



Research Article

Vowel space area in later childhood and adolescence: Effects of age, sex and ease of communication



Michèle Pettinato^{a,b,*}, Outi Tuomainen^b, Sonia Granlund^b, Valerie Hazan^b

^a Department of Linguistics, University of Antwerp, Prinsstraat 13, B-2000 Antwerpen, Belgium

^b Speech, Hearing & Phonetic Sciences, Chandler House, 2 Wakefield Street, London WC1N 1PF, UK

ARTICLE INFO

Article history:

Received 4 August 2014

Received in revised form

2 July 2015

Accepted 8 July 2015

Keywords:

Vowel space development

Childhood and adolescence

Vowel hyperarticulation

Spontaneous speech

Clear speech

Fine phonetics

Articulatory development

ABSTRACT

This study investigated vowel space area (VSA) development in childhood and adolescence and its impact on the ability to hyperarticulate vowels. In experiment 1, 96 participants aged 9–14 years carried out an interactive task when communication was easy (no barrier, 'NB') and difficult (the speech of one participant was filtered through a vocoder, 'VOC'). Previous recordings from 20 adults were used as reference. Measures of VSA (ERB²), F1 and F2 ranges (ERB) and articulation rate were obtained. Children's VSA were significantly larger than adults'. From the age of 11, vowel hyperarticulation was evident in VOC, but only because VSA were gradually reducing with age in NB. The results suggest that whilst large VSA do not prevent children from hyperarticulating vowels, the manner in which this is achieved may not be adult-like. Experiment 2 was conducted to verify that large VSA were not a by-product of children being unable to see each other. Thirteen participants carried out the same task face-to-face with their interlocutor. Comparisons to matched participants from experiment 1 showed no differences in VSA, indicating that the audio-only modality did not influence results. Possible reasons for larger VSA in the spontaneous speech of children and adolescents are discussed.

© 2015 The Authors. Published by Elsevier Ltd. All rights reserved.

1. Introduction

The present study is concerned with the production of vowels in conversational speech produced with communicative intent by children aged 9–14 years. The focus is on how vowel space area (VSA) is affected by age and sex within this age span. We also investigate whether children and adolescents use vowel space expansion as a clear speech strategy when interacting with an interlocutor who is receiving a distorted speech signal.

Typically-developing children are able to produce vowels contrastively as early as age three years (Donegan, 2002; Stoel-Gammon & Herrington, 1990), in spite of having to compensate for changes in vocal tract size throughout childhood and adolescence (Kent & Vorperian, 1995). Anatomically, children have smaller vocal tracts which result in higher formant values; formant values decrease throughout childhood as vocal tract size increases (Eguchi & Hirsh, 1969; Lee, Potamianos & Narayan, 1999; Perry, Ohde, & Ashmead, 2001). However, children's vowels do not seem to be merely shifted towards higher frequencies: they are also typified by greater acoustic variability. Also, the VSA in children's speech are significantly larger than those of adults. These characteristics are unlikely to be solely due to smaller vocal tracts.

Amongst studies of vowel development, Lee et al.'s (1999) influential study covers the widest (cross-sectional) age ranges from 5 to 18 years. In their study, 10 American English monophthongal vowels were elicited in a carrier sentence and compared to tokens produced by adults. Importantly, this study not only charted developing formant values, spectral envelopes and durations for individual vowels, but focused on acoustic variability within categories. Lee et al. showed that segmental acquisition was characterized by a gradual reduction of variability for spectral and durational aspects throughout childhood, with substantial reduction between ages 8 and 14 years. Adult-like variability was reached in spectral measures by around 14 years and in durational measures by around 12 years. Insofar as acoustic variability approximates articulatory variability, it was suggested that younger

* Correspondence to: Computational Linguistics & Psycholinguistics Research Center, Universiteit Antwerpen – Stadscampus, Lange Winkelstraat 40-42, B-2000 Antwerpen, Belgium. Tel.: +32 3 265 5895.

E-mail address: michele.pettinato@uantwerpen.be (M. Pettinato).

children did not have robust articulatory targets yet and that neuro-muscular control for articulation continues to develop until early teens. The authors also noted that the VSA of children (i.e., the $F1$ – $F2$ quadrilaterals) were larger than adults'.

The difference between children's and adults' VSA was explored by [Vorperian and Kent \(2007\)](#), who carried out a meta-analysis of child and adult studies of vowel development. They reported a gradual lowering of formants through childhood and adolescence and increasing stability of formant values; this was accompanied by a continuous reduction in VSA size until adulthood. Formant and fundamental frequency ($F0$) values are expected to be higher in children and pre-pubertal children because of their smaller vocal tracts and shorter vocal folds. More intriguing is the concomitant finding of a larger VSA in children. Vorperian and Kent's findings are unlikely to be a by-product of combining separate studies with varying dialects and methods; indeed, [Flipsen and Lee \(2012\)](#) confirmed the pattern of gradual lowering and shrinking in VSA in their re-analysis of a subset of data from [Lee et al. \(1999\)](#). Admittedly these three studies (i.e. [Lee et al., 1999](#); [Vorperian & Kent, 2007](#); [Flipsen & Lee, 2012](#)) compared raw formant measures and did not apply normalization procedures to control for size differences in speakers. Nevertheless, acoustic theory only predicts a shifting of the values towards higher frequencies, not towards the extremes of the vowel area. Rather than being due to the acoustics of the vocal tract, the larger VSA in children may have an articulatory origin. It may be a corollary of children's larger speech movements in early childhood, which may result in vowels being articulated at the extremes of the vowel area ([Green, Moore, & Reilly, 2002](#); [Green & Nip, 2010](#); [Walsh & Smith, 2002](#)). Indeed, large and inefficient movements (i.e., an 'overshoot' of the target) typify immature motor control (see e.g., [Goldfield, Kay & Warren, 1993](#); [Jeannerod, 1988](#)).

