent produced in matched items for both groups (TD: 84 productions; HFA: 42 productions). H* was the second most frequent PA (TD: 11; HFA: 6), followed by L*+H, which occurred less often but still above the minimum frequency threshold (TD: 7; HFA: 4). The remaining categories (H+L*, L+H*, T:) were extremely rare (≤ 4 productions each), and were therefore collapsed for analysis. The chi-square test indicated no significant group difference in the distribution of Adj accents ($\chi^2(5, N = 161) = 4.59$, $p = .467$, Cramer's $V = 0.169$). Overall distribution of PA produced when contrastive focus is correctly realized in N1 and Adj is illustrated in Figure 4.4.

Table 4.2. Observed frequencies and row percentages of pitch accent types produced on N1 (animal) in matched contrastive focus productions for the HFA and TD groups.

		Match responses: Pitch Accents on N1(Animal)						
Population		**H***	**H*+L**	**H+L***	**L*+H**	**L+H***	**T:**	**Total**
HFA	Observed	10	17	1	5	1	1	35
	% within row	28.6%	48.6%	2.9%	14.3%	2.9%	2.9%	100.0%
TD	Observed	8	40	7	2	1	4	62
	% within row	12.9%	64.5%	11.3%	3.2%	1.6%	6.5%	100.0%
Total	Observed	18	57	8	7	2	5	97
	% within row	18.6%	58.8%	8.2%	7.2%	2.1%	5.2%	100.0%

Table 4.3. Observed frequencies and row percentages of pitch accent types produced on Adj (Color) in matched contrastive focus productions for the HFA and TD groups.

		Match responses: Pitch Accents on Adj (Color)						
Population		**H***	**H*+L**	**H+L***	**L*+H**	**L+H***	**T:**	**Total**
HFA	Observed	6	42	0	4	3	0	55
	% within row	10.9%	76.4%	0.0%	7.3%	5.5%	0.0%	100.0%
TD	Observed	11	84	1	7	1	2	106
	% within row	10.4%	79.2%	0.9%	6.6%	0.9%	1.9%	100.0%
Total	Observed	17	126	1	11	4	2	161
	% within row	10.6%	78.3%	0.6%	6.8%	2.5%	1.2%	100.0%

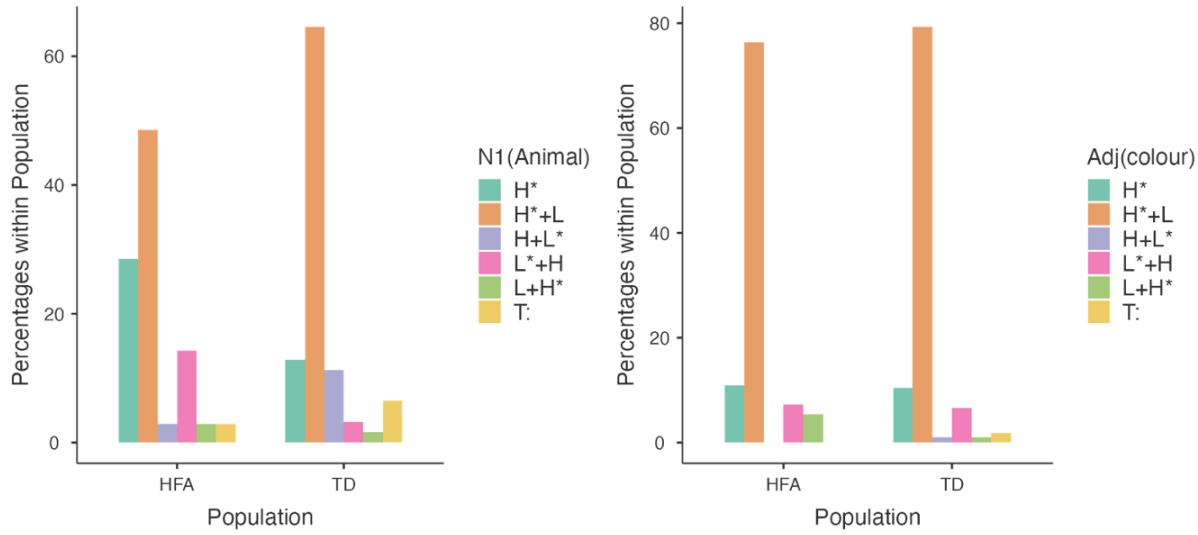


Figure 4.4. Bar plot showing the distribution of pitch accent types produced on N1 and Adj in matched contrastive focus productions for the HFA and TD groups. Percentages are calculated within group.

4.3.4.2. Mismatch responses: Pitch Accents in N1, Adj and N2

For mismatch productions on N1, pitch accent distributions are reported in Table 4.4. Across both groups H* was the most frequent accent, accounting for 33.6% of productions in the HFA group (n = 38) and 45.8% in the TD group (n = 33). The next most frequent strategies were T: (TD = 11; HFA = 25), L*+H (TD = 12; HFA = 21) with H*+L being comparatively less common when prominence was erroneously placed on N1 (TD = 12; HFA = 10). Lengthened syllables (T:) were used more than twice as often by children with HFA than by TD children when prominence was mistakenly realized on N1 (HFA = 25; TD = 11). To meet the assumptions of the chi-square test of independence, pitch accent categories with low expected frequencies were collapsed. Specifically, H+L* and L+H* were merged into a single *Other* category, while the remaining pitch accent types were retained as separate levels. Following this procedure, the different distribution of pitch accents in N1 resulted statistically significant ($\chi^2(4, N = 185) = 9.60$, $p = .048$, Cramer's V = 0.228).

Pitch accent distributions in cases of prominence misplacement on the adjective are illustrated in Table 4.5. In the HFA group, H* was the most frequently produced PA, accounting for 32.2% of productions (n = 38), followed by H*+L (22.9%, n = 27) and lengthened syllables (T:) (19.5%, n = 23). In contrast, in the TD group H*+L emerged as preferred PA, representing 42.5% of productions (n = 45), followed by H* (23.6%, n = 25) and L*+H (15.1%, n = 16). Other pitch accent types occurred with lower frequencies in both groups. Low-frequency pitch accent categories were again collapsed to satisfy the assumptions of the chi-square test, with H+L* and L+H* merged into a single *Other* category. Under these conditions, the distribution of pitch accents used when prominence was misplaced on the adjective differed between groups ($\chi^2(4, N = 224) = 14.6$, $p = .006$, Cramer's V = 0.255).

Finally, pitch accent distributions on the object (N2) are reported in Table 4.6. In the HFA group, productions were distributed across multiple pitch accent categories, with H*+L **(**30.1%**,** *n* = 31) and H+L* **(**26.2%**,** *n* = 27) occurring at comparable rates, alongside a substantial proportion of productions realized with less frequent pitch accents, including H***,** L***,** L*+H**,** L+H***,** and T:. TD children, however, preferred H+L* when prominence was misplaced on N2, which accounted for 63.3% of productions (*n* = 19), whereas H*+L occurred less frequently (10.0%**,** *n* = 3). Following category collapsing, the distribution of pitch accents on N2 differed significantly between groups, χ^2 (2, $N = 133$) = 14.6, $p < .001$, Cramer's V = 0.332). Figure 4.5 illustrates the overall distribution of pitch accent types in cases of prominence misplacement on N1, Adj, and N2.

Table 4.4. Observed frequencies and row percentages of pitch accent types produced on N1 when prominence was misplaced, for the HFA and TD groups.

		Mismatch responses: Pitch Accents on N1(Animal)						
Population		**H***	**H*+L**	**H+L***	**L*+H**	**L+H***	**T:**	**Total**
HFA	Observed	38	10	7	21	12	25	113
	% within row	33.6%	8.8%	6.2%	18.6%	10.6%	22.1%	100.0%
TD	Observed	33	12	4	12	0	11	72
	% within row	45.8%	16.7%	5.6%	16.7%	0.0%	15.3%	100.0%
Total	Observed	71	22	11	33	12	36	185
	% within row	38.4%	11.9%	5.9%	17.8%	6.5%	19.5%	100.0%

Table 4.5. Observed frequencies and row percentages of pitch accent types produced on Adj when prominence was misplaced, for the HFA and TD groups

		Mismatch responses: Pitch Accents on Adj (Colour)						
Population		**H***	**H*+L**	**H+L***	**L*+H**	**L+H***	**T:**	**Total**
HFA	Observed	38	27	6	16	8	23	118
	% within row	32.2%	22.9%	5.1%	13.6%	6.8%	19.5%	100.0%
TD	Observed	25	45	2	16	1	17	106
	% within row	23.6%	42.5%	1.9%	15.1%	0.9%	16.0%	100.0%
Total	Observed	63	72	8	32	9	40	224
	% within row	28.1%	32.1%	3.6%	14.3%	4.0%	17.9%	100.0%

Table 4.6. Observed frequencies and row percentages of pitch accent types produced on N2 when prominence was misplaced, for the HFA and TD groups

Population		Mismatch responses: Pitch Accents on N2 (Object)							
		H*	H*+L	H+L*	L*	L*+H	L+H*	T:	Total
HFA	Observed	12	31	27	7	11	1	14	103
	% within row	11.7 %	30.1 %	26.2 %	6.8%	10.7 %	1.0 %	13.6 %	100.0 %
TD	Observed	2	3	19	5	1	0	0	30
	% within row	6.7%	10.0 %	63.3 %	16.7 %	3.3%	0.0 %	0.0%	100.0 %
Total	Observed	14	34	46	12	12	1	14	133
	% within row	10.5 %	25.6 %	34.6 %	9.0%	9.0%	0.8 %	10.5 %	100.0 %

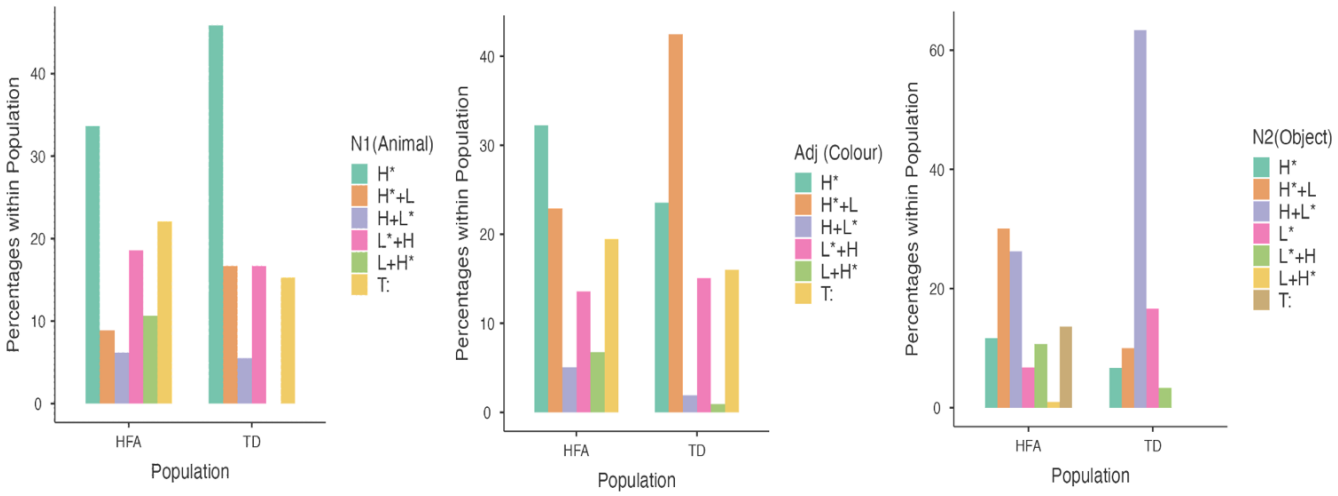


Figure 4.5. Bar plot showing the distribution of pitch accent types produced when prominence was misplaced in N1, Adj and N2 in contrastive focus productions for the HFA and TD groups. Percentages are calculated within group.

4.3.4.3. Boundary Tones in Match and Mismatch responses

In matched responses, boundary tone use was highly consistent across groups. In both HFA and TD children, productions most frequently ended with L%, accounting for 95.9% of utterances in the HFA group (n = 70) and 95.2% in the TD group (n = 139). The use of H%

and !H% was rare overall, with only 3 productions in the HFA group (1 H% and 2 !H%) and 7 productions in the TD group (all H%). In mismatched responses, while L% remained the most frequent boundary tone in both groups, its use was reduced in the HFA group (83.7%, $n = 175$) compared to TD children (96.9%, $n = 127$). Conversely, H% occurred more frequently in the HFA group (15.3%, $n = 32$) than in the TD group (3.1%, $n = 4$), whereas !H% remained extremely rare (HFA= 2 productions; TD = 0). Given that H% and !H% were very sparse as categories they were collapsed into a single *Other* category for the chi-square analysis of matched responses. Under this classification, boundary tone distributions did not differ between groups in matched responses ($\chi^2(1, N = 219) = 0.052$, $p = .819$) but differed when prominence was misplaced ($\chi^2(2, N = 340) = 14.3$, $p < .001$).

4.3.4.4. Focus Accuracy Index and Receptive Vocabulary

Given the group difference observed in receptive vocabulary (PPVT), an additional analysis was conducted to examine whether group differences in focus production accuracy persisted after controlling for PPVT. A linear regression model was fitted with Focus Accuracy Index (FAI) as the dependent variable, group (HFA vs. TD) as a predictor, and PPVT as a covariate. The model accounted for a moderate proportion of variance in focus accuracy ($R^2 = .321$). After controlling for PPVT, group membership was no longer a significant predictor of FAI ($\beta = 0.132$, SE = 0.099, $t = 1.33$, $p = .192$). In contrast**,** PPVT emerged as a significant positive predictor of focus accuracy ($\beta = 0.0039$, SE = 0.0014, $t = 2.83$, $p = .008$).

4.4. Discussion

A number of studies have indicated that the speech of children with ASD is characterized by unusual prosody (Fosnot & Jun, 1999; Grossman et al., 2010; Grice et al., 2023). In particular, the production of contrastive stress has been reported to be considerably impaired, as it relies more on contextual and pragmatic information rather than on grammatical rules (Fine et al.,

1991; Nadig & Shaw 2015; McCann et al., 2007). Although previous research on this population has reported inappropriate use of stress (Shriberg et al., 2001, Peppé et al., 2006), it is still unclear where prominence is misplaced within the utterance. Moreover, almost no studies have provided a detailed phonological account of the intonation patterns of CF in children with Autism Spectrum Disorder. Using the Focus Expression subtest of the PEPS-C, the present study investigated how children with High-Functioning Autism (HFA) and typically developing (TD) peers produced contrastive focus (CF) in European Portuguese (EP), with a specific focus on (i) accuracy of prominence placement in CF, (ii) patterns of prominence misplacement, (iii) strategies used to mark focus in both successful and unsuccessful productions, and (iv) the role of receptive vocabulary in explaining differences in focus production.

