An acoustic investigation of the developmental trajectory of lexical stress contrastivity in Italian

Joanne Arciuli [a,*], Lucia Colombo [b]

[a] *Australian Research Council and Faculty of Health Sciences, The University of Sydney, PO Box 170, Lidcombe, 1825, NSW, Australia*
[b] *Department of General Psychology, The University of Padova, Via Venezia 8, 35131, Padova, Italy*

Abstract

We examined whether typically developing Italian children exhibit adult-like stress contrastivity for word productions elicited via a picture naming task ($n = 25$ children aged 3–5 years and 27 adults). Stimuli were 10 trisyllabic Italian words; half began with a weak–strong (WS) pattern of lexical stress across the initial 2 syllables, as in *patata*, while the other half began with a strong–weak (SW) pattern, as in *gomito*. Word productions that were identified as correct via perceptual judgement were analysed acoustically. The initial 2 syllables of each correct word production were analysed in terms of the duration, peak intensity, and peak fundamental frequency of the vowels using a relative measure of contrast—the normalised pairwise variability index (PVI). Results across the majority of measures showed that children's stress contrastivity was adult-like. However, the data revealed that children's contrastivity for trisyllabic words beginning with a WS pattern was not adult-like regarding the PVI for vowel duration: children showed less contrastivity than adults. This effect appeared to be driven by differences in word-medial gemination between children and adults. Results are compared with data from a recent acoustic study of stress contrastivity in English speaking children and adults and discussed in relation to language-specific and physiological motor-speech constraints on production.

1. Introduction

Speech production is a truly remarkable feat. While we have made great progress in understanding the production of individual speech sounds, less is known about the prosodic aspects of speech production. The distinction between strong and weak syllables within single words is a type of prosody known as lexical stress (note that there is no lexical stress in monosyllabic words). For example, in English, *incense* has different meanings depending on whether the pattern of lexical stress is strong–weak (first syllable stress) or weak–strong (second syllable stress). Similarly, the meaning of *ancora* in Italian changes depending on the assignment of lexical stress. Consider the English word *zebra* which has a strong–weak pattern of stress; if one were to produce this word with a

weak–strong pattern one may not be understood. This is because stress plays a role in lexical access during spoken word recognition on the part of the listener (see studies of adults by Arciuli and Cupples, 2004, Arciuli and Slowiaczek, 2007, and Cooper et al., 2002, as well as studies of infants and children by Curtin, 2010, and Quam and Swingley, 2014). Thus, appropriate production of lexical stress is critical for intelligibility during spoken communication (Field, 2005; Klopfenstein, 2009).

The ability to achieve stress contrastivity during word production is more vulnerable to developmental influences than previously thought. For example, it can be atypical in some individuals with the neurodevelopmental disability of autism (see Arciuli, 2014, for a review). A recent study of English speaking typically developing children aged 3–7 years revealed that even when word productions were identified as correct via perceptual judgement there were fine-grained acoustic differences in the way children realise stress

---
* Corresponding author.
E-mail address: joanne.arciuli@sydney.edu.au (J. Arciuli).

contrastivity in their production of particular words by comparison with adults (Ballard et al., 2012). To date, much of the research on children's production of lexical stress has focussed only on speakers of English (but see DePaolis et al., 2008, for a cross-linguistic acoustic study of lexical stress in infants exposed to English, Finnish, French or Welsh at the onset of word use at 10–18 months of age). Questions remain regarding whether *language-specific factors* (such as children's familiarity with certain patterns of lexical stress that are most common in their language) or *physiological constraints* associated with the developing motor-speech system (that may affect words with particular patterns of lexical stress regardless of the language in question), or both, might be at play when children are mastering the production of lexical stress. In the current study, we focussed on speakers of standard Italian, comparing the word productions of typically developing preschool children aged 3–5 years with those of healthy adults, in order to further explore this question.

1.1. Measuring stress contrastivity

Listeners perceive lexical stress as the distinction between strong and weak syllables within single words. This distinction is realised acoustically in a number of ways and, to a certain extent, depends on the language in question. Even so, the key acoustic correlates of lexical stress relate to the duration, intensity, and fundamental frequency of vowels.