Vowel development also shows emergence of female–male differences. The age at which these appear varies between studies. [Lee et al. \(1999\)](#) noted differentiation of $F2$ and $F3$ at 11 years, but this only became significant around age 15; $F0$ differences had reached significance by 12 years. [Eguchi and Hirsh \(1969\)](#) and [Perry et al. \(2001\)](#) found similar ages for onset of sex differences. However, other studies have found much earlier effects: [Busby and Plant \(1995\)](#) and [Whiteside and Hodgson \(2000\)](#) report significant effects of sex on vowel formants and/or $F0$ in children as young as 5 years. In an anatomical study of vocal tract size and volume between the ages 2 and 25 years ([Fitch & Giedd, 1999](#)), significant sex dimorphism emerged around 10 years and became highly significant around 14 years. The authors suggested that, before age 10, sex differences in vowels are due to behavioural and social factors. In terms of VSA, [Flipsen and Lee \(2012\)](#) report larger vowel spaces in the female speakers in [Lee et al.'s \(1999\)](#) study from age 16 onwards; larger VSA are also found for adult females ([Hillenbrand, Getty, Clark, & Wheeler, 1995](#); [Whiteside, 2001](#)). Again, a physiological explanation is not fully satisfactory. Instead, [Simpson \(2009\)](#) has suggested that this may be part of an altogether more hyperarticulated speech style for women.

In summary, children's acoustic signatures for vowels are more variable, their VSA is larger and their vowels are longer than adults'. Furthermore, articulation rate in children is slower than in adults, gradually increasing towards adult levels over the first 10 years of life ([Logan, Byrd, Mazzocchi, & Gillam, 2011](#)), although the exact age at which adult level articulation rates are found varies (see e.g., [Lee et al., 1999](#); [Sturm & Seery, 2007](#)). These characteristics may have an effect on children's ability to implement changes at the level of fine phonetic detail. Clear speech, a variety of speech that is produced in response to a communication barrier in the environment (e.g., noise or reverberation) or affecting the interlocutor (e.g. hearing impairment or lack of native mastery of the language), is partly reliant on alterations at fine phonetic level. [Lindblom's \(1990\)](#) Hyper–Hypo theory of speech production posits speech production as being guided by the needs of the interlocutor within an interactive setting. Lindblom suggests that there are conflicting forces between the need for a talker to provide clear information for the listener and the need to minimize effort in her/his articulatory gestures. In these terms, clear speech is a style which is at the hyper-extreme of the continuum, where differences between categories are exaggerated. As linguistic categories are still in the process of being fully established in childhood, it is important to investigate if Lindblom's Hyper–Hypo theory also applies to children's speech.

Clear speech phenomena have been roughly divided into global acoustic alterations which affect the discriminability of the speech signal, and into finer phonetic changes, which increase discriminability between meaningful linguistic units. Global acoustic alternations include a reduction in speech rate, more frequent pausing, increases in intensity, raised mean $F0$ and greater $F0$ range ([Cooke, King, Garnier, & Aubanel, 2014](#); [Krause & Braid, 2004](#); [Smiljanic & Bradlow, 2005](#)). Fine phonetic changes have been shown for Voice Onset Time measures in consonants ([Smiljanic & Bradlow, 2008](#); [Sundberg, 2001](#); [Synnestvedt, Bernstein-Ratner, & Newman, 2010](#)) and terms of increased in the $F1$ – $F2$ vowel space ([Ferguson & Kewley-Port, 2007](#); [Hazan & Baker, 2011](#); [Krause & Braid, 2004](#)). For comprehensive reviews of clear speech phenomena in adults and their effects on the listener, see [Cooke et al., 2014](#); [Smiljanic & Bradlow, 2009](#) and [Uchanski, 2005](#).

There are large individual differences in how effectively clear speech is produced ([Ferguson, 2004](#); [Hazan & Markham, 2004](#)). Although most of this variability is arbitrary, a recurrent finding has been that female speakers are perceived as being clearer than males on average, both in casual and clear speech settings (see e.g. [Bradlow, Torretta, & Pisoni, 1996](#); [Ferguson, 2004](#), [Hazan & Markham, 2004](#)). Authors have conjectured that greater intelligibility may, in fact, be a feature of women's sociolect ([Fant, 1975](#)). For a review of phonetic differences between male and female speakers, see [Simpson \(2009\)](#).