With respect to our first research question, the initial hypothesis was supported: children with ASD showed lower accuracy in placing prominence on the intended target during the contrastive focus task, producing misplacement patterns in approximately 74% of utterances (vs. TD: 43.7%) However, critically, this group difference did not persist once receptive vocabulary was taken into account. When PPVT scores were included as a covariate, group membership was no longer a significant predictor of FAI, indicating that variability in contrastive focus production accuracy was better explained by language abilities than by diagnostic group per se. This result directly addresses our fourth research question and suggests that reduced focus accuracy in HFA is not independent of broader linguistic competence. Consequently, our findings contrast with studies reporting comparable contrastive focus performance between ASD and TD children, in which no substantial group differences were observed when groups were matched on language abilities (Jarvinen-Pasley et al., 2008; Diehl & Paul, 2011). However, Shriberg et al. (2001) reported differences in speech performance between ASD and TD children even when groups were matched on vocabulary measures,

specifically in the contrastive focus task. Conversely, Filipe et al. (2016) found comparable performance in contrastive focus production between ASD and TD children despite the absence of vocabulary matching. In this context, the present findings align with a growing body of work indicating that prosodic functions are indeed closely related to language abilities (DePape et al., 2012). An additional point of interest concerns the performance of the TD group. Despite expectations of high levels of accuracy, approximately half of the TD children's productions were classified as mismatches (47.3%), indicating that the CF task was challenging even for TD speakers. Previous developmental research has shown that the ability to produce contrastive focus continues to develop throughout the school years (Chen, 2011), and that its developmental trajectory varies across languages. In English, focus marking through prosodic prominence appears to be in place by around five years of age (Wells et al., 2004). In contrast, in Spanish, where CF is primarily expressed through word order rather than intonation, adult-like production of CF has been reported much later, around adolescence (Martínez-Castilla and Peppé, 2008). In Italian, it is still unclear at what age this ability fully develops, with estimates ranging from around 10 years to adulthood (Zanchi et al., 2022). Regarding European Portuguese (EP), Filipe and colleagues (2016) showed that from a developmental point of view adult-like production of CF may only be fully established between 18 and 20 years of age. Given that the age range of the present sample spanned from 5 to 15 years, it is plausible that even TD children were still in the process of consolidating contrastive focus production, which may account for the unexpected proportion of mismatches observed. It has also been argued that EP places particular demands on the production of CF, as focus is conveyed through intonational cues and focused elements are often followed by additional post-focal accentuation rather than by complete deaccenting (Frota, 2014; Frota & Prieto, 2015a). Consequently, producing contrastive focus in European Portuguese requires children to manage both focal and post-focal prosodic structure, which may increase the difficulty of the

PEPS-C Focus Production task for Portuguese-speaking children. These already high demands for TD may be particularly challenging for children with ASD, for whom difficulties in executive functioning have been also reported (Filipe et al., 2018). Accordingly, regarding our second research question, the two groups differed markedly in their patterns of prominence misplacement. Whereas TD children most often misplaced prominence by accenting a single non-target word, typically the adjective, children with HFA tended to distribute prominence across the utterance, with the most frequent error type involving simultaneous emphasis on three constituents. In addition, children with ASD exhibited greater variability in error types than TD peers, including flat intonation contours, prominence placement on non-target constituents such as N2 or Adj+N2 combinations, and utterances that were completely unrelated to the task (14% of total HFA productions). In our sample CF was overall easier to realize when the adjective was the intended target rather than the subject noun. To our knowledge, only a limited number of studies have examined lexical factors in contrastive stress production in ASD, most notably those by Baltaxe (1984; Baltaxe and Guthrie, 1987) and Peppé et al. (2007). Baltaxe reported, like in our study, a high proportion of simultaneous stress misassignment in children with ASD, and further suggested that contrastive stress sentences were easier for the ASD group when prominence fell on the first constituent of the sentence (i.e., the subject in an SVO structure). This preference was interpreted as reflecting an egocentric stage of cognitive development in the children. The later study by Peppé and colleagues, also using the PEPS-C task, confirmed Baltaxe's findings, showing that children with ASD made fewer errors when prominence was placed on the adjective, which in English occurs in sentence-initial position as the first lexical word within the noun phrase *(e.g., the GREEN cow has the ball)*. In the present study, the adjective was likewise favored in mismatch responses; however, in EP adjectives occur post-nominally within the noun phrase [e.g., *ovelha VERMELHA* ('sheep- red') *tem a bola* (has the ball)]. We hypothesize that this pattern does

not reflect a preference for the adjective as a lexical category per se, but rather an effect of prosodic position within the utterance. Crucially, EP is a right-headed language in which the nuclear accent in broad focus contexts typically falls toward the right edge of the sentence and similarly the most prominent word within a prosodic phrase is the one at its right-edge (Frota, 2014). From this perspective, placing prominence in a non-default position, such as early in the utterance and towards the left edge of the prosodic phrase as required by the PEPS-C contrastive focus task, may be inherently more demanding, potentially leading both TD and HFA children to favor later constituents in the utterance. This interpretation is further supported by the findings related to our third research question.

The analysis of prosodic strategies revealed that group differences emerged primarily when prominence was misplaced, rather than when it was correctly assigned. In matched productions, in fact, both HFA and TD children relied mostly on the pitch accent typically associated with contrastive focus in European Portuguese (H*+L: Frota, 2014*).* This result suggests that when the focus is correctly realized (i.e., match responses), TD and HFA speakers rely mostly on the same prosodic strategy, despite some variability in the HFA group, which also produced a higher proportion of H* accents on N1.

A different pattern emerged when children did not place prominence on the correct target word. In mismatched productions, prosodic strategies appear to be different according to the position on which prominence was misplaced. When prominence was erroneously assigned to N1, both groups most frequently produced H*, with limited use of the target H*+L (HFA: n = 10; TD: n = 12) suggesting that, when prominence was placed early in the utterance, children tended to rely on a simpler accentual strategy from an articulatory-phonetic perspective (monotonal vs bitonal). On the adjective, TD children tended to preserve the target contrastive focus accent H*+L even when the focus location was incorrect, indicating that they retained knowledge of the appropriate accentual form despite the misalignment. In contrast, children

with ASD continued to rely more heavily on H*, although use of H*+L increased relative to N1 (n = 27 vs n = 38). When prominence was misplaced on N2 (a constituent that should not have carried prominence at all) children in the HFA group showed an increased use of the target pitch accent H*+L (n = 31), alongside frequent production of H+L*, the falling accent typically associated with broad-focus declaratives in EP (Frota, 2014). TD children predominantly produced H+L* in this position (n = 19): this is a pattern that is prosodically well-formed in European Portuguese, where post focal constituents are not obligatorily deaccented (Frota, 2000; 2002; Frota and Prieto, 2015a). Overall, these findings point to a positional effect in mismatch contexts: as prominence shifted toward the right edge of the utterance, children, particularly in the HFA group, more frequently produced canonical accentual patterns, including the target H*+L and the default declarative H+L*. With respect to lengthened syllables (T:) no dedicated statistical analyses were conducted to assess their functional status as prosodic strategy; however descriptive patterns indicate that they occurred mainly in mismatch conditions on N1 and the adjective and were more frequent in the HFA group than in TD children. Finally, the analysis of boundary tones showed different patterns in matched versus unmatched productions. When children placed prominence correctly, both groups behaved similarly and relied almost exclusively on L%, the expected boundary tone for CF in statements. However, in unmatched productions, although L% remained the most common boundary in both populations, HFA children produced many more H% endings than TD children. It is likely that, for HFA speakers, misplacing prominence also disrupts the stability of phrase-final marking. TD children, on the other hand, typically produced the low boundary tone even when the prominence was misplaced, proving that they preserved the proper phrase-final marking despite the accentual mistake.

In summary, children with HFA showed lower accuracy in identifying the intended focus target than TD peers; however, this difference disappeared once receptive vocabulary was

considered, indicating that group differences in accuracy were closely tied to language abilities. Nonetheless, qualitative differences remained in error patterns: when prominence was misplaced, children with HFA tended to distribute prominence across multiple constituents, whereas TD children more often produced simpler errors involving a single non-target word (the adjective). Importantly, when children with HFA successfully placed prominence on the target, they largely employed the same target CF accent as TD children, although such target-like realizations were relatively infrequent. In mismatch conditions, group differences were observed, alongside a positional trend where prosodic realizations became more target-like as prominence moves toward the end of the utterance. For children with ASD, placing prominence outside the default position may be particularly challenging, given the need to also maintain post-focal accentuation. From a clinical perspective, prosody plays a key role in successful communication, and contrastive focus is particularly important for signaling a speaker's intended contrasts and beliefs. The present fine-grained analysis shows that moving beyond overall accuracy is informative, as it reveals how children with HFA mismark focus and, crucially, where intervention efforts should be directed. Considering the limited availability of prosody focused interventions in ASD (Filipe, 2014), these findings point to the need for ad hoc training aimed at strengthening prosody-pragmatics mapping, especially the ability to constrain prominence to a single constituent in early position in the sentence. However, this study has some limitations. The PEPS-C Focus Production task is highly structured and repetitive and therefore may not fully capture the complexity of prosodic behavior in spontaneous interaction. It has been argued that PEPS-C tasks are relatively easy and that children with ASD tend to perform better in structured settings than in naturalistic contexts (Boraston & Blakemore, 2007; Nadig et al., 2010). However, this was not the case in the present study, where both groups showed difficulties despite the controlled nature of the task. The repetitive and constrained format of the PEPS-C may represent a potential strength, as it

could be exploited as a training tool, for instance, to practice placing prominence in non-default, sentence-initial positions. Additional limitations concern the statistical approach adopted for the analyses of prominence misplacement patterns and prosodic strategies (RQ2 and RQ3). These analyses were conducted at the utterance level using chi-square tests in order to capture variation in error types and intonational choices. While this approach allowed a detailed characterization of prosodic behavior, it does not explicitly model the hierarchical structure of the data, whereby multiple utterances are produced by the same speaker and may share speaker-specific properties. For this reason, results related to RQ2 and RQ3 should be interpreted with caution and regarded as exploratory. Relatedly, given the number of chi-square tests performed, no correction for multiple comparisons was applied. Future studies should address these limitations by adopting analytical approaches that account for the nested nature of the data, such as aggregating measures at the participant level or using mixed-effects models with utterances nested within speakers and items, allowing a more conservative assessment of group differences. Future studies should also include cross-cultural comparisons and carefully matching groups on language abilities.

4.5. Conclusions

Salience of contrast is reduced in children with HFA with simultaneous prominence on multiple words and more variable intonational categories. Focus production was overall more accurate when prominence occurred toward the right edge of the utterance, reflecting phrase-final prominence in European Portuguese.

4.6. Data Availability Statement

The full annotated corpus, including audio recordings from the PEPS-C Focus Expression task and the corresponding prosodic annotations, is openly available on the Open Science Framework (OSF) at: https://doi.org/10.17605/OSF.IO/QR74X.

CHAPTER 5 | Persuasive Prosody of Italian TV Commercials: An Acoustic Brief Study.

D'Aloia V., Di Berardino C., Palumbo R., Fasolo M.

Abstract

Advertising is a form of persuasive communication aimed at shaping consumer attention, attitudes, and behavioral intentions within highly constrained temporal windows. While a large body of research has examined linguistic and rhetorical strategies in advertising, acoustic studies of broadcast media remain limited. This gap is particularly evident in Italian, where the prosodic realization of persuasive discourse in TV commercials has not yet been investigated. The present study investigates whether and how Italian advertising speech differ acoustically from non-persuasive broadcast speech, focusing on prosodic features associated with pitch variability, temporal structuring, and speech tempo. Using the CLIPS corpus (Corpora e Lessici dell'Italiano Parlato e Scritto), Italian TV advertising clips (ADV) and informational-cultural broadcast clips (INFO) were analyzed. A total of 50 speech units (25 ADV, 25 INFO) were selected and analyzed using Praat. The interpausal unit (IPU) was adopted as the unit of analysis to allow for comparable segmentation across speech types. Three prosodic measures were extracted: F0 range (in semitones), IPU duration (in seconds), and speech rate (syllables per second). Results showed that advertising speech exhibited greater pitch variability and shorter speech units than informational speech, while no differences emerged in speech rate between the two conditions. These findings suggest that persuasive discourse in Italian television does not rely on faster delivery, but rather on strategically timed units and prosodic cues linked to emotional impact.

Key words: *persuasive speech, prosody, Italian Tv commercials, advertising, acoustics*.

5.1. Introduction

Advertising is a form of persuasive communication that aims to inform potential customers about products or services while also encouraging purchase behavior (Leech, 1966). Because advertising messages are delivered within very limited time windows, they must quickly capture attention, maintain engagement, and prompt action (Kanner, 1999; Fletcher, 2008). Research on language in advertising has traditionally focused on lexical and syntactic strategies used to achieve these goals. Studies have described the extensive use of ellipsis [e.g. "*Communication Unlimited*" (Motorola, America)], short and simple words, slogans, and repetition, as well as phonological devices such as rhyme and alliteration, all of which contribute to attention capture and brand recall (McQuarrie & Mick, 1996; Zou, 2018). In parallel, phonetics and psychology studies suggested that the voice itself is a major channel of persuasive intent. Notably reviews of persuasive communication underlined that vocal delivery conveys information beyond the content of communication and intentions are often associated with differences in prosodic and vocal characteristics (Guyer et al, 2019; Gomes et al., 2023).

From an acoustic perspective, several prosodic features have been identified as particularly relevant for the perception of persuasion, most notably speech rate (i.e. tempo), fundamental frequency (F0, i.e. pitch), and F0 contour (Blanc & Dominey, 2003; Grichkovtsova et al., 2009). Evidence from political communication showed that vocal characteristics can influence electoral outcomes, with lower pitch being consistently associated with leadership traits and perceived competence (Klofstad & Anderson, 2018). Similar effects have been reported in commercial contexts: salespeople using falling pitch contours outperform those using rising contours, with sentence-final pitch lowering being linked to higher sales performance (Peterson et al., 1995). In advertising, attitudes toward brands have been shown to be more favorable when the advertising voice has a lower pitch and perceived as masculine (Montrey, 2005; Pogačar et al., 2018), whereas higher F0 values are generally perceived as less attractive in

vocal perception research (Feinberg et al., 2005; Quené et al., 2016). At the same time, increased pitch variability, measured as a wider F0 range, have been associated with greater perceived charisma (Berger et al., 2017). Supporting this view, Niebuhr and colleagues (2016) analyzed acoustic features in product presentations by former Apple CEO Steve Jobs, widely regarded as a highly charismatic speaker, and found that his speech was characterized by an F0 range approximately six semitones wider than reference values. In addition to pitch-related cues, speech rate has been shown as well to positively influence persuasiveness. Faster speech has been associated with higher perceived confidence, competence, and communicator effectiveness (Peterson et al., 1995; Charoenruk & Olson, 2018), and studies report greater persuasive impact for messages delivered at higher rates compared to slower speech (Apple et al., 1979; Guyer et al., 2019). At the same time, findings are mixed, as other works suggested that excessively fast speech may delay message comprehension and reduce persuasion by impairing information processing (Smith & Shaffer, 1995; Moore et al., 1986).