One acoustic measure that can be used to gauge the amount of stress contrastivity across adjacent syllables is the pairwise variability index (PVI). In the literature there are references to the 'raw PVI' and the 'normalised PVI' (different equations are used to derive these values, as explained by Nolan and Asu, 2009). The normalised PVI (where the pairwise difference between syllables is divided by the average value of the pair) is preferable to the raw PVI and far more useful than raw data for individual syllables. This is because both the raw PVI and raw data from individual syllables reflect individual differences amongst speakers. For example, speakers vary in their rate of speech, which is reflected in duration values. Speakers also vary in how loudly they speak which is reflected in intensity values. Finally, speakers vary in terms of the pitch of their voice which is reflected in values for fundamental frequency. In effect, the normalised PVI neutralises these kinds of differences, including any differences which may relate to gender, age, emotional state and so on.

By way of background, the PVI was originally used to examine variability in vowel duration across successive syllables of read sentences in order to understand the rhythmic patterning of different languages (e.g., Grabe and Low, 2002; Low et al., 2000; Tan and Low, 2014). It is one of the so called 'rhythm metrics' that has been useful in exploring the idea that languages are either stress-timed (duration of successive syllables can be uneven) or syllable-timed (duration of successive syllables is even). This idea, originally put forward by Pike (1945) and Abercrombie (1967) was based on perceptual evaluations; many direct investigations of the acoustic signal in the decades since have failed to support a strict

dichotomy (e.g., Roach, 1982; Dauer, 1983, and Nespor, 1990, but see Ramus et al., 1999). A number of acoustic studies, that have specifically compared English (traditionally described as stress-timed) and Italian (traditionally described as syllable-timed), have indicated that both of these languages exhibit variability in vowel duration across successive syllables although there are differences in the amount of variability (e.g., Braun and Geiselmann, 2011; Mairano and Romano, 2007; Vayra et al., 1987).

Rhythm metrics have uses beyond their original purpose. The normalised PVI has been applied beyond the analysis of vowel duration, to include analysis of the intensity and fundamental frequency of vowels, in order to explore stress contrastivity in the production of English words. For example, Arciuli et al. (2014) used the normalised PVI to provide new insights regarding speech produced in noise by showing that adults do not uniformly increase the intensity of adjacent syllables within words when speaking in noise; rather, they increase the amount of stress contrastivity. Ballard et al. (2012) used the normalised PVI to examine the developmental trajectory of stress contrastivity in the word productions of children versus adults. The methodology of the current study follows the study by Ballard and colleagues, which is described in full in the next section.

Another measure of stress contrastivity is the lexical stress ratio (LSR) which provides a single score per word production (Shriberg et al., 2003). It is a weighted composite of acoustic data relating to the duration, intensity, and fundamental frequency of vowels in adjacent syllables derived from a principal components analysis. The speech that was used in its formulation was elicited from English speaking children, with either speech delay or suspected apraxia of speech, who produced 8 words with initial syllable stress via an imitation task. The PVI has broader applicability than the LSR because it has been used to measure words with differing stress patterns (not just those with stress on the initial syllable), produced by both typical and atypical speakers. In addition, the normalised PVI can be used to derive separate scores relating to duration, intensity, and fundamental frequency whereas the LSR derives a single score which combines these acoustic measures.

1.2. Children's mastery of stress contrastivity

Comparison of child versus adult speech reveals developmental trajectories and can help us understand when children reach 'mastery'. This approach has been used to explore fine-grained acoustic differences in the production of individual segments (e.g., Li, 2012; Nicholson et al., 2015) and stress contrastivity across adjacent syllables (e.g., Schwartz et al., 1996; Ballard et al., 2012).

In an acoustic study of young children's production of lexical stress in English Schwartz et al. (1996) examined the production of bisyllabic nonwords by 14 typically developing two year olds via unsolicited imitation. Nonwords were constructed so that there was a trochaic or strong–weak (SW) version as well as an iambic or weak–strong (WS) version. The