Since children's VSA seem to be larger than those of adults', possibly due to less skilled articulatory movements, the question of how this affects their ability to exaggerate the vowel space for clear speech becomes relevant. Indeed, children have been shown to have remarkable command of fine phonetics; for example, [Foulkes and Docherty \(2006\)](#) discuss the gradual emergence of gender-specific patterns of pre-aspiration in prepausal /t/ in 3-year old children from a working-class community in Newcastle. Also, gender-specific patterns in the production of sibilant fricatives have been documented in children as young as 3 years in American English ([Holliday, Beckman, & Mays, 2010](#)). However, there is a temporal and functional difference between clear speech and regional, socio-economic or gendered variation. The latter presumably belong to the default setting of a speaker's phonetics and are stored in some long-term modality (although their expression can be varied according to the setting/listener and are by no means fixed, see [Labov,](#)

1986, 1994). In contrast, clear speech is a temporary modification in response to a specific disruption. Therefore, the study of the development of clear speech abilities presents the opportunity of observing whether children are flexible and skilled at calibrating their linguistic and phonetic output in response to short-term changes in their environment. We can also observe children's ability to respond to the communicative needs of their conversational partners.

The topic of children's clear speech abilities is relatively under-researched. Redford and Gildersleeve-Neuman (2007, 2009) compared the same words produced by children aged 3–5 years in casual interaction and under formal instruction to “speak clearly.” Adult raters were only able to perceive acoustic differences between children's ‘casual’ and ‘clear’ tokens from age 4 onwards. Acoustic measurements gave evidence of some strategies which deviated from typical adult clear speech: from age 4, the clear tokens had a faster rate and overall lower F_0 , although F_0 range did not differ between styles. The ability to enhance at the segmental level showed gradual emergence in the speech of 5 year olds: although initial stops did not differ between clear and casual speech, clear speech had a greater percentage of final stop release. Similarly, whilst the VSA was not expanded, F_1 was higher overall for clear speech vowels. Interestingly, the VSA in the casual condition was not reduced, therefore not showing the articulatory shortcuts typical of casual speech. The authors comment that the somewhat anomalous F_0 and rate results may have been due to the children interpreting the instructions to ‘speak clearly’ and ‘with your big girl/boy voice’ as requesting them to speak like an adult. Indeed, using a different elicitation method where children had to teach a puppet to speak, Syrett and Kawahara (2013) report that preschoolers produced words that had longer and more intense vowels, as well as a more dispersed VSA and expanded F_0 range than in normal speech. Therefore, it seems that children are able to produce vowel hyperarticulation, though it is not clear whether this hyperarticulation is comparable to and as effective as that produced by adults.

Most of the studies on vowel development have used a form of laboratory speech where children were instructed to say simple phrases in a particular style, and it is possible that this artificial experimental design may have influenced some of the findings (see e.g. Hazan & Baker, 2011; Lam & Tjaden, 2013). For example, in Lee et al.'s (1999) study, vowels were measured in monosyllabic keywords that were embedded in a carrier sentence. Far less is known about vowel development in conversational speech; because of connected speech phenomena, VSA may well be less enlarged in children's natural conversations. It would be of interest therefore to compare vowel measures obtained using conversational speech with Lee et al.'s data. Similarly, clear speech phenomena are sensitive to the type of instruction given (Lam & Tjaden, 2013), and the studies reviewed here have either used a child-friendly version of laboratory speech or fairly constrained scripted exchanges. As, according to Lindblom's (1990) Hyper–Hypo theory, speech adaptations are made specifically for a listener's needs, it is important to use an interactive task in which a real interlocutor is present; previous studies (e.g. Scarborough, Brenier, Zhao, Hall-Lew, & Dmitrieva, 2007) have shown that speech elicited to an imagined interlocutor may indeed differ to that elicited in a real communicative situation involving a listener. The work presented here therefore took an interactional approach and used a conversational task to elicit speech samples. This allowed us to examine speech development in a more naturalistic setting that is more likely to reflect children's everyday experience. Syrett and Kawahara (2013) showed that children can enhance their speech. However, the question then is whether they do this spontaneously during interactions when it is appropriate for their interlocutor's needs. The 96 participants in our study were aged between 9 and 14 years, covering a period of marked development in the fine phonetic detail of vowels and speech in general (Lee et al., 1999), converging towards adult values during this time. Moreover, other studies have also located the emergence of gender effects within this age range (e.g. see Romeo, Hazan, & Pettinato, 2013 for fricative development).

For the first study, pairs of participants carried out the diapix task (Van Engen et al., 2010; Baker & Hazan, 2011) in a laboratory, where participants communicated via headsets and sat in different booths during the recordings, without sight of each other. Diapix is a spot-the-difference task which was used in good and adverse listening conditions to elicit casual and clear conversations. The data from the children were compared to recordings for 20 adults recordings from a previously collected corpus using the same methodology (see Baker & Hazan, 2011; Hazan & Baker, 2011). In order to verify whether the lack of visual feedback and the unusual setting may have had a more deleterious effect on conversational ease in children than in adults, a second diapix study was carried out in a classroom setting where child participants were face-to-face during the task. Vowel measures were compared across the child laboratory and child classroom studies.