Despite this converging evidence on voice and persuasion, acoustic studies of broadcast media, and advertising speech in particular, remain relatively limited. Existing work indicates that broadcast registers tend to adopt a highly dynamic speaking style, characterized by elevated pitch levels, increased pitch variability, emphatic intonation, fast speech rate, and short pauses, all aimed at capturing and sustaining listener attention (Cotter, 1993; Castro et al., 2010). More specific research by Rodero and colleagues (2017) on radio advertising showed that commercials displaying a high-low (HL) pitch pattern in declarative sentences, are associated with stronger sympathetic nervous system activation in listeners as well as enhanced recall and recognition of the message. The same line of research also demonstrated that speech tempo plays a role in message processing, with a rate of approximately 180 words per minute (WPM) emerging as optimal in radio commercials and yielding higher levels of comprehension and information retention (Rodero, 2020).

Crucially for the present work, Italian media communication remains a major blind spot. One of the few available studies on Italian broadcast media is the diachronic analysis of television newsreading by Savino and colleagues, which shows that contemporary broadcasts are characterized by faster speech rates, longer sequences of continuous speech, and fewer pauses. These changes pointed to a reorganization of temporal structuring in broadcast speech over time, while overall melodic complexity remains largely stable (Savino et al., 2024).

Despite these findings, Italian advertising speech remains unexplored from an acoustic perspective. This gap is particularly striking given the availability of broadcast corpora that include advertising material. The CLIPS corpus (*Corpora e Lessici dell'Italiano Parlato e Scritto;* Albano Leoni et al., 2006; Savy & Cotugno, 2009), for instance, provides structured broadcast data comprising television advertising alongside other speech genres, yet it has not been exploited for acoustic research of Italian persuasive prosody in TV commercials. The present study addresses these gaps by examining the prosodic correlates of persuasive speech in Italian television commercials through a comparative acoustic approach. We compare advertising speech with informational-cultural broadcast speech to test whether persuasive commercials in Italian are characterized by (i) greater pitch variability, (ii) differences in temporal structuring, and (iii) differences in speech tempo.

5.2. Methods

5.2.1. Materials

For the purpose of the study, we used the CLIPS corpus *(Corpora e Lessici dell'Italiano Parlato e Scritto)* a linguistically stratified spoken Italian corpus developed within the CLIPS project coordinated at the University of Naples "Federico II" between 2000 and 2006 and publicly available online (Abano Leoni et al., 2006; Savy & Cotugno, 2009). CLIPS includes approximately 100 hours of naturally occurring spoken Italian balanced by sex of speakers,

sampled across multiple communicative contexts (broadcast, read, dialogic and telephone speech) and designed to capture diatopic variations through recordings of 15 different localities. The hierarchical structure of the corpus consists of five main sub-corpora: Broadcast (*Radiotelevisivo*), Dialogic (*Dialogico*), Read (*Letto*) and Telephone speech (*Ortofonico*). Within the Broadcast sub-corpus, recordings are further distinguished by medium (television vs. radio). For the present analysis, we selected television broadcasts only and restricted the sample to the *national* subset to minimize regional dialectal variation. From the *national* television subsection, we focused on the PB (*publicità*) and DC (*divulgazione e cultura*) categories, which correspond to advertisement and informational-cultural speech respectively.

The sample consisted of five advertisement clips (ADV; PB) and five informational-cultural clips (INFO; DC), covering different topics. Within each clip, we selected a set of speech segments for a total of 50 units (25 ADV and 25 INFO; four produced by female speakers, evenly distributed across speech types). Unit selection was not random but guided by functional relevance. In advertisement clips, we included segments containing the product or brand name, slogans, and, where present, call-to-action elements (e.g. "*scegli ora*" ("choose now"); Stewart, 2006). In informational clips, we selected spoken units introducing the main topic, presenting key informational content (e.g., definitions), and expressions inviting the listener to engage with the topic (e.g., "*andiamo a scoprire il Colosseo*" ("let's explore the Colosseum").

Considering the different communicative goals between advertising and informational speech (persuasion vs information delivery), we adopted the interpausal unit (IPU) as the unit of analysis, a standard unit in corpus-based spoken-language research and the unit adopted in the available acoustic study of Italian broadcast speech (Savino et al., 2024). An interpausal unit (IPU) is defined as a stretch of speech bounded by presence of pauses (Koiso et al., 1998; Buhmann et al., 2002) and has been argued to be related to speech planning processes

(Gerstenberg et al. 2018). Because IPUs are delimited by pauses, their duration provides an index of how speech is temporally segmented into planning units, reflecting both the organization of the speech flow and cognitive load during production (Oomen & Postma, 2001) In line with Savino et al. (2024), in the present study pauses were operationalized as silences ≥ 200 milliseconds (ms). Given the possible presence of background music and non-speech audio in both advertising and informational-cultural clips, we restricted the sample to recordings in which speech was clearly audible and not overlapped or masked by music or other acoustic interference. This criterion ensured reliable segmentation and analysis of the speech signal.

5.2.2. Procedure

Portions of speech delimited by silent pauses of at least 200 ms. were identified and annotated as interpausal units (IPUs) using Praat software (Boersma & Weenink; 2009). A total of 50 IPUs were analyzed, and annotation was carried out on three Praat tiers according to Figure 5.1: the lowest tier contained the full orthographic transcription of each IPU, the intermediate tier included word-level segmentation for segmentation support, and the top tier contained syllable-level segmentation. For each IPU, the following prosodic measures were extracted:

- F0 range (*pitch variability*): measured as the span between the maximum and minimum F0 values across the entire IPU, expressed in semitones. Fundamental frequency values were extracted in Hertz in Praat and subsequently converted to semitones using a logarithmic transformation based on the difference between maximum and minimum F0 values (Snow & Balog, 2002).
- IPU duration (*temporal structuring*): overall duration of the interpausal unit, automatically computed in Praat as the temporal extent of the unit and expressed in seconds. As IPUs are defined by the presence of pauses, their duration reflects how speech is chunked into temporally organized units.

- Speech rate (*speech tempo*; syllables per second, syll/s): calculated as the number of syllables produced within each IPU divided by its duration (Zmarich, 1996). A syllable-based measure was preferred over word-based measures because Italian is commonly described as a syllable-timed language (Bertinetto, 1981) and because syll/s provides a more fine-grained and robust estimate of speech tempo, avoiding confounds related to word length variability (monosyllabic vs polysyllabic). Syllables were manually segmented based on the actual phonic chain rather than on the syllabification of lexical forms, then extracted using a Praat script and automatically counted.

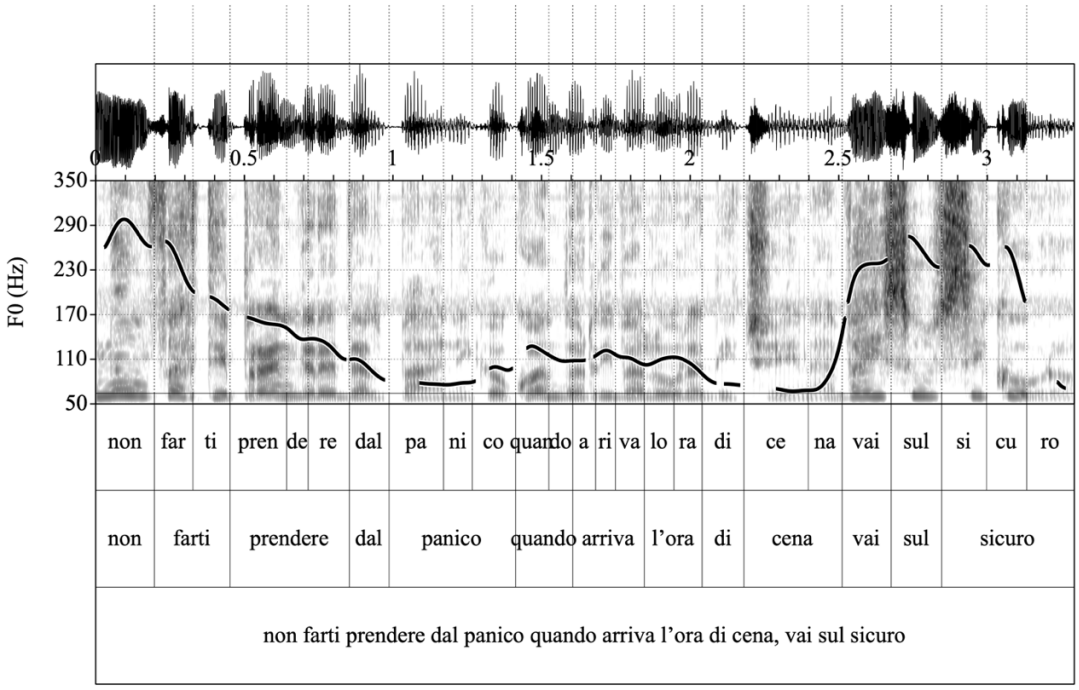


Figure 5.1. Example of IPU annotation in Praat showing the three annotation tiers: orthographic transcription (bottom), word-level segmentation (intermediate), and syllable-level segmentation (top).

5.2.3. Reliability

Inter-coder reliability was assessed on a randomly selected subset comprising 20% of the annotated material. A second trained annotator independently identified interpausal unit (IPU) boundaries and performed syllable segmentation on this subset. Agreement on IPU delimitation

was high (Cohen's K = .88), indicating consistent identification of pause-defined speech units. Reliability for syllable segmentation was assessed by correlating the number of syllables per unit across annotators, yielding a strong correlation (Pearson's r = .91). In cases of disagreement, annotations were discussed and resolved by consensus.

5.2.4. Plan of Analysis

To test whether advertising and informational speech differ in prosodic realization, three linear mixed-effects models were fitted, with F0 range (semitones), IPU duration (seconds) and speech rate (syllables per second) as dependent variables, indexing pitch variability, temporal structuring, and speech tempo, respectively. In all models, *Speech Type* (Advertisement vs. Informational) was included as a fixed effect, and a random intercept for *Clip* was added to account for the non-independence of IPUs extracted from the same audiovisual clip and to estimate the contribution of clip-level variability to the overall variance. All analyses were conducted with statistical software Jamovi (version 2.3.28.0; Jamovi, 2023)).

5.3. Results

The distribution of prosodic measures across advertising and informational speech is summarized in Table 5.1 and illustrated in Figure 5.2. Inspection of the distributions and Shapiro-Wilk tests indicated no significant deviations from normality for any of the prosodic measures in either speech type (all *p*s > .05; see Table 5.1). Linear mixed-effects models revealed clear speech type-related differences in pitch variability and temporal structuring, but not in speech rate. Specifically, advertising IPUs exhibited substantially greater pitch variability than informational IPUs, with a higher mean F0 range in ADV speech (M = 18.37 st, SD = 5.08) than in INFO speech (M = 10.22 st, SD = 2.81). This difference was confirmed by the mixed-effects model, which showed a strong effect of *Speech Type* on F0 range (β = 8.15 st, SE = 1.18, t = 6.93, p < .001), with speech type accounting for a large proportion of

explained variance (marginal R^2 = .50). In contrast, clip-level variability was minimal for this measure (ICC = .006), indicating that pitch variability differences were primarily driven by speech condition rather than by individual clips.

Temporal structuring also differed between conditions. Informational IPUs were longer on average (M = 3.87 s, SD = 1.50) than advertising IPUs (M = 2.44 s, SD = 1.31), and this pattern was supported by the mixed-effects analysis, which revealed a significant effect of *Speech Type* on IPU duration (β = 1.38 s, SE = 0.51, t = 2.72, p = .026). For this measure, *Speech Type* accounted for a moderate proportion of variance (marginal R^2 = .19), while clip-level variability contributed a small but meaningful share of the total variance (ICC = .14).

By contrast, speech rate did not differ between the two conditions. Mean speech rate was 6.57 syll/s (SD = 1.71) in ADV IPUs and 6.49 syll/s (SD = 1.37) in INFO IPUs, with largely overlapping distributions. Consistent with this pattern, the mixed-effects model showed no effect of *Speech Type* on speech rate (β = 0.08 syll/s, SE = 0.44, t = 0.18, p = .861), with no variance explained by Speech Type (marginal R^2 = .001) and no clip-related variability (ICC = 0).

Table 5.1: Means, standard deviations, ranges, and Shapiro-Wilk normality test results for the three prosodic measures (F0 range, IPU duration, and speech rate) across ADV (advertising) and INFO (informational) speech.

							Shapiro-Wilk	
	Speech Type	N	Mean	SD	Minimum	Maximum	W	p
F0range (semitones)	ADV	25	18.37	5.08	10.544	27.29	0.952	.282
	INFO	25	10.22	2.81	3.664	15.18	0.980	.886
IPU_Duration (s)	ADV	25	2.44	1.31	0.860	6.04	0.921	.054
	INFO	25	3.87	1.50	1.910	7.85	0.928	.078
Speech rate (Syll/s)	ADV	25	6.57	1.71	2.811	10.82	0.977	.831
	INFO	25	6.49	1.37	3.465	8.38	0.951	.268

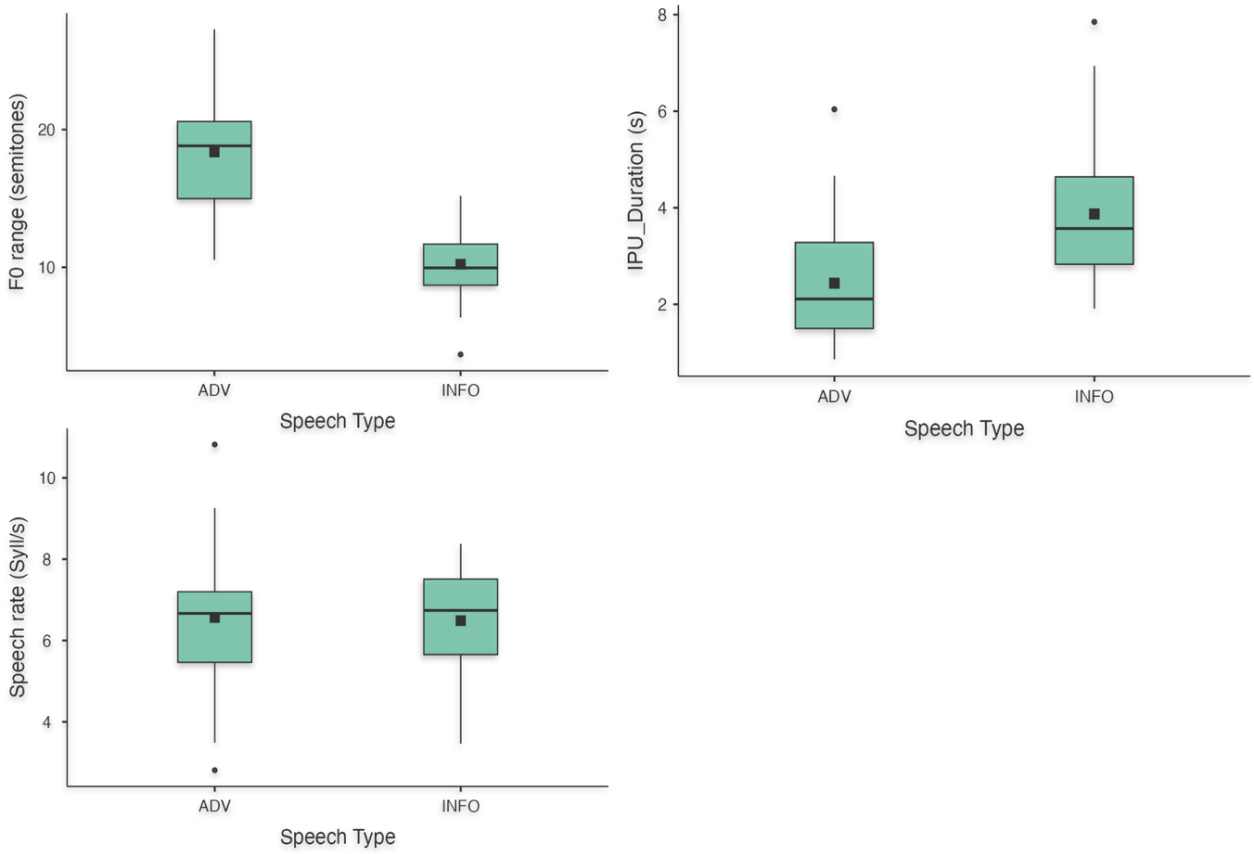


Figure 5.2. Distribution of F0 range (semitones), speech rate (syllables per second), and IPU duration (seconds) across advertising and informational speech. Boxplots show the interquartile range (boxes) and the median (horizontal line); squares indicate mean values.

5.4. Discussion

The aim of this study was to investigate whether Italian television advertising differed acoustically from non-persuasive informational-cultural speech, focusing on prosodic features associated with pitch variability, temporal structuring, and speech rate. By comparing the two speech types, the study sought to identify prosodic strategies that may characterize persuasive communication in Italian broadcast media. Advertising IPUs exhibited a substantially wider F0 range than informational IPUs, with a mean difference of approximately 8 semitones, corresponding to nearly an octave of additional pitch span, indicating a more dynamically modulated pitch profile. Importantly, this effect was consistent across clips, indicating that increased pitch variability characterizes advertising speech in general rather than specific commercials. This finding aligns closely with previous research showing that expanded pitch

range is a salient feature of persuasive and attention-oriented speech across different domains, including political communication, public speaking, and advertising (Berger, 2017; Rodero, 2017). It is worth noting that another speech register characterized by increased pitch variability is infant-directed speech (IDS). In this register, exaggerated pitch modulation has been consistently associated with the expression of positive emotional states and has been shown to make speech more engaging and emotionally attractive, thereby contributing to the creation and maintenance of an emotional bond between caregivers and infants (Spinelli et al., 2017). In this respect, the use of expanded pitch range in advertising speech appears to rely on acoustic cues that are closely linked to emotional expression. Consistent with this view, Valigura et al. (2020) argue that a defining feature of television advertising, compared to other informational media genres, is its emotional impact, which contributes to shaping a mental image in a person's mind, and relies more on a subconscious sphere rather a rational one.

A second key finding concerns temporal structuring, as indexed by interpausal unit duration. IPUs in television commercials were significantly shorter than those in informational speech, indicating a more segmented temporal organization of speech. This pattern suggests that advertising speech may rely on shorter continuous speech spans, potentially reflecting a tighter organization of the spoken message within limited time constraints to facilitate rapid processing (Fletcher, 2008). Direct comparisons with previous advertising studies are not possible, as IPU duration has not been systematically examined in this domain. However, evidence from other spoken-language contexts indicates that IPU duration is sensitive to differences in speech planning and production. Shorter IPUs have been reported in less proficient second-language speakers compared to more proficient ones (Judkins, 2022), as well as in older speakers relative to younger adults (Gerstenberg et al., 2018) pointing to IPU duration as a marker that may be sensitive to speech planning and cognitive chunking processes. An interesting comparison emerges when considering the study by Savino and

colleagues (2024), which also examined IPU duration but in Italian broadcast speech. Savino reports mean IPU durations of approximately 5 seconds in television newsreading broadcast, a register primarily oriented toward stable information delivery. In the present study, informational-cultural speech showed shorter IPUs on average (M = 3.87 s), while advertising speech exhibited even shorter IPUs (M = 2.44 s). While speculative, this pattern may point to a continuum across broadcast registers, with longer IPUs characterizing more information-dense speech (e.g., news), intermediate durations in informational-cultural content, which is designed to both inform and entertain, and progressively shorter IPUs in advertising, where communicative goals are more strongly oriented toward persuasion and emotional impact in a short time window.

Finally, contrary to expectations, speech rate did not differ between advertising and informational speech. Both registers exhibited comparable syllable rates (ADV: 6.57 syll/s vs INFO 6.49 syll/s) indicating that speech rate was not modulated as a function of persuasive intent. This finding is particularly informative given previous work suggesting that faster speech rates may enhance perceived persuasiveness and speaker competence (Peterson, 1995; Charoenruk & Olson, 2018). Previous work on Italian spoken language reported mixed values regarding average speech rate values (syll/s). Federico et al., (2004) reported a mean of approximately 7.17 syll/s in natural Italian speech, whereas Mairano & Romano (2010) described Italian samples as relatively homogeneous, with speech rates ranging between 5 and 6.5 syll/s. Similarly, Pellegrino and colleagues (2011) reported an average syllabic rate of 6.99 syll/s for Italian. Across these studies, speech rate was calculated as syllables per second over pause-delimited stretches of speech, dividing the number of syllables by the duration of the phonic chain, as in our study. Against this background, the speech rate observed in TV commercials in the present study appears to fall well within the range reported for Italian spoken language, rather than standing out as unusually fast or slowed. Importantly, however,

the goal of the present study was not to establish whether advertising speech is faster or slower in absolute terms, but to assess whether speech rate differentiates persuasive from non-persuasive broadcast speech. The absence of differences between advertising and informational speech suggests that, in Italian television advertising, persuasion does not rely on increased speech tempo.

In summary, persuasive speech in Italian television advertising appears to rely on acoustic cues that are also known to encode emotional states, most notably enhanced pitch variability, together with shorter speech units between pauses, pointing to a more expressive and temporally segmented prosodic organization. These features are consistent with previous evidence showing that pitch variation plays a key role in attention capture, emotional engagement, and emphasis during speech processing (e.g., Lieberman & Pierrehumbert, 1984; Hill & Miller 2010). By contrast, speech rate did not differ from that of non-persuasive informational speech and remained within the typical range reported for Italian. Taken together, these results suggest that, in Italian broadcast advertising, persuasion is not achieved by speaking faster, but by speaking more melodically and by structuring speech into shorter, strategically timed units that support cognitive and emotional impact.

This study has some limitations. First, the small sample size limits the generalizability of the findings. In addition, the lack of prior acoustic studies on Italian television advertising restricts direct comparability with existing work. Nonetheless, given the absence of previous acoustic research on Italian television advertising, the present study should be regarded as an initial, exploratory contribution, establishing reference points and motivating more extensive investigations. Future research should examine larger and more diverse samples and extend the analysis to additional prosodic dimensions, including pause structure and intonational contour types, to assess whether specific contour patterns are preferentially used to realize persuasive intent in Italian broadcast speech.

5.5. Conclusions

Despite the availability of publicly accessible corpora of Italian broadcast media, the prosodic realization of persuasive speech in Italian television advertising remains largely unexplored. Our findings indicate that persuasive commercial speech differs from non-persuasive in emotional prosodic cues and in shorter speech units, but not in speed, pointing to a persuasion style that prioritizes emotional salience and temporal efficiency over faster delivery.

CHAPTER 6 | General Discussion

The overall aim of this thesis was to demonstrate that prosody underlies human communication across development, typicality, and communicative contexts, and can be studied as an observable and quantifiable aspect of speech. Across four empirical studies, the thesis examined prosodic processing and production from prenatal auditory sensitivity through early vocal development, as well as prosodic strategies in atypical and persuasive speech. By adopting an integrative approach that combines acoustic-phonetic measures with phonological representations, the thesis provides converging evidence that prosody functions as a central structural component of human communication across the lifespan, rather than as a peripheral aspect of speech.

6.1. Overview of the main findings

Chapter 2 provided preliminary evidence that auditory sensitivity to suprasegmental features is already present in utero. Exposure to structured acoustic stimuli elicited changes in fetal physiological indices, including fetal heart rate and umbilical artery pulsatility index. Importantly, the musical stimuli consisted of contemporary music segments differing in beats per minute (BPM), thereby varying in rhythmic structure and temporal organization. The observed decrease in fetal heart rate (FHR) may be indicative of an orienting response, reflecting not only attentional engagement with the external auditory input but also an early capacity for auditory discrimination, supporting the view that sensitivity to complex acoustic patterns begins prenatally. Concurrently, the increase in umbilical artery pulsatility index (UA-PI) suggests a modulation of fetal blood flow potentially linked to fetal well-being, possibly facilitating enhanced sensory processing. Together, these findings indicate that rhythmic and prosodic properties of sound are not only detected before birth but could modulate cardiovascula responses even in the prenatal period. In this study suprasegmental perception,

operationalized in terms of rhythmic structure, was assessed though physiological reactions reflecting perceptual processing, namely changes in fetal heart rate and umbilical artery pulsatility. Given the shared temporal and rhythmic foundations of music and speech prosody, these results support the notion that prenatal exposure to structured acoustic input may contribute to the early shaping of prosodic sensitivity, laying also the groundwork for later speech and language processing.

Chapter 3 demonstrated that prosodic differentiation is already present in pre-lexical vocalizations, well before the stabilization of words, by tracking the longitudinal development of pitch-related features in Italian infants between 4 and 16 months of age. Analyses of F0 mean and F0 range revealed systematic developmental patterns, alongside substantial inter-individual variability. Across the age range considered, mean pitch level showed a slight overall increase over time, consistent with the growing role of vocalizations in social interaction and communicative engagement with the caregiver. In contrast, pitch variation followed a non-linear trajectory: infants initially exhibited wider pitch excursions during the earliest months, followed by a gradual reduction after approximately the seventh month, suggesting increasing control over vocal production and a progressive shift toward more speech-like prosodic modulation. Importantly, distinct types of pre-lexical vocalizations were associated with different acoustic profiles, with production types linked to clearer communicative functions. Together, these findings indicate that early vocal behavior is not random but systematically organized along prosodic dimensions, while also highlighting the need to account for individual developmental trajectories when investigating the emergence of prosody in infancy.

Chapter 4 showed that children with Autism Spectrum Disorder (ASD) exhibit different prosodic strategies when producing contrastive focus in European Portuguese. When errors occurred, typically developing children tended to emphasize a single non-target constituent,

whereas children with ASD more frequently distributed prominence across multiple constituents within the utterance, often spreading emphasis over three elements rather than one. In addition, productions by children with ASD were characterized by greater variability in prosodic strategies, including less consistent use of pitch accent types and increased instability in the marking of phrase boundaries through boundary tones. Taken together, these patterns suggest that the cumulative variability observed in ASD reduces the perceptual salience of contrastive focus for the listener, making the intended contrast less clearly identifiable. From a pragmatic perspective, this is particularly relevant, as contrastive focus serves to correct and specify meaning; difficulties in producing it appropriately may therefore reflect challenges in expressing communicative intent or establishing conversational boundaries, which appear to be especially demanding for individuals with ASD. Critically, these results could only be fully interpreted through an integrated phonetic-phonological approach. Acoustic measures alone would have obscured the structural nature of the prosodic deviations, while phonological analysis without acoustic grounding would have lacked empirical precision.

Chapter 5 examined prosodic features of persuasive speech in Italian television advertising using a corpus-based approach, comparing advertising clips with non-persuasive broadcast speech. Results showed that advertising speech was characterized by substantially greater pitch variability and shorter interpausal units, while no differences emerged in speech rate between the two conditions. These findings indicate that persuasive communication in Italian television does not rely on faster delivery, but rather on a more melodically modulated and temporally segmented prosodic organization. Increased pitch variation, a well-established acoustic correlate of emotional expression, appears to function as a key cue for attention capture and emotional impact, whereas shorter speech units suggest a strategy of tighter temporal structuring within constrained communicative windows. Importantly, this study represents the first acoustic investigation of Italian persuasive broadcast speech and demonstrates that

variation in pitch and temporal chunking, rather than speech tempo, define the prosodic profile of persuasion. By adopting a corpus-based design, the study captures patterns of prosodic behavior as they naturally occur in real communicative contexts, strengthening the link between prosody, persuasion, and observable communicative behavior.

6.2. Broader implications and future directions

Across the lifespan and across communicative contexts, prosody emerges as a foundational mechanism for communication. Prenatally, its cues are already perceivable and reflected in physiological responses; from infancy through childhood, it structures interaction and supports the gradual emergence of pragmatic competence; and in adulthood, it enables persuasion, emotional expression, influence of others' emotional and attentional states.

The findings of this thesis also have implications for applied domains. In clinical contexts, the demonstrated relevance of prosody for social interaction and pragmatic competence, together with its measurability through acoustic indices, opens the possibility of developing objective biomarkers for communicative disorders. Acoustic measures of prosody may complement traditional clinical assessments by providing reproducible, detailed indicators of communicative behavior, with potential benefits for early identification, monitoring of developmental trajectories, and outcome evaluation. This is particularly relevant given the limited availability of prosody-focused intervention programmes, for instance in Autism Spectrum Disorder (Filipe, 2014), despite clear evidence that prosodic deficits affect every day social functioning, including relationship building and maintenance, educational outcomes, and professional life (Van Bourgondien & Woods, 1992).

Beyond clinical settings, these findings are relevant for speech technology and natural language processing, where the inclusion of prosodic features can enhance the naturalness, intelligibility, and communicative effectiveness of automatic systems, ranging from

conversational agents to language-learning applications (Nouza, 1999; Delmonte, 2000). Similarly, in marketing and persuasive communication, identifying the acoustic correlates of effective persuasion provides a basis for evidence-based training practices. Rather than relying on intuition or stylistic conventions, communicative effectiveness may be systematically taught and optimized, with applications in advertising, sales, and public speaking.

Despite its contributions, this thesis has several limitations. Sample sizes were necessarily limited in some studies, particularly in the prenatal investigation and in the analysis of advertising speech, reflecting both methodological constraints and the exploratory nature of these datasets. Moreover, the thesis adopts a multi-study model spanning highly diverse populations, methodologies, and communicative contexts. While this span of empirical domains may be a strength of this dissertation, it also limits direct comparability across studies, as not all measures and designs are fully overlapping, and integration necessarily occurs at a conceptual rather than experimental level.

A particularly promising direction concerns the development of automatic classification systems capable of identifying prosodic profiles associated with specific developmental stages or communicative conditions, potentially capturing subtle acoustic patterns that are difficult to detect perceptually or visually on the spectrogram. Such advances would further strengthen the role of prosody as a key interface between linguistic structure, cognition, and social communication across the lifespan and across different contexts.