The following research questions were addressed:

- 1) Vowel space area development: is an expanded VSA (relative to adults) also characteristic of conversational speech in children? How does it evolve with age? Because of connected speech phenomena and the more casual setting of a conversational task, we predicted the differences in adults and children to be attenuated. We expected to see a gradual shrinking in VSA, with possible adult-like values by the age of 14 years.
- 2) Vowel hyperarticulation: can children use this strategy to clarify their speech? If by default children's speech movements are bigger and their VSA accordingly enlarged, further expansion may be difficult to achieve, especially for younger children. However, since Syrett and Kawahara (2013) documented vowel hyperarticulation in preschoolers, and segmental variability is expected to have reached adult levels around age 14 (Lee et al., 1999), we predicted that the oldest participants would have enlarged VSA. A secondary concern focused on the relationship between speech rate and vowel expansion. As faster rates may be associated with more mature articulation skills (Redford, 2014), we hypothesized that faster talkers may be more able to hyperarticulate.
- 3) Sex: the adult literature reports sex effects on vowel space (Hillenbrand et al., 1995; Whiteside, 2001) and rate (e.g., Jacewicz, Fox, & Wei, 2010; Quené, 2008), with women having a slower speech rate and a more expanded VSA. Can we identify these effects in children and if so, when do they emerge?

2. Experiment 1

The first study examined the development of VSA and articulation rate in conditions where communication was easy ('no barrier', NB) or difficult ('vocoded', VOC). The extent to which participants used vowel hyperarticulation and its relation with articulation rate was examined by comparing VSA, formant ranges and articulation rates between NB and VOC conditions.

2.1. Method

2.1.1. Participants

Fifty single-sex pairs of children between the ages of 9 and 14 years inclusive were recruited for this study. The pairs knew each other and were friends. Data from two male pairs could not be included because of non-completion of the recording sessions, resulting in a total of 96 child participants (46M, 50F, mean age: 11;8 years, range 9;0 to 15;0 years). Recordings from 20 adults (10M, 10F, mean age: 22;04, range 18;0–29;0 years) performing the same diapiX task from the LUCID corpus (see Baker & Hazan, 2011; Hazan & Baker, 2011) were used as an adult reference. Participants were non-bilingual native Southern British English speakers¹ who reported no history of hearing or language impairments. All participants passed a hearing screen at 25 dB HL or better at octave frequencies between 250 and 8000 Hz in both ears.

2.1.2. Procedure

During the recording, two participants sat in different sound-treated rooms at University College London (UCL) and communicated via headsets fitted with a condenser cardioid microphone (Beyerdynamic DT297) whilst playing an interactive 'spot the difference game' (diapiXUK; Baker & Hazan, 2011). The speech of each participant was recorded on a separate channel at a sampling rate of 44,100 Hz (16 bit) using an EMU 0404 USB audio interface and Adobe Audition. The experimenter sat outside the rooms and monitored the conversations over headphones. Participants were informed of this and told that the recording would be interrupted or terminated if one of them was uncomfortable. In the NB condition, the two speakers could hear each other without difficulty. In order to elicit clear speech for the VOC condition, the voice of one of the speakers ('talker A') was distorted via a 3-channel noise-excited vocoder before being channelled in real time to talker B. The vocoder was a three-channel version that was described in Rosen, Faulkner and Wilkinson (1999). It was expected that in this vocoder condition, talker A would have to enhance his/her speech for the benefit of talker B. Fig. 1 illustrates how the sentence 'and is there a boy' is affected by vocoding: the upper panel displays the original waveform and spectrogram, and the lower panel shows these after 3-channel vocoding.

The data were collected as part of a wider protocol, but only relevant tasks will be described here (for the adult protocol, see Hazan & Baker, 2011). To familiarise participants with the roles of talker A and B and the nature of differences typically found in the picture sets, children began by receiving training on the diapiX task with a set of pictures that was never used in the recordings. They were each given a picture and seated so that they could not see each other's picture. They were told that the pictures contained 12 differences which they had to find. One child was designated 'the leader' (talker A) and instructed to do most of the talking, whereas the other child (talker B) was mainly there to ask questions and make suggestions. Talker A was instructed as follows: "your job is to describe the picture, so that together you can find the differences. You're the leader, so you should do most of the talking. Your friend can help you with questions and by telling you what s/he sees." With the younger age group, the experimenter gave some of the more reticent participants hints during the training phase (e.g. "ask him/her how many pears there are on the tree"). After they had found several differences, children were allowed to look at each other's pictures and continue comparing them. The participants were told they would take turns at the talker A and B roles in separate recordings, and that they had 10 min to find the 12 differences. They were also told that the recording would either stop when all 12 differences had been found, or when 10 min had passed. They were informed that, during some of the recordings, the voice of talker A would be distorted, and that the experimenter would inform them when this was about to happen. Talker A was told: "you have to find a way to help X to understand you." Moreover, participants were told they would first get to practice their understanding of this distorted speech via a training program prior to carrying out the task in the VOC condition. The experimenter explained to them that the distorted speech they would be hearing from their partner in this condition would be very similar to the training program. They were also asked to be patient with each other, as the speech might not be easy to understand.