6.3. General conclusion

This thesis highlights the central role of the prosody in human communication, showing that speakers convey meaning not only through *what* they say, but also through *how* they say it. Prosodic modulation allows communicative messages to be enriched with structural, pragmatic, and affective information that goes beyond lexical content. Sensitivity to these vocal

cues emerges early, already at the prenatal auditory level, and continues to develop across the lifespan. From early childhood onward, prosody supports interaction and social communication, contributing to the successful establishment and maintenance of interpersonal relationships. Importantly, deficits in prosodic abilities are associated with pragmatic and communicative difficulties, which can in turn affect personal, social, and professional outcomes. Conversely, when prosodic skills are effectively deployed, they can enhance communicative impact and influence listeners' emotional and attentional states. The findings of this thesis show that these processes are observable and measurable through objective and quantifiable indices. As such, prosody can be systematically studied and applied across multiple domains, including clinical contexts, behavioral sciences, and technological applications.

General Bibliography

Albano Leoni A., F., Cutugno, F., Savy, R., et al. (2006). *Corpora e lessici dell'italiano parlato e scritto (CLIPS)* [Electronic corpus]. Università degli Studi di Napoli Federico II. http://www.clips.unina.it (accessed 18/01/2026)

Amano, S., Nakatani, T., & Kondo, T. (2006). Fundamental frequency of infants' and parents' utterances in longitudinal recordings. *Journal of the Acoustical Society of America, 119*(3), 1636.

American Psychiatric Association. 2013. *Diagnostic and statistical manual of mental disorders.* (5th ed.) Author.

Apple, W., Streeter, L. A., & Krauss, R. M. (1979). Effects of pitch and speech rate on personal attributions. *Journal of Personality and Social Psychology, 37*(5), 715–727.

Applebaum, L., Coppola, M., & Goldin-Meadow, S. (2014). Prosody in a communication system developed without a language model. *Sign Language & Linguistics, 17*(2), 181–212. https://doi.org/10.1075/sll.17.2.02app.

Aristotle. (1991). *Rhetoric: A theory of civic discourse* (G. A. Kennedy, Trans.). Oxford University Press.

Asghari, M., Khodadadi, A., Eslami, Z., Hosseini, S. M., & Khazaie, H. (2021). Distinctive prosodic features of people with autism spectrum disorder: A systematic review and meta-analysis study. *Scientific Reports, 11*, 23087. https://doi.org/10.1038/s41598-021-02487-6

Astruc, L., Payne, E., Post, B., del Mar Vanrell, M., II, & Prieto, P. (2013). Tonal targets in early child English, Spanish, and Catalan. *Language and Speech, 56*(2), 229–253. https://doi.org/10.1177/0023830912460494.

Audacity Team. (2023). *Audacity* (Version 3.2.5) [Software]. https://www.audacityteam.org/

Aureli, T., Spinelli, M., Fasolo, M., Garito, M. C., Perucchini, P., & D'Odorico, L. (2017). The pointing–vocal coupling progression in the first half of the second year of life. *Infancy, 22*(6), 801–818.

Baltaxe, C. (1984). Use of contrastive stress in normal, aphasic, and autistic children. Journal of Speech and Hearing Research, 27,97-105.

Baltaxe, C. A., & Guthrie, D. (1987). The use of primary sentence stress by normal, aphasic, and autistic children. *Journal of Autism and Developmental Disorders, 17*(2), 255–271.

Bambini, V., Arcara, G., Bechi, M., Buonocore, M., Cavallaro, R., & Bosia, M. (2016). The communicative impairment as a core feature of schizophrenia: Frequency of pragmatic deficit, cognitive substrates, and relation with quality of life. *Comprehensive Psychiatry, 71*, 106–120. https://doi.org/10.1016/j.comppsych.2016.08.012.

Bänziger, T., Patel, S., & Scherer, K. R. (2015). The role of the voice in emotional communication. In *The Oxford Handbook of Voice Perception* (pp. 185–201). Oxford University Press.

Baron-Cohen, S. (1995). *Mindblindness: An Essay on Autism and Theory of Mind* (MIT Press, Cambridge), 171 pp.

Baron-Cohen, S., and Staunton, R. (1994). Do children with autism acquire the phonology of their peers? An examination of group identification through the window of bilingualism. *First Language*. 14, 241–248.

Bates, E., & MacWhinney, B. (1979). *A functionalist approach to the acquisition of grammar*. In E. Ochs & B. Schieffelin (Eds.), *Developmental pragmatics*. Academic Press.

Beckman, M. E., & Pierrehumbert, J. B. (1986). Intonational structure in Japanese and

English. *Phonology Yearbook, 3*, 255–309.

Behrens, H., & Gut, H. (2005). The relationship between prosodic and syntactic organization in early multiword speech. *Journal of Child Language, 32*(1), 1–34.

Belyk, M., & Brown, S. (2014). *Perception of affective and linguistic prosody: An ALE meta-analysis of neuroimaging studies*. Social Cognitive and Affective Neuroscience, 9(9), 1395–1403. https://doi.org/10.1093/scan/nst124

Bennett, S. (1983). A 3-year longitudinal study of school-aged children's fundamental frequencies. *Journal of Speech & Hearing Research, 26*(1), 137–141.

Berger, S., Niebuhr, O., & Peters, B. (2017). Winning over an audience: A perception-based analysis of prosodic features of charismatic speech. In *Proceedings of the 43rd Deutsche Jahrestagung für Akustik (DAGA)* (pp. 1454–1457).

Bertinetto, P. M. (1981). *Strutture prosodiche dell'italiano*. Accademia della Crusca.

Blanc, J. M., & Dominey, P. F. (2003). Identification of prosodic attitudes by a temporal recurrent network. *Cognitive Brain Research, 17*(3), 693–699. https://doi.org/10.1016/S0926-6410(03)00195-2.

Boersma, P., & Weenink, D. (2009). *Praat: Doing phonetics by computer* (Version 6.4.25) [Computer software]. http://www.praat.org.

Bolz, N., Kalache, K. D., Proquitte, H., Slowinski, T., Hartung, J. P., Henrich, W., & Bamberg, C. (2013). Value of Doppler sonography near term: can umbilical and uterine artery indices in low-risk pregnancies predict perinatal outcome? *Journal of perinatal medicine*, *41*(2), 165–170. https://doi.org/10.1515/jpm-2012-0042

Bonanno M, Fabiani M, Cavallini F, & Iacomini S. (2025). Prosodic Perception and Production in Autism: A Scoping Review. *Journal of Clinical and Developmental Psychology*, *2025*(1), 116–135. https://doi.org/10.13129/2612-4033/0110-4799.

Boraston, Z., & Blakemore, S.-J. (2007). The application of eye-tracking technology in the

study of autism. *The Journal of Physiology, 581*(3), 893–898. https://doi.org/10.1113/jphysiol.2007.133587.

Brisson, J., Martel, K., Serres, J., Sirois, S., & Adrien, J. L. (2014). Acoustic analysis of oral productions of infants later diagnosed with autism and their mother. *Infant Mental Health Journal, 35*(3), 285–295. https://doi.org/10.1002/imhj.21442.

Brosnan, M., Ashwin, C., & Lewton, M. (2017). Brief report: Intuitive and reflective reasoning in autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 47(8), 2595–2601. https://doi.org/10.1007/s10803-017-3131-3.

Buhmann, J., Caspers, J., van Heuven, V. J., Hoekstra, H., Martens, J.-P., & Swerts, M. (2002). Annotation of prominent words, prosodic boundaries and segmental lengthening by non-expert transcribers in the spoken Dutch corpus. In *Proceedings of the LREC 2002* (pp. 779–785). ELRA.

Camaioni, L. (1992). Comunicazione e linguaggio. In L. Camaioni & V. Volterra (Eds.), *Lo sviluppo della comunicazione e del linguaggio* (pp. 45–78). Il Mulino.

Carvalho, A., Dautriche, I., Millotte, S., & Christophe, A. (2018). Early perception of phrasal prosody and its role in syntactic and lexical acquisition. In P. Prieto & N. Esteve-Gibert (Eds.), *The development of prosody in first language acquisition* (pp. 18–30). John Benjamins. https://doi.org/10.1075/tilar.23.02car.

Carvalho, E., Rosado Miranda Justo, J. M., Gratier, M., & Rodrigues, H. (2019). The Impact of Maternal Voice on the Fetus: A Systematic Review. *Current Women's Health Reviews*, *15*(3),196-206.Article 3. https://doi.org/10.2174/1573404814666181026094419.

Castro, L., Serridge, B., de Moraes, J. A., Freitas, M., & Botinis, A. (2010). The prosody of the TV news speaking style in Brazilian Portuguese. In *Proceedings of ExLing 2010*

(pp. 17–20)

Chafe, W. (1976). Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In C. N. Li (Ed.), *Subject and topic* (pp. 25–56). Academic Press.

Charoenruk, N., & Olson, K. (2018). Do listeners perceive interviewers' attributes from their voices, and do perceptions differ by question type? *Field Methods, 30*(4), 312–328. https://doi.org/10.1177/1525822X18781957.

Chen, A. (2011). Tuning information packaging: Intonational realization of topic and focus in child Dutch. *Journal of Child Language*, *38*, 1055–1083. https://doi.org/10.1017/S0305000910000541.

Chen, A., & Fikkert, P. (2007). Intonation of early two-word utterances in Dutch. In *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 315–320). Saarbrücken, Germany.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. Harper & Row.

Cotter, C. (1993). Prosodic aspects of broadcast news register. In *Proceedings of the 19th Annual Meeting of the Berkeley Linguistics Society* (pp. 90–100).

Couture, S. M., Penn, D. L., & Roberts, D. L. (2006). The functional significance of social cognition in schizophrenia: A review. *Schizophrenia Bulletin, 32*(1), 44–63. https://doi.org/10.1093/schbul/sbl029.

Crary, M. A., & Tallman, V. L. (1993). Production of linguistic prosody by normal and speech-disordered children. *Journal of Communication Disorders, 26*(4), 245–262. https://doi.org/10.1016/0021-9924(93)90019-7.

Cruttenden, A. (1986). *Intonation*. Cambridge University Press.

Cruttenden, A. (1997). *Intonation*. Cambridge University Press.

Crystal, D. (1969). *Prosodic systems and intonation in English*. Cambridge University Press.

Crystal, D. (1978). *The analysis of intonation in young children*. Edward Arnold.

Crystal, D. (1981). *Clinical linguistics*. Springer.

D'Odorico, L. (1984). Non-segmental features in prelinguistic communications: An analysis of some types of infant cry and non-cry vocalizations. *Journal of Child Language, 11*(1), 17–27.

D'Odorico, L., & Franco, F. (1991). Selective production of vocalization types in different communication contexts. *Journal of Child Language, 18*(3), 475–499.

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. *Enfance*, (3), 317–327.

De Boer, J. N., Voppel, A. E., Brederoo, S. G., Schnack, H. G., Truong, K. P., Wijnen, F. N. K., & Sommer, I. E. C. (2023). Acoustic speech markers for schizophrenia-spectrum disorders: A diagnostic and symptom-recognition tool. *Psychological Medicine, 53*(4), 1302–1312. https://doi.org/10.1017/S0033291721002804

De Carvalho, A., He, A. X., Lidz, J., & Christophe, A. (2019). Prosody and function words cue the acquisition of word meanings in 18-month-old infants. *Psychological Science, 30*(3), 319–332.

DeCasper, A. J., & Fifer, W. P. (1980). Of human bonding: Newborns prefer their mothers' voices. *Science, 208*(4448), 1174–1176. https://doi.org/10.1126/science.7375928.

Delmonte, R. (2000). SLIM prosodic automatic tools for self-learning instruction. *Speech Communication, 30*(2–3), 145–166.

DePape, A. M. R., Chen, A., Hall, G. B., & Trainor, L. J. (2012). Use of prosody and information structure in high functioning adults with autism in relation to language ability. *Frontiers in Psychology, 3*, Article 72.

https://doi.org/10.3389/fpsyg.2012.00072.

Diehl, J. J., & Paul, R. (2013). Acoustic and perceptual measurements of prosody production on the Profiling Elements of Prosodic Systems in children with autism spectrum disorders. *Applied Psycholinguistics, 34*(1), 135–161. https://doi.org/10.1017/S0142716411000646.

Diehl, J. J., Watson, D., Bennetto, L., McDonough, J., & Gunlogson, C. (2009). An acoustic analysis of prosody in high-functioning autism. *Applied Psycholinguistics, 30*(3), 385–404. https://doi.org/10.1017/S0142716409090201.

Drimalla, H., Scheffer, T., Landwehr, N., Baskow, I., Roepke, S., Behnia, B., & Dziobek, I. (2020). Towards the automatic detection of social biomarkers in autism spectrum disorder: introducing the simulated interaction task (SIT). *NPJ digital medicine*, *3*, 25. https://doi.org/10.1038/s41746-020-0227-5.

Drukker, L., Staines-Urias, E., Villar, J., Barros, F. C., Carvalho, M., Munim, S., McGready, R., Nosten, F., Berkley, J. A., Norris, S. A., Uauy, R., Kennedy, S. H., & Papageorghiou, A. T. (2020). International gestational age-specific centiles for umbilical artery Doppler indices: a longitudinal prospective cohort study of the INTERGROWTH-21st Project. *American journal of obstetrics and gynecology*, *222*(6), 602.e1–602.e15. https://doi.org/10.1016/j.ajog.2020.01.012.

Dunn, D. M., & Dunn, L. M. (2007). *Peabody Picture Vocabulary Test* (4th ed.). Pearson.

Dynel, M. (2019). Ironic intentions in action and interaction. *Language Sciences, 75*, 1–14. https://doi.org/10.1016/j.langsci.2019.06.005.

Esposito, G., Pini, N., Tagliaferri, S., Campanile, M., Zullo, F., Magenes, G., Maruotti, G. M., & Signorini, M. G. (2021). An integrated approach based on advanced CTG parameters and Doppler measurements for late growth restriction management. *BMC pregnancy and childbirth*, *21*(1), 775. https://doi.org/10.1186/s12884-021-04235-0.

Esteve-Gibert, N., & Prieto, P. (2013). Prosody signals the emergence of intentional communication in the first year of life. *Journal of Child Language, 40*(5), 919–944. https://doi.org/10.1017/S0305000912000359.

Esteve-Gibert, N., & Prieto, P. (2018). *The development of prosody in first language acquisition*. Cambridge University Press.

Fairbanks, G. (1942). An acoustical study of the pitch of infant hunger wails. *Child Development, 13*, 227–232.

Fasolo, M., D'Odorico, L., Costantini, A., & Cassibba, R. (2010). The influence of biological, social, and developmental factors on language acquisition in pre-term born children. *International Journal of Speech-Language Pathology, 12*(6), 461–471.

Fasolo, M., Majorano, M., & D'Odorico, L. (2008). Babbling and first words in children with slow expressive development. *Clinical Linguistics & Phonetics, 22*(2), 83–94.

Fay, W., and Schuler, A. L. (1980). *Emerging Language in Autistic Children*. Baltimore: University Park Press.

Federico, A., Mori, L., & Paoloni, A. (2004). La velocità di articolazione come parametro identificativo: Potenzialità e limiti per la caratterizzazione del parlante. In *Atti del Convegno AISV* (pp. 869–876).

Feinberg, D. R., Jones, B. C., Little, A. C., Burt, D. M., & Perrett, D. I. (2005). Manipulations of fundamental and formant frequencies influence the attractiveness of human male voices. *Animal Behaviour, 69*(3), 561–568. https://doi.org/10.1016/j.anbehav.2004.06.012.

Ferdousi, M. A., Sharif, M. M., Mohiuddin, A. S., & Shegufta, F. (2013). Normal value of pulsatility index of umbilical artery in second and third trimester of pregnancy. *Bangladesh Med. Res. Counc. Bull., 39*(1), 42–44. https://doi.org/10.3329/bmrcb.v39i1.15809.

Filipe, M. G., Frota, S., & Vicente, S. G. (2018). Executive functions and prosodic abilities in children with high-functioning autism. *Frontiers in psychology*, *9*, 359. https://doi.org/10.3389/fpsyg.2018.00359.

Filipe, M. G., Frota, S., Castro, S. L., & Vicente, S. G. (2014). Atypical prosody in Asperger syndrome: Perceptual and acoustic measurements. *Journal of Autism and Developmental Disorders, 44*(8), 1972–1981. https://doi.org/10.1007/s10803-014-2073-2.

Filipe, M. G., Peppé, S., Frota, S., & Vicente, S. G. (2017). Prosodic development in European Portuguese from childhood to adulthood. *Applied Psycholinguistics, 38*(5), 1045–1070. https://doi.org/10.1017/S0142716417000030.

Filipe, M., Frota, S., Villagomez, A., & Vicente, S. G. (2016). Prosody in Portuguese Children with HighFunctioning Autism. In *Intonational grammar in Ibero-Romance: Approaches across linguistic subfields* (pp. 277-294). John Benjamins Publishing Company. https://doi.org/10.1075/ihll.6.13fil.

Fine, J., Bartolucci, G., Ginsberg, G., & Szatmari, P. (1991). The use of intonation to communicate in pervasive developmental disorders. *Journal of Child Psychology and Psychiatry*, 32, 771-782. https://doi.org/10.1111/j.1469-7610.1991.tb01901.x.

Fisher, C., & Tokura, H. (1996). Prosody in speech to infants: Direct and indirect acoustic cues to syntactic structure. In *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 343–363).

Flax, J., Lahey, M., Harris, K., & Boothroyd, A. (1991). Relations between prosodic variables and communicative functions. *Journal of Child Language, 18*(1), 3–19. https://doi.org/10.1017/S030500090001326X.

Fletcher, W. (2008). *Powers of persuasion: The inside story of British advertising, 1951–2000*. Oxford University Press

Fosnot, S. M., & Jun, S. (1999, August). Prosodic characteristics in children with stuttering or autism during reading and imitation. In *Proceedings of the 14th international congress of phonetic sciences* (pp. 1925-1928).

Fox, D. B. (1990). An analysis of the pitch characteristics of infant vocalizations. *Psychomusicology, 9*(1), 21–30. https://doi.org/10.1037/h0094161.

Frith, U. (1991). *Autism and Asperger Syndrome*. Cambridge: Cambridge University Press.

Frota, S. (2000). *Prosody and focus in European Portuguese: Phonological phrasing and intonation*. New York, NY: Garland Publishing.

Frota, S. (2002). *The prosody of focus: A case study with cross-linguistic implications*. In B. Bel & I. Marlien (Eds.), *Speech prosody 2002: Proceedings of the 1st International Conference on Speech Prosody* (pp. 277–280). Laboratoire Parole et Langage. https://doi.org/10.21437/SpeechProsody.2002-64.

Frota, S. (2014). The intonational phonology of European Portuguese. In S.-A. Jun (Ed.), *Prosodic typology II: The phonology of intonation and phrasing* (pp. 6–42). Oxford University Press.

Frota, S., & Butler, J. (2018). Early development of intonation: Perception and production. In P. Prieto & N. Esteve-Gibert (Eds.), *The development of prosody in first language acquisition* (pp. 146–160). John Benjamins.

Frota, S., & Prieto, P. (2015a). *Intonation in Romance*. Oxford University Press.

Frota, S., & Prieto, P. (Eds.). (2015). *Intonation in Romance*. Oxford University Press.

Frota, S., Cruz, M., Matos, N., & Vigário, M. (2016). Early prosodic development. In S. Frota & P. Prieto (Eds.), *Intonational grammar in Ibero-Romance* (pp. 295–324). John Benjamins.

Frota, S., Oliveira, P., Cruz, M., & Vigário, M. (2015). *P-ToBI: Tools for the transcription of*

Portuguese prosody. Laboratório de Fonética, CLUL/FLUL. http://labfon.letras.ulisboa.pt/InAPoP/P-ToBI/.

Fusaroli, R., Lambrechts, A., Bang, D., Bowler, D. M., & Gaigg, S. B. (2017). "Is voice a marker for Autism spectrum disorder? A systematic review and meta-analysis." *Autism Research*, *10*(3), 384–407. https://doi.org/10.1002/aur.1678.

García González, J., Ventura Miranda, M. I., Manchon García, F., Pallarés Ruiz, T. I., Marin Gascón, M. L., Requena Mullor, M., Alarcón Rodriguez, R., & Parron Carreño, T. (2017). Effects of prenatal music stimulation on fetal cardiac state, newborn anthropometric measurements and vital signs of pregnant women: A randomized controlled trial. *Complementary therapies in clinical practice*, *27*, 61–67. https://doi.org/10.1016/j.ctcp.2017.03.004

Gebuza, G., Zaleska, M., Kaźmierczak, M., Mieczkowska, E., & Gierszewska, M. (2018). The effect of music on the cardiac activity of a fetus in a cardiotocographic examination. *Advances in clinical and experimental medicine: official organ Wroclaw Medical University*, *27*(5), 615–621. https://doi.org/10.17219/acem/68693

Gerhardt, K. J., Otto, R., Abrams, R. M., Colle, J. J., Burchfield, D. J., & Peters, A. J. (1992). Cochlear microphonics recorded from fetal and newborn sheep. *American Journal of Otolaryngology, 13*(4), 226–233. https://doi.org/10.1016/0196-0709(92)90026-P.

Gerstenberg, A., Fuchs, S., Kairet, J. M., Frankenberg, C., & Schröder, J. (2018). A cross-linguistic, longitudinal case study of pauses and interpausal units in spontaneous speech corpora of older speakers of German and French [Paper presentation]. *Speech Prosody 2018*. https://doi.org/10.21437/SpeechProsody.2018-43.

Gervain, J. (2015). Plasticity in early language acquisition. *Current Opinion in Neurobiology, 35*, 13–20.

Gervain, J., Christophe, A., & Mazuka, R. (2020). Prosodic bootstrapping. In C.

Gussenhoven & A. Chen (Eds.), *The Oxford handbook of language prosody*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198832232.013.36

Gili Fivela, B., Avesani, C., Barone, M., Bocci, G., Crocco, C., D'Imperio, M., Giordano, R., Marotta, G., Savino, M. Sorianello, P. (2015). Intonational phonology of the regional varieties of Italian. In S. Frota & P. Prieto (Eds.), *Intonation in Romance* (pp. 140–197). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199685332.001.0001.

Gleitman, L., Gleitman, H., Landau, B., Wanner, E., & Newmeyer, F. (1988). *Linguistics: The Cambridge survey. Vol. 3*. Cambridge University Press.

Goldstein, M. H., Schwade, J. A., & Bornstein, M. H. (2009). The value of vocalizing. *Child Development, 80*(3), 636–644. https://doi.org/10.1111/j.1467-8624.2009.01287.x.

Gomes, D. Â. D. V., Dornelas, R., Souza, L. M. S. D., Strauss, Y. C. D. S., Silva, E. M. D., & Celeste, L. C. (2023). Acoustics of persuasion: A systematic review of prosodic features on the communicative attitudes of confidence. *DELTA: Documentação de Estudos em Linguística Teórica e Aplicada, 39*(3), e202339345326. https://doi.org/10.1590/1678-460X202339345326.

Graham, F. K. (1992). Attention: The heartbeat, the blink, and the brain. In B. A. Campbell, H. Hayne, & R. Richardson (Eds.), *Attention and information processing in infants and adults: Perspectives from human and animal research* (pp. 3–29). Lawrence Erlbaum Associates, Inc.

Granier-Deferre, C., Ribeiro, A., Jacquet, A. Y., & Bassereau, S. (2011). Near-term fetuses process temporal features of speech. *Developmental science*, *14*(2), 336–352. https://doi.org/10.1111/j.1467-7687.2010.00978.x

Gratier, M., & Devouche, E. (2011). Imitation and repetition of prosodic contour in vocal interaction at 3 months. *Developmental Psychology, 47*(1), 67–76.

Green, H., & Tobin, Y. (2009). Prosodic Analysis is Difficult... But Worth It: A Study in High Functioning Autism. *International Journal of Speech-Language Pathology*, *11*(4), 308-315. https://doi.org/10.1080/17549500903003060.

Grice, M., & Baumann, S. (2007). An introduction to intonation: Functions and models. In J. Trouvain & U. Gut (Eds.), *Non-native prosody: Phonetic description and teaching practice* (pp. 25–51). De Gruyter Mouton.

Grice, M., Wehrle, S., Krüger, M., Spaniol, M., Cangemi, F., & Vogeley, K. (2023). Linguistic prosody in autism spectrum disorder—An overview. In *Language and Linguistics Compass* (Vol. 17, Issue 5). John Wiley and Sons Inc. https://doi.org/10.1111/lnc3.12498.

Grichkovtsova, I., Morel, M., & Lacheret, A. (2009). *Perception of affective prosody in natural and synthesized speech: Which methodological approach?* In S. Hancil (Ed.), *The role of prosody in affective speech* (pp. 371-390). Peter Lang.

Groome, L. J., Mooney, D. M., Holland, S. B., Bentz, L. S., Atterbury, J. L., & Dykman, R. A. (1997). The heart rate deceleratory response in low-risk human fetuses: effect of stimulus intensity on response topography. *Developmental psychobiology*, *30*(2), 103–113. https://doi.org/10.1002/(sici)1098-2302(199703)30:2<103::aid-dev2>3.0.co;2-u

Grossman, R. B., Bemis, R. H., Skwerer, D. P., & Tager-Flusberg, H. (2010). Lexical and affective prosody in children with high-functioning autism.

Gupta, Shikha & Gupta, Varad. (2018). Effect of Classical Music on Foetoplacental Circulation in Pregnancy Induced Hypertension Using Color Doppler Indices. *International Journal of Physiology*. https://doi.org/10.5958/2320-

608X.2018.00103.8.

Gussenhoven, C. (2004). The Phonology of Tone and Intonation. Cambridge: Cambridge University Press.

Guyer, J. J., Fabrigar, L. R., & Vaughan-Johnston, T. I. (2019). Speech rate, intonation, and pitch: Investigating the bias and cue effects of vocal confidence on persuasion. *Personality and Social Psychology Bulletin, 45*(3), 389–403. https://doi.org/10.1177/0146167218787805

Guyer, J. J., Fabrigar, L. R., & Vaughan-Johnston, T. I. (2019). Speech rate, intonation, and pitch: Investigating the bias and cue effects of vocal confidence on persuasion. *Personality and Social Psychology Bulletin, 45*(3), 389–403. https://doi.org/10.1177/0146167218787805.

Halliday, M. K. K. (1967). Intonation and Grammar in British English. The Hague: Mouton.

Hill, K. T., & Miller, L. M. (2010). Auditory attentional control and selection during cocktail party listening. *Cerebral Cortex, 20*(3), 583–590. https://doi.org/10.1093/cercor/bhp124.

Hirschberg, J., Benus, S., Gravano, A., & Levitan, R. (2020). Prosody in discourse and speaker state. In C. Gussenhoven & A. Chen (Eds.), *The Oxford handbook of language prosody* (pp. 468–476). Oxford University Press.

Hirst, D., & Looze, C. D. (2021)**.** Fundamental frequency and pitch. In R.-A. Knight & J. Setter (Eds.), *The Cambridge handbook of phonetics* (pp. 336–361). Cambridge University Press.

Hsu, H.-C., Fogel, A., & Cooper, R. B. (2000). Infant vocal development during the first 6 months. *Infant and Child Development, 9*(1), 1–16. https://doi.org/10.1002/(SICI)1522-7219(200003)9:1<1::AID-ICD210>3.0.CO;2-V.

Hubbard, D. J., Faso, D. J., Assmann, P. F., & Sasson, N. J. (2017). Production and

perception of emotional prosody by adults with autism spectrum disorder. *Autism Research, 10*(12), 1991–2001. https://doi.org/10.1002/aur.1847.

Iyer, S. N., & Oller, D. K. (2008). Fundamental frequency development in typically developing infants and infants with severe-to-profound hearing loss. *Clinical Linguistics & Phonetics, 22*(12), 917–936. https://doi.org/10.1080/02699200802316776.

Jamovi. (2023). *Jamovi (Version 2.3.28.0)* [Computer software]. https://www.jamovi.org/.

Järvinen-Pasley, A., Peppé, S., King-Smith, G., & Heaton, P. (2008). The relationship between form and function level receptive prosodic abilities in autism. *Journal of autism and developmental disorders*, *38*(7), 1328-1340.

Jeličić, L., Janković-Raznatović, S., Sovilj, M., & Subotić, M. (2023). Change in middle cerebral artery pulsatility index over time due to fetal response to defined sound stimulation: Human fetal responses to airborne sound stimulation. *International Journal of Prenatal & Life Sciences*, *1*(1), 1-7. https://doi.org/10.24946/IJPLS/202312205.

Judkins, L., Alazard-Guiu, C., & Astésano, C. (2022). How do we chunk and pause in non-native vs. native speech? In *Speech Prosody 2022* (pp. 792–796). https://doi.org/10.21437/SpeechProsody.2022-161.

Jun, S.-A. (Ed.). (2005a). *Prosodic typology: The phonology of intonation and phrasing*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199249633.001.0001.

Kahane, J. C., & Kahn, A. R. (1984). Weight measurements of infant and adult intrinsic laryngeal muscles. *Folia Phoniatrica et Logopaedica, 36*(3), 129–133.

Kanner, B. (1999). *The 100 best TV commercials… and why they worked*. Times Business.

Kanner, B. (1999). *The 100 best TV commercials… and why they worked*. Times Business.

Kanner, L. (1943). Autistic disturbances of affective contact. Nerv. Child 2, 217–250.

Kawakami, K., Takai-Kawakami, K., Masuda, N., Suzuki, M., Shimizu, Y., & Yanaihara, T. (2002). Measuring human fetal responses to sounds by umbilical and middle-cerebral artery velocity waveforms: A preliminary study. *Hokkaido University Repository*. https://eprints.lib.hokudai.ac.jp/dspace/bitstream/2115/25352/1/24_P25-33.pdf

Kent, R. D. (1976). Anatomical and neuromuscular maturation of the speech mechanism. *Journal of Speech, Language, and Hearing Research, 19*(3), 421–447. https://doi.org/10.1044/jshr.1903.421.

Kent, R. D., & Murray, A. D. (1982). Acoustic features of infant vocalic utterances. *Journal of the Acoustical Society of America, 72*(2), 353–365.

Kišiček, G. (2018). Persuasive power of prosodic features. *Argumentation and Advocacy*. https://doi.org/10.1080/10511431.2019.1525003.