Because there is a significant learning effect when listening to vocoded speech (Bent, Buchwald, & Pisoni, 2009; Davis, Johnsruide, Hervais-Adelman, Taylor, & McGettigan, 2005), all participants were given a 10-min vocoder training session using the software described in Faulkner, Rosen, and Green (2012). For the training, the software presented sentences of a story one at a time. The sentences were passed through the same vocoder that was used in the VOC condition. After each sentence, the programme displayed a selection of words on a computer screen and the participants had to click on those words which had been part of the sentence. Every time they clicked on a word that was incorrect, the sentence was replayed. Once they had clicked on all the correct words, the sentence was printed on the screen and replayed. The next sentence was then played. The training lasted 10 min since previous research identified this to be sufficient for significant learning of vocoded speech (Bent et al., 2009).

Every pair of children/adolescents carried out six recordings with different sets of pictures: two without a communication barrier (NB) and two where the speech of talker A was filtered through the vocoder. The final two recordings were for an additional condition with adult multi-talker babble (see Hazan & Baker, 2011), but only results from NB and VOC conditions are reported here. During the recording session, although talker A did not experience the effect of vocoding directly, s/he nevertheless had a sense of the communication barrier talker B was experiencing, as both speakers had undergone the training programme.

¹ All 96 children are schooled in England and used English as their primary language for communication. However, five children (5%) reported an additional language used at home, (e.g. by grandparents), but none were fluent and they did not actively use their L2 at the time of the recordings.

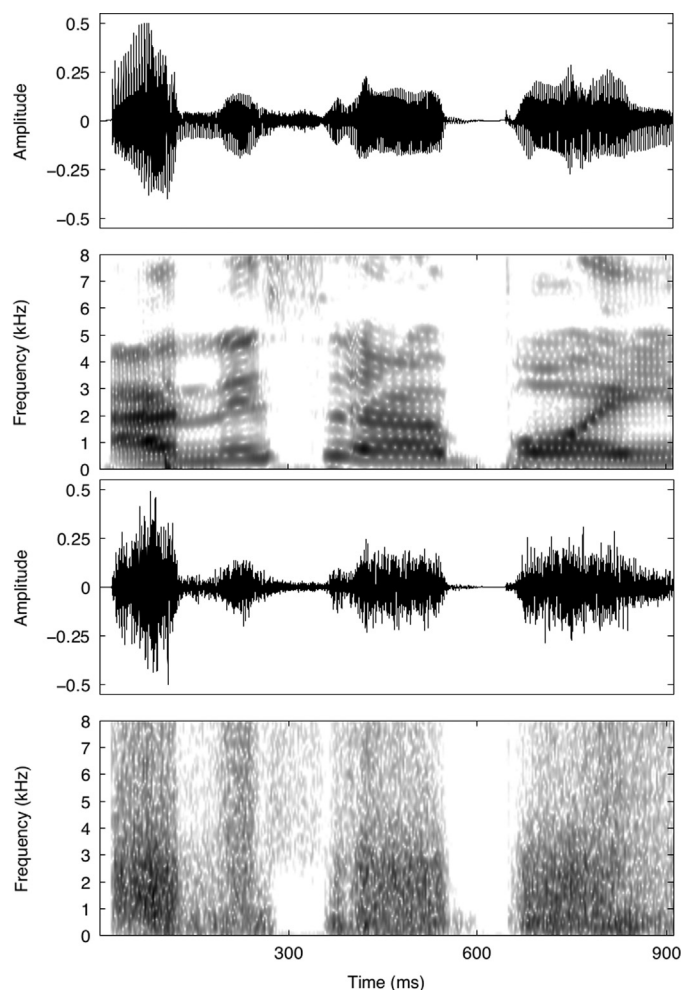


Fig. 1. Waveforms and spectrograms of the utterance “and is there a boy” produced by a female child in the current study. The upper two panels represent the speech in the no-barrier condition (no distortions applied), and the lower two panels represent the speech after being distorted through a 3-channel vocoder.

Every pair started out with a recording in NB, and pairs of participants were counterbalanced between doing two VOC recordings first or the two babble recordings (not reported here). Everyone ended with the second NB recording. Participants switched roles between recordings in each condition, so that every child was talker A once in NB and VOC. Recordings were stopped once all differences had been found or after 10 min. The recording equipment for the adults was the same, but the order of tasks varied slightly from the children (see [Baker & Hazan, 2011](#) for more details); the adults carried out three picture tasks per condition but only the first of these was used so that the amount of data were comparable to that collected for the child participants.

2.1.3. Data processing

For all recordings, each channel was transcribed using freeware transcription software from Northwestern University's Linguistics Department (Wavescroller) to a set of transcription guidelines based on those used by [Van Engen et al. \(2010\)](#) with minor adaptations for the coding of pauses. Word- and phoneme-level alignment software which was developed in-house at UCL, based on the Hidden Markov Model Toolkit ([HTK Team, 2012](#)) was used to automatically align the transcriptions and create Praat Textgrids with separate word and phoneme tiers. Alignment was manually checked and corrected in two stages. First, the word level alignment of all the files was manually checked and adjusted where necessary. Words for which any boundary changes occurred while checking were extracted from each Textgrid file, and their phoneme level was automatically re-aligned to adjust for changed word boundaries. As the aligner was observed not to always be entirely accurate in determining vowel boundaries, the alignment of the three vowels included in this study was then verified and corrected by hand where necessary in the new textgrid files. Recordings lasted approximately 10 min, yielding around 4 min of analysable speech for talker A once silences, fillers, non-speech sounds such as laughter and sections with background noise had been excluded.

Only talker A's speech was analysed. VSA was examined by analysing three corner vowels in content words: [i:], [æ] and [ɔ:]. These vowels were chosen amongst available monophthongs because (1) they were the most frequent per individual participant recordings – recall that this was spontaneous, unscripted speech (2) they had the best differentiation in terms of front-back and high-low distinctions and therefore covered the largest distances in the $F1$ – $F2$ space. Most clear speech studies (e.g., [Ferguson & Kewley-Port, 2007](#); [Smiljanic & Bradlow, 2005](#)) use three or four corner vowels to calculate VSA. Consonantal context for the vowels could not be controlled for, even though flanking consonants can have an effect on vowel formants ([Strange et al., 2007](#)). Nevertheless, it

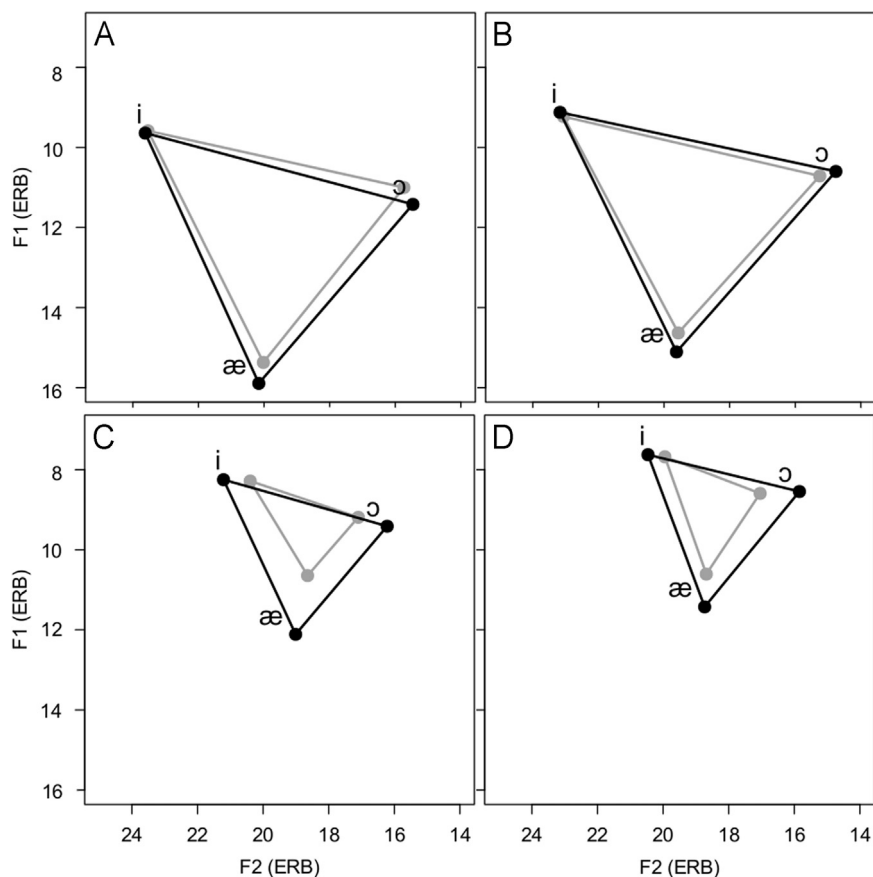


Fig. 2. Vowel space area (ERB²) between vowels [i:], [ɔ:] and [æ] for the 'no barrier' (NB) (grey) and communication barrier (VOC) (black) conditions in the girls (A), boys (B), women (C) and men (D).

was expected that the large numbers for each vowel (5212 for [i:], 3660 [æ] and 2742 [ɔ:]) would counteract the potential noise in the measurements due to different consonantal contexts.