Kisilevsky, B. S., Pang, L., & Hains, S. M. (2000). Maturation of human fetal responses to airborne sound in low- and high-risk fetuses. *Early human development*, *58*(3), 179–195. https://doi.org/10.1016/s0378-3782(00)00075-x.

Kisilevsky, S., Hains, S. M., Jacquet, A. Y., Granier-Deferre, C., & Lecanuet, J. P. (2004). Maturation of fetal responses to music. *Developmental science*, *7*(5), 550–559. https://doi.org/10.1111/j.1467-7687.2004.00379.x.

Kjelgaard, M. A., & Tager-Flusberg, H. (2001). An investigation of language profiles in autism: Implications for genetic subgroups. Language and Cognitive Processes, 15, 1-22. https://doi.org/10.1080/01690960042000058.

Klofstad, C. A., & Anderson, R. C. (2018). Voice pitch predicts electability but does not signal leadership ability. *Evolution and Human Behavior, 39*(3), 349–354. https://doi.org/10.1016/j.evolhumbehav.2018.01.004.

Ko, E.-S., et al. (2016). Entrainment of prosody in the interaction of mothers with their young children. *Journal of Child Language, 43*(2), 284–309.

Koiso, H., Horiuchi, Y., Tutiya, S., Ichikawa, A., & Den, Y. (1998). An analysis of turn-taking and backchannels based on prosodic and syntactic features in Japanese map task dialogs. *Language and Speech, 41*(3–4), 295–321. https://doi.org/10.1177/002383099804100404.

Krüger, M., Cangemi, F., Vogeley, K., & Grice, M. (2018). Prosodic marking of information status in adults with autism spectrum disorders. In *Proceedings of the 9th International Conference on Speech Prosody* (pp. 182–186).

Kügler, F., & Calhoun, S. (2020). Prosodic encoding of information structure – A typological perspective. In C. Gussenhoven & A. Chen (Eds.), *The Oxford handbook of language prosody* (pp. 454–467). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198832232.013.30.

Kulinski, J., Ofori, E. K., Visotcky, A., Smith, A., Sparapani, R., & Fleg, J. L. (2022). Effects of music on the cardiovascular system. *Trends in cardiovascular medicine*, *32*(6), 390–398. https://doi.org/10.1016/j.tcm.2021.06.004

Kuperberg, G. R. (2010a). Language and schizophrenia. Part 1: An introduction. *Language and Linguistics Compass, 4*, 576–589. https://doi.org/10.1111/j.1749-818X.2010.00216.x.

Ladd, D. R. (2008). *Intonational phonology* (2nd ed.). Cambridge University Press.

Landa, R. (2000). Social language use in Asperger syndrome and high-functioning Autism. In A. Klin, F. R. Volkmar, & S. Sparrow (Eds.), *Asperger syndrome* (pp. 125–158). Guilford Press.

Laufer, M. Z., & Horii, Y. (1977). Fundamental frequency characteristics of infant non-distress vocalization. *Journal of Child Language, 4*(2), 171–184. https://doi.org/10.1017/S0305000900001586.

Laws, G., & Bishop, D. V. M. (2004). Verbal deficits in Down's syndrome and specific language impairment: A comparison. *International Journal of Language & Communication Disorders, 39*, 423–451.

Lecanuet, J. P., Granier-Deferre, C., Jacquet, A. Y., & Busnel, M. C. (1992). Decelerative cardiac responsiveness to acoustical stimulation in the near-term fetus. *The Quarterly journal of experimental psychology. B, Comparative and physiological psychology*, *44*(3-4), 279–303. https://doi.org/10.1080/02724999208250616.

Lecanuet, J. P., Graniere-Deferre, C., Jacquet, A. Y., & DeCasper, A. J. (2000). Fetal discrimination of low-pitched musical notes. *Developmental psychobiology*, *36*(1), 29–39.

Lecanuet, J. P., Graniere-Deferre, C., Jacquet, A. Y., & DeCasper, A. J. (2000). Fetal discrimination of low-pitched musical notes. *Developmental psychobiology*, *36*(1), 29–39.

Lee, S., Potamianos, A., & Narayanan, S. (1999). Acoustics of children's speech. *Journal of the Acoustical Society of America, 105*(3), 1455–1468. https://doi.org/10.1121/1.426686.

Leech, G. N. (1966). *English in advertising*. Longman.

Lehiste, I. (1970). *Suprasegmentals*. MIT Press.

Liberman, M. Y., & Pierrehumbert, J. (1984). Intonational invariance under changes in pitch range and length. In M. Aronoff & R. T. Oehrle (Eds.), *Language sound structure: Studies in phonology presented to Morris Halle* (pp. 157–233). MIT Press.

Lieberman, P. (1967). *Intonation, perception, and language*. MIT Press.

Lord, C., Rutter, M., & Le Couteur, A. (1994). The Autism Diagnostic Interview-Revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders, 24*(5), 659–685.

Lord, C., Rutter, M., DiLavore, P. C., & Risi, S. (2000). *Autism Diagnostic Observation Schedule (ADOS).* Western Psychological Services.

Lord, C., Rutter, M., DiLavore, P. C., Risi, S., Gotham, K., & Bishop, S. L. (2012). *Autism Diagnostic Observation Schedule* (2nd ed.). Western Psychological Services.

Mairano, P., & Romano, A. (2010). Un confronto tra diverse metriche ritmiche usando Correlatore. In *La dimensione temporale del parlato* (Vol. 5, pp. 79–100).

Malamitsi-Puchner, A., Nikolaou, K. E., & Puchner, K. P. (2006). Intrauterine growth restriction, brain-sparing effect, and neurotrophins. *Annals of the New York Academy of Sciences*, *1092*, 293–296. https://doi.org/10.1196/annals.1365.026.

Mampe, B., et al. (2009). Newborns' cry melody is shaped by their native language. *Current Biology, 19*(23), 1994–1997.

Mampe, B., Friederici, A. D., Christophe, A., & Wermke, K. (2009). Newborns' cry melody is shaped by their native language. *Current Biology, 19*(23), 1994–1997. https://doi.org/10.1016/j.cub.2009.09.064.

Marie, C., & Trainor, L. J. (2013). Development of simultaneous pitch encoding: Infants show a high-voice superiority effect. *Cerebral Cortex, 23*, 660–669. https://doi.org/10.1093/cercor/bhs050.

Marsál K. (2009). Obstetric management of intrauterine growth restriction. *Best practice & research. Clinical obstetrics & gynaecology*, *23*(6), 857–870. https://doi.org/10.1016/j.bpobgyn.2009.08.011

Martínez-Castilla, P., & Peppé, S. (2010). Cross-linguistic expression of contrastive accent:

Clinical assessment in Spanish and English. *Clinical Linguistics & Phonetics*, *24*, 955–962. https://doi.org/10.3109/02699206.2010.511411.

Masterson, J. J., & Kamhi, A. G. (1992). Linguistic trade-offs in school-age children. *Journal of Speech, Language, and Hearing Research, 35*(5), 1064–1075.

McAlpine, A., Plexico, L., Plumb, A., & Cleary, J. (2014). Prosody in young verbal children with autism spectrum disorder. *Contemporary Issues in Communication Science and Disorders, 41*, 120–132. https://doi.org/10.1044/cicsd_41_S_120.

McCann, J., Peppe´, S., Gibbon, F. E., O'Hare, A., & Rutherford, M. (2007). Prosody and its relationship to language in school-aged children with high-functioning autism. *International Journal of Language & Communication Disorders*, 42, 682-702. https://doi.org/10.1080/13682820601170102.

McQuarrie, E. F., & Mick, D. G**.** (1996). Figures of rhetoric in advertising language. *Journal of Consumer Research, 22*(4), 424–438. https://doi.org/10.1086/209459.

McRoberts, G. W., & Best, C. T. (1997). Accommodation in mean F0 during parent–infant interactions. *Journal of Child Language, 24*(3), 719–736.

Mehrabian, A. (1995). Theory and evidence bearing on a scale of trait arousability. *Current Psychology, 14*(1), 3–28.

Meise, C., Germer, U., & Gembruch, U. (2001). Arterial Doppler ultrasound in 115 second- and third-trimester fetuses with congenital heart disease. *Ultrasound in obstetrics & gynecology: the official journal of the International Society of Ultrasound in Obstetrics and Gynecology*, *17*(5), 398–402. https://doi.org/10.1046/j.1469-0705.2001.00397.x.

Montrey, J. (2005). An investigation of the effects of speakers' vocal characteristics on ratings of confidence and persuasion [Master's thesis, University of Central Florida]. STARS. https://stars.library.ucf.edu/etd/595.

Montrey, J. (2005). An investigation of the effects of speakers' vocal characteristics on ratings of confidence and persuasion [Master's thesis, University of Central Florida]. STARS. https://stars.library.ucf.edu/etd/595.

Moore J. K. (2002). Maturation of human auditory cortex: implications for speech perception. *The Annals of otology, rhinology & laryngology. Supplement*, *189*, 7–10. https://doi.org/10.1177/00034894021110s502

Moore, D. L., Hausknecht, D., & Thamodaran, K. (1986). Time compression, response opportunity, and persuasion. *Journal of Consumer Research, 13*(1), 85–99. https://doi.org/10.1086/209044.

Movalled, K., Sani, A., Nikniaz, L., & Ghojazadeh, M. (2023). The impact of sound stimulations during pregnancy on fetal learning: a systematic review. *BMC pediatrics*, *23*(1), 183. https://doi.org/10.1186/s12887-023-03990-7

Mozziconacci, S. (2002, April). Prosody and emotions. In *Proceedings of Speech Prosody 2002* (pp. 1–9). ISCA & Laboratoire Parole et Langage. https://doi.org/10.21437/SpeechProsody.2002-1.

Murry, T., Hoit-Dalgaard, J., & Gracco, V. L. (1983). Infant vocalization. *Folia Phoniatrica, 35*(5), 245–253. https://doi.org/10.1159/000265770.

Nadel, L. (Ed.). (1988). *The psychobiology of Down syndrome: Issues in the biology of language and cognition*. MIT Press.

Nadig, A., & Shaw, H. (2015). Acoustic marking of prominence: How preadolescent speakers with and without high-functioning autism mark contrast in an interactive task. *Language, Cognition and Neuroscience, 30*(1–2), 32–47. https://doi.org/10.1080/01690965.2012.753150.

Nadig, A., Lee, I., Singh, L., Bosshart, K., & Ozonoff, S. (2010). How does the topic of

conversation affect verbal exchange and eye gaze? A comparison between typical development and high-functioning autism. *Neuropsychologia, 48*(9), 2730–2739. https://doi.org/10.1016/j.neuropsychologia.2010.05.020.

Nespor, M., & Vogel, I. (1986). *Prosodic phonology*. Foris Publications.

Niebuhr, O., Voße, J., & Brem, A. (2016). What makes a charismatic speaker? A computer-based acoustic–prosodic analysis of Steve Jobs' tone of voice. *Computers in Human Behavior, 64*, 366–382. https://doi.org/10.1016/j.chb.2016.06.059.

Niebuhr, O., Voße, J., & Brem, A. (2016). What makes a charismatic speaker? A computer-based acoustic–prosodic analysis of Steve Jobs' tone of voice. *Computers in Human Behavior, 64*, 366–382. https://doi.org/10.1016/j.chb.2016.06.059.

Nouza, J. (1999). Computer-aided spoken-language training with enhanced visual and auditory feedback. In *Proceedings of Eurospeech '99* (pp. 183–186). Budapest, Hungary.

Ochi, K., Ono, N., Owada, K., Kojima, M., Kuroda, M., Sagayama, S., & Yamasue, H. (2019). Quantification of speech and synchrony in the conversation of adults with autism spectrum disorder. *PLOS ONE, 14*(12), e0225377. https://doi.org/10.1371/journal.pone.0225377.

Oller, D. K., et al. (1976). Infant babbling and speech. *Journal of Child Language, 3*(1), 1–11.

Oomen, C. C., & Postma, A. (2001). Effects of divided attention on the production of filled pauses and repetitions. *Journal of Speech, Language, and Hearing Research, 44*(5), 997–1004. https://doi.org/10.1044/1092-4388(2001/078).

Papaeliou, C., & Trevarthen, C. (2006). Prelinguistic pitch patterns. *Journal of Child Language, 33*(1), 163–178.

Papaeliou, C., Minadakis, G., & Cavouras, D. (2002). Acoustic patterns of infant vocalizations. *Journal of Speech, Language, and Hearing Research*.

Papoušek, M., & Papoušek, H. (1989). Forms and functions of vocal matching. *First Language, 9*(6), 137–157.

Patel, S. P., et al. (2020). An acoustic characterization of prosodic differences in autism spectrum disorder and first-degree relatives. *Journal of Autism and Developmental Disorders, 50*(9), 3032–3045. https://doi.org/10.1007/s10803-020-04392-9.

Paul, R., Augustyn, A., Klin, A., & Volkmark, F. R. (2005). Perception and production of prosody by speakers with autism spectrum disorders. *Journal of Autism and Development Disorders*, 35(2), 205-220. https://doi.org/10.1007/s10803-004-1999-1.

Paul, R., Augustyn, A., Klin, A., & Volkmark, F. R. (2005). Perception and production of prosody by speakers with autism spectrum disorders. *Journal of Autism and Development Disorders*, 35(2), 205-220. https://doi.org/10.1007/s10803-004-1999-1.

Paul, R., Bianchi, N., Augustyn, A., Klin, A., & Volkmar, F. (2008). Production of Syllable Stress in Speakers with Autism Spectrum Disorders. *Research in autism spectrum disorders*, *2*(1), 110–124. https://doi.org/10.1016/j.rasd.2007.04.001.

Pellegrino, F., Coupé, C., & Marsico, E. (2011). *Across-language perspective on speech information rate*. Language, 87(3), 539–558. https://doi.org/10.1353/lan.2011.0057.

Peppé, S., & McCann, J. (2003). Assessing intonation and prosody in children with atypical language development: The PEPS-C test and the revised version. *Clinical Linguistics & Phonetics*, 17(4–5), 345–354.

Peppé, S., McCann, J., Gibbon, F., O'Hare, A., & Rutherford, M. (2007). Receptive and expressive prosodic ability in children with high-functioning autism. Journal of Speech, Language, and Hearing Research, 50, 1015-1028. https://doi.org/10.1044/1092-4388(2007/071).

Peppé, S., McCann, J., Gibbon, F., O'Hare, A., & Rutherford, M. (2006). Assessing prosodic and pragmatic ability in children with high-functioning autism. *Journal of Pragmatics*, *38*(10), 1776-1791. https://doi.org/10.1016/j.pragma.2005.07.004.

Peppé, S., McCann, J., Gibbon, F., O'Hare, A., & Rutherford, M. (2007). Receptive and expressive prosodic ability in children with high-functioning autism. Journal of Speech, Language, and Hearing Research, 50, 1015-1028. https://doi.org/10.1044/1092-4388(2007/071).

Peterson, R. A., Cannito, M. P., & Brown, S. P. (1995). An exploratory investigation of voice characteristics and selling effectiveness. *Journal of Personal Selling & Sales Management, 15*(1), 1–15. https://doi.org/10.1080/08853134.1995.10754008.