On average, 29 [i:], 21 [æ] and 15 [ɔ:] vowel tokens were included in the calculations of vowel measures per talker for NB and 25 [i:], 18 [æ] and 14 [ɔ:] tokens for the VOC condition. Formant tracking algorithms by Burg in Praat were used to get the two lowest formants (F_1 , F_2) in all content words with [i:], [æ], [ɔ:] vowels over 50 ms in duration. Formant frequencies were measured from the midpoint of the vowel and the maximum frequency of formant calculation algorithm was based on sex (Male: 500 and 1485 Hz; Female: 550 and 1650 Hz). Outliers were determined by using a ± 2 SD cut-off criterion within each individual. Formant estimates were converted to Equivalent Rectangular Bandwidth (ERB) values to reduce the effect of anatomical differences due to sex and age, and median F_1/F_2 ERB values were calculated per vowel per talker. This was achieved using the `normalizeVowels` function in `phonR` package (McCloy, 2015) that employs the Glasberg and Moore (1990) formula. The ERB normalization method transforms formant values to be similar to those perceived by the auditory system in relative terms, and was chosen over other available methods because some of them require a full set of vowels for the transformation (e.g., most vowel-extrinsic methods), and other methods require reliable $F_1-F_2-F_3$ measures (e.g., Nearey-2 method) which were not possible to achieve from spontaneous speech recordings. Furthermore, previous studies comparing various vowel normalization methods have either reported that there are very little difference between vowel-intrinsic scaling methods (Adank, Smits, & van Hout, 2004; Clopper, 2009) and that, amongst these, the ERB scale is one of the most successful in reducing inter-speaker variability (Calamai, 2006).

For each speaker, a measure of F_1 range (in ERB) was derived by subtracting F_1 [i:] from F_1 [æ], giving an indication of how much vowels were differentiated in terms of height. The degree to which the front/back distinction was instantiated was explored by examining the F_2 range obtained by subtracting F_2 [ɔ:] from F_2 [i:]. The VSA was derived for individual speakers from F_1-F_2 values of the three vowels ([i:], [æ], [ɔ:]) separately for NB and VOC conditions. We first derived the Euclidean distance between pairs of vowels. Heron's formula was then used to calculate VSA across the three vowels (see Sapir, Ramig, Spielman, & Fox, 2010).

Articulation rate measurements were derived from the number of syllables per overall duration of speech regions for each speaker. Speech regions were automatically extracted using Praat scripts and excluded all non-speech segments (such as laughter, coughing, and singing) and silent intervals (pauses) exceeding 500 ms. Syllable counts were obtained using the `qdap` package for R (Rinker, 2013).

2.2. Results

2.2.1. Statistical analyses

Analyses were based on linear mixed-effects modelling using the `lme` function in the `nlme` package for R (R Development Core Team, 2007). We first compared the child VSA to the adult VSA by grouping the participants into four age bands (9–10 years old,

Table 1

Mean first and second formant (*F1* and *F2*) values in Hertz (mean, standard deviation) for the 'no barrier' (NB) and communication barrier (VOC) conditions for the three groups of children and for boys and girls separately.

Age band	9–10 Boys <i>N</i> =16		9–10 Girls <i>N</i> =14		11–12 Boys <i>N</i> =8		11–12 girls <i>N</i> =16		13–14 Boys <i>N</i> =22		13–14 Girls <i>N</i> =20	
	NB	VOC	NB	VOC	NB	VOC	NB	VOC	NB	VOC	NB	VOC
[i:] <i>F1</i>	441 (63)	439 (62)	465 (72)	464 (74)	441 (61)	442 (66)	457 (67)	463 (86)	386 (46)	372 (50)	433 (64)	427 (71)
[i:] <i>F2</i>	2676 (353)	2662 (420)	2808 (276)	2783 (342)	2672 (124)	2795 (289)	2798 (244)	2848 (232)	2460 (248)	2506 (262)	2651 (298)	2707 (262)
[ɔ:] <i>F1</i>	545 (81)	567 (106)	584 (93)	620 (107)	558 (83)	513 (67)	553 (72)	590 (98)	495 (98)	494 (90)	533 (74)	563 (89)
[ɔ:] <i>F2</i>	998 (165)	983 (126)	1092 (193)	1083 (241)	1038 (106)	919 (138)	1079 (181)	1034 (145)	968 (242)	912 (215)	1014 (167)	995 (212)
[æ] <i>F1</i>	991 (147)	1043 (154)	1084 (144)	1115 (162)	948 (113)	1010 (110)	1024 (133)	1085 (165)	838 (126)	900 (122)	945 (126)	1032 (154)
[æ] <i>F2</i>	1764 (135)	1806 (141)	1841 (166)	1852 (155)	1747 (77)	1731 (123)	1807 (137)	1865 (168)	1584 (137)	1592 (135)	1728 (143)	1754 (150)

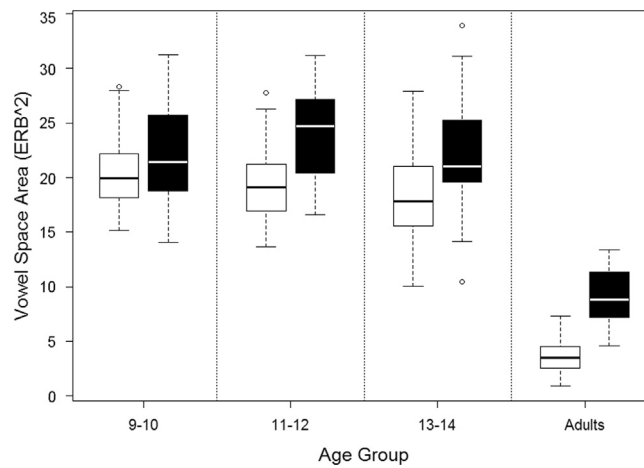


Fig. 3. Box-plots showing the size of the vowel space area (ERB²) for 'no barrier' (NB) and communication barrier (VOC) conditions in three groups of children and in adults.