Pfanner, L., Tancredi, R., & Marcheschi, M. (2008). Comunicazione e linguaggio nei disturbi

Pierrehumbert, J. B. (1980). *The phonology and phonetics of English intonation* (Doctoral dissertation, Massachusetts Institute of Technology). Indiana University Linguistics Club.

Pierrehumbert, J. B., & Beckman, M. E. (1988). *Japanese tone structure*. MIT Press.

Pierrehumbert, J. B., & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. R. Cohen, J. L. Morgan, & M. E. Pollack (Eds.), *Intentions in communication* (pp. 271–311). MIT Press.

Pogačar, R., Shrum, L. J., & Lowrey, T. M. (2018). The effects of linguistic devices on consumer information processing and persuasion: A language complexity × processing mode framework. *Journal of Consumer Psychology, 28*(4), 689–711. https://doi.org/10.1002/jcpy.1046.

Price, P. J., Ostendorf, M., Shattuck-Hufnagel, S., & Fong, C. (1991). The use of prosody in syntactic disambiguation. *The Journal of the Acoustical Society of America, 90*(6),

2956–2970. https://doi.org/10.1121/1.401770.

Prieto, P., & Vanrell, M. d. M. (2007). Early intonational development in Catalan. *Proceedings of ICPhS XVI*.

Prieto, P., Estrella, A., Thorson, J., & del Mar Vanrell, M. (2012). Is prosodic development correlated with grammatical and lexical development? *Journal of Child Language, 39*(2), 221–257. https://doi.org/10.1017/S030500091100002X.

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

Quené, H., Boomsma, G., & van Erning, R. (2016). Attractiveness of male speakers: Effects of voice pitch and speech tempo. In *Proceedings of Speech Prosody 2016* (pp. 1086–1089). ISCA. https://www.isca-archive.org/speechprosody_2016/quene16_speechprosody.pdf.

Querleu, D., Renard, X., Boutteville, C., & Crepin, G. (1989). Hearing by the human fetus? *Seminars in perinatology*, *13*(5), 409–420.

Rasbash, J., et al. (2005). *A user's guide to MLwiN Version 2.0*. University of Bristol.

Raven, J. C. (1995). *Manual for the Colored Progressive Matrices (Revised*). Windsor: NFRE-Nelson.

Robb, M. P., & Saxman, J. H. (1985). Developmental trends in vocal fundamental frequency. *Journal of Speech, Language, and Hearing Research, 28*(3), 421–427. https://doi.org/10.1044/jshr.2803.421.

Robb, M. P., Saxman, J. H., & Grant, A. A. (1989). Vocal fundamental frequency characteristics. *Journal of the Acoustical Society of America, 85*(4), 1708–1717. https://doi.org/10.1121/1.397960.

Rodero, E. (2020). Do your ads talk too fast to your audio audience? How speech rates of audio commercials influence cognitive and physiological outcomes. *Journal of Advertising Research, 60*(3), 337–349. https://doi.org/10.2501/JAR-2019-038.

Rodero, E., Potter, R. F., & Prieto, P. (2017). Pitch range variations improve cognitive processing of audio messages. *Human Communication Research, 43*(3), 397–413. https://doi.org/10.1111/hcre.12110.

Rooth, M. (1992). "A theory of focus interpretation," *Natural Language Semantics* 1, 75–116.

Rothgänger, H. (2003). Analysis of the sounds of the child. *Early Human Development, 75*(1–2), 55–69. https://doi.org/10.1016/j.earlhumdev.2003.09.003

Ruesch, J. (1957). *Disturbed communication: The clinical assessment of normal and pathological communicative behaviour*. W. W. Norton & Company.

Samek-Lodovici, V. (2018). Contrast, contrastive focus, and focus fronting. *UCL Working Papers in Linguistics.*, *2018*, 57-77.

Savino, M. (2012). *The intonation of polar questions in Italian: Where is the rise?* Journal of the International Phonetic Association, 42(1), 23–48. https://doi.org/10.1017/S002510031100048X.

Savino, M., Wehrle, S., & Grice, M. (2024). The prosody of Italian newsreading: A diachronic analysis. In *Proceedings of Speech Prosody 2024* (pp. 836–840). https://doi.org/10.21437/SpeechProsody.2024-169.

Savy, R., & Cutugno, F. (2009). Diatopic, diamesic and diaphasic variations in spoken Italian. In *Proceedings of the 5th Corpus Linguistics Conference* (CL2009) (pp. 20–23).

Schulman, H., Fleischer, A., Stern, W., Farmakides, G., Jagani, N., & Blattner, P. (1984).

Umbilical velocity wave ratios in human pregnancy. *American journal of obstetrics and gynecology*, *148*(7), 985–990. https://doi.org/10.1016/0002-9378(84)90541-6

Selkirk, E. (1984). *Phonology and syntax*. Cambridge, MA: MIT Press.

Shahidullah, S., & Hepper, P. G. (1994). Frequency discrimination by the fetus. *Early human development*, *36*(1), 13–26. https://doi.org/10.1016/0378-3782(94)90029-9.

Shattuck-Hufnagel, S., & Turk, A. E. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research, 25*(2), 193–247. https://doi.org/10.1007/BF01708572.

Shriberg, L. D. (1993). Four new speech and prosody–voice measures for genetics research and other studies in developmental phonological disorders. *Journal of Speech and Hearing Research, 36*(1), 105–140. https://doi.org/10.1044/jshr.3601.105.

Shriberg, L. D., Paul, R., McSweeny, J. L., Klin, A., Cohen, D. J., & Volkmar, F. R. (2001). Speech and prosody characteristics of adolescents and adults with high-functioning autism and Asperger syndrome. Journal of Speech, Language, and Hearing Research, 44, 1097-1115. https://doi.org/10.1044/1092-4388(2001/087).

Simões, M. (2000). Investigações no âmbito da aferição nacional do Teste das Matrizes Progressivas Coloridas de Raven (M.P.C.R). Lisboa: Fundação Calouste Gulbenkian.

Sjöström, K., Valentin, L., Thelin, T., & Marsál, K. (1997). Maternal anxiety in late pregnancy and fetal hemodynamics. *European journal of obstetrics, gynecology, and reproductive biology*, *74*(2), 149–155. https://doi.org/10.1016/s0301-2115(97)00100-0.

Smith, A., & Zelaznik, H. N. (2004). Development of functional synergies. *Developmental Psychobiology, 45*(1), 22–33. https://doi.org/10.1002/dev.20009

Smith, S. L., Gerhardt, K. J., Griffiths, S. K., Huang, X., & Abrams, R. M. (2003). Intelligibility of sentences recorded from the uterus of a pregnant ewe and from the fetal inner ear. *Audiology and Neuro-Otology, 8*(6), 347–353. https://doi.org/10.1159/000073519.

Smith, S. M., & Shaffer, D. R. (1995). Speed of speech and persuasion: Evidence for multiple effects. *Personality and Social Psychology Bulletin, 21*(10), 1051–1060. https://doi.org/10.1177/0146167295211003.

Snow, D. (2004). Falling intonation in infant utterances. *Journal of Phonetics, 32*(3), 373–393. https://doi.org/10.1016/S0095-4470(03)00038-X.

Snow, D. (2006). Regression and reorganization of intonation. *Child Development, 77*(2), 281–296. https://doi.org/10.1111/j.1467-8624.2006.00870.x.

Snow, D., & Balog, H. L. (2002). Do children produce the melody before the words? A review of developmental intonation research. *Lingua, 112*(12), 1025–1058. https://doi.org/10.1016/S0024-3841(02)00060-8.

Snow, D., & Ertmer, D. J. (2012). Children's development of intonation after cochlear implant. *Clinical Linguistics & Phonetics, 26*(1), 51–70.

Soderstrom, M., et al. (2008). Acoustical cues and grammatical units. *Journal of Child Language, 35*(4), 869–902.

Sorianello, P. (2006). Prosodia: Modelli e ricerca empirica. Carocci

Sorianello, P. (2012). *Linguaggio e sindrome di Down*. FrancoAngeli.

Sorianello, P. (2024). *The prosodic impairment*. In *Speech sciences entries*. Speech Prosody Studies Group. https://gepf.falar.org/entries/69.

Spinelli, M., Fasolo, M., & Mesman, J**.** (2017). Does prosody make the difference? A meta-

analysis on relations between prosodic aspects of infant-directed speech and infant outcomes. *Developmental Review, 44*, 1–18. https://doi.org/10.1016/j.dr.2017.02.001.

Stark, R. E., et al. (1993). Vocal communication in the first 18 months. *Journal of Speech, Language, and Hearing Research, 36*(3), 548–558.

Stewart, D. W. (2006). Putting financial discipline in marketing: A call to action. *Corporate Finance Review, 11*(2), 14.

Stoel-Gammon, C. (2011). Relationships between lexical and phonological development. *Journal of Child Language, 38*(1), 1–34.

Tager-Flusberg, H., Paul, R., & Lord, C. (2005). Communication in autism. In F. Volkmar, A. Klin, R. Paul, & D. Cohen (Eds.), Handbook of autism and pervasive developmental disorders (3rd ed., pp. 335- 364). N.Y.: Wiley & Sons.

Teixidó, M., François, C., Bosch, L., & Männel, C. (2018). The role of prosody in early speech segmentation and word–referent mapping. In P. Prieto & N. Esteve-Gibert (Eds.), *The development of prosody in first language acquisition* (pp. 80–94). John Benjamins.

Tracy, D. K., Ho, D. K., O'Daly, O., Michalopoulou, P., Lloyd, L. C., Dimond, E., & Shergill, S. S. (2011). It's not what you say but the way that you say it: An fMRI study of differential lexical and non-lexical prosodic pitch processing. *BMC Neuroscience, 12*(1), Article 128. https://doi.org/10.1186/1471-2202-12-128. ù

Trevarthen, C. (1977). Descriptive analyses of infant communicative behaviour. In H. R. Shaffer (Ed.), *Studies in mother–infant interaction* (pp. 227–270). Academic Press.

Tristao, R., Jesus, J., Lemos, M., & Freire, R**.** (2020). *Foetal music perception: A comparison study between heart rate and motor responses assessed by APIB scale in ultrasound exam*. Preprints. https://doi.org/10.20944/preprints202007.0345.v1.

Valigura, O., Kozub, L., & Monashnenko, A**.** (2020). Prosody of English television advertising: Sociolinguistic features and pragmatic potential. *Euromentor Journal, 11*(1), 13–38.

Van Bourgondien, M. E., & Woods, A. V. (1992). Vocational possibilities for high-functioning adults with autism. In E. Schopler & G. B. Mesibov (Eds.), *High-functioning individuals with autism* (pp. 227–239). Plenum Press.

Vihman, M. M. (2014). *Phonological development: The first two years*. Wiley.

Vihman, M. M., et al. (1985). From babbling to speech. *Language, 61*(2), 397–445. https://doi.org/10.2307/414151

Wagner, M., & Watson, D. G. (2010). Experimental and theoretical advances in prosody: A review. *Language and Cognitive Processes, 25*(7–9), 905–945. https://doi.org/10.1080/01690961003589492.

Wand, M. P., & Jones, M. C. (1995). *Kernel smoothing*. Chapman & Hall/CRC.

Wasserman, L. (2006). *All of nonparametric statistics*. Springer.

Watson, D. G. (2010) The many roads to prominence: Understanding emphasis in conversation. In B. Ross (Ed.), The Psychology of Learning and Motivation (Vol. 52, pp. 163-183). San Diego, CA: Academic Press.

Wechsler, D. (1991). Wechsler Intelligence Scale for Children (WISC-III): Manual, 3rd Edn. San Antonio, TX: The Psychological Corporation.

Wells, B., & Peppé, S. (2003). Intonation abilities of children with speech and language impairments. *Journal of Speech, Language, and Hearing Research, 46*(1), 5–20. https://doi.org/10.1044/1092-4388(2003/001).

Wells, B., Peppé, S., & Goulandris, A. (2004). Intonation development from five to thirteen. *Journal of Child Language*, *31*, 749–778. https://doi.org/10.1017/S030500090400652X .

Whalen, D. H., et al. (1991). Intonational differences in babbling. *Journal of Child Language, 18*(3), 501–516. https://doi.org/10.1017/S0305000900011223.

Yip, M. (2002). *Tone*. Cambridge University Press. https://doi.org/10.1017/CBO9780511486371.011.

Yue, W., Han, X., Luo, J., Zeng, Z., & Yang, M. (2021). Effect of music therapy on preterm infants in neonatal intensive care unit: Systematic review and meta-analysis of randomized controlled trials. *J. Adv. Nurs.*, 77(2), 635–652. https://doi.org/10.1111/jan.14630.

Zampini, L., et al. (2015). Prosodic skills in children with Down syndrome. *International Journal of Language & Communication Disorders*.

Zampini, L., Fasolo, M., Spinelli, M., Zanchi, P., Suttora, C., & Salerni, N. (2016). Prosodic skills in children with Down syndrome and in typically developing children. *International Journal of Language & Communication Disorders, 51*(1), 74–83.

Zanchi, P., et al. (2016). Syntax and prosody in narratives. *First Language, 36*(2), 124–139.

Zanchi, P., et al. (2016). The role of acoustic and contextual features in the recognition of crying causes. *Psicologia Clinica dello Sviluppo, 20*(1), 103–123.

Zanchi, P., Peppé, S., & Zampini, L. (2022). Prosodic development from 4 to 10 years: Data from the Italian adaptation of the PEPS-C. *Speech Communication*, *144*, 10–19. https://doi.org/10.1016/j.specom.2022.08.007.

Zhou, W. (2018). *Stylistic analysis of English advertising language*. In *Proceedings of the International Conference on Contemporary Education, Social Sciences and Ecological Studies (CESSES 2018)* (pp. 399–403). Atlantis Press. https://doi.org/10.2991/cesses-18.2018.89.

Zimmermann, M. (2007). Contrastive focus. In C. Féry & M. Krifka (Eds.), The notions of information structure (pp. 147–159). Universitätsverlag Potsdam. http://opus.potsdam.de/opus/volltexte/2008/1652/.

Zmarich, C., Magno Caldognetto, E., & Ferrero, F. (1997). Analisi confrontativa di parlato spontaneo e letto: Fenomeni macroprosodici e indici di fluenza. In *Atti delle VII Giornate di Studio del GFS* (pp. 111–139). Napoli, Italy.

Acknowledgements

I would like to thank my supervisor, Prof. Mirco Fasolo, for encouraging me to move beyond the technical practice of linguistics and to place it within a broader psychological perspective. He constantly pushed me to look past the acoustic details of speech analysis and to consider the wider picture of how the human brain processes language. His guidance helped me develop the interdisciplinary approach that shapes this thesis.

I would also like to thank Professors Sónia Frota and Marisa Filipe at the Lisbon Centre for Linguistics for giving me the tools to deepen my research in prosody. Their support helped me combine my phonetic and acoustic background with phonological theory, and to better approach the study of both typical and atypical prosody. Their expertise and guidance were essential to my growth as a researcher.