11–12 years old, 13–14 years old and adults). The best-fitting model for each individual analysis was chosen with hierarchical approaches, that is, adding one predictor at a time to a baseline model that includes no predictors other than the intercept. Condition (2: NB and VOC), sex (2: Girls, Boys) and Age bands were entered one by one as fixed effects and Participant was a random effect. Likelihood ratio tests were used to determine which effects were needed in the model. Orthogonal planned contrasts were used to evaluate if the VSA differed between the three groups of children and adults. Tukey's HSD was used for any subsequent post-hoc comparisons. However, after establishing the child VSA relative to adult VSA, the subsequent analyses focused on developmental aspects of the acoustic–phonetic characteristics of conversational speech (namely VSA, *F1* and *F2* range and articulation rate) and were, therefore, conducted for the child data only by treating age (in months) as a continuous variable. The change in *F1* (i.e., [æ] *F1* minus [i:] *F1*) and *F2* ranges (i.e., [i:] *F2* minus [ɔ:] *F2*) was analysed in order to get a better sense of which dimension was primarily affecting VSA size. Again, the best-fitting model for each individual analysis (VSA, *F1*/*F2* range, articulation rate) was chosen with bottom-up hierarchical approaches by entering Condition (2: NB and VOC), sex (2: Girls, Boys) and Age (months) one by one as fixed effects and Participant as a random effect.

First we wanted to verify that the VOC condition increased task difficulty by comparing the transaction times (i.e., time in minutes to find eight differences) between the conditions. Our results showed that it took significantly longer to find eight differences in the VOC condition ($M=6.7$, $SD=2.0$) than in the NB condition ($M=4.0$, $SD=1.5$; $\chi^2(1)=95.88$, $p<.001$) validating our assumption about good and adverse listening conditions.

Fig. 2 illustrates the size of the vowel triangles in the 'no barrier' (NB) and communication barrier (VOC) conditions for child and adult speakers, separately for females and males. Group mean values were derived by calculating a median across all tokens for each vowel and each speaker and then calculating an overall average for each age band. These are presented in Hertz in Table 1, although all statistical evaluations were done using ERB values.

2.2.2. VSA in children and adults

We compared VSA size in the NB and VOC conditions between children and adults (see Fig. 3). The linear mixed effects model's type III Wald Chi square test revealed a significant main effect of Condition, ($\chi^2(1)=75.99$, $p<.001$) and Age band ($\chi^2(3)=141.77$, $p<.001$) and a significant interaction between Condition and Age band ($\chi^2(3)=16.75$, $p=.008$). The orthogonal planned contrasts for

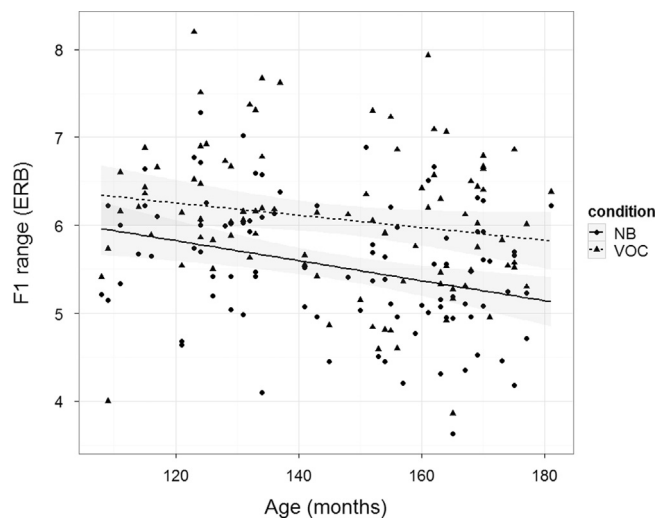


Fig. 4. Scatterplots showing the F1 range (ERB) as a function of age in children for 'no barrier' (NB) and communication barrier (VOC) conditions. Shaded areas represent 95% confidence intervals.

Table 2
Descriptive data for the main effects of Condition and Gender for F1 and F2 range (ERB) measures.

	NB	VOC	BOYS	GIRLS
F1 range	5.53 (.73) ^{***}	6.07 (.82) ^{***}	5.65 (.78) [*]	5.94 (.84) [*]
F2 range	7.68 (.86) ^{***}	8.13 (1.09) ^{***}	7.94 (1.19)	7.88 (.79)

^{**} $p < .001$.

^{*} $p < .05$.

^{***} $p < .0001$.

the three groups of children versus adults demonstrated that adults had significantly smaller VSA than all three groups of children (all comparisons, $p < .05$).

Furthermore, post-hoc tests on the Condition and Age band interaction showed that VSA were bigger for the VOC condition than for the NB condition for all four age bands (all comparisons, $p < .001$), except for the 9–10 year olds where the difference between NB and VOC was approaching significance ($p = .051$; see Fig. 3). However, as apparent from Fig. 2, not all vowels contributed to the expansion of the VSA in VOC to a similar degree in children. For example, F1 frequency increased significantly for vowel [æ] between NB and VOC conditions in all three age bands (all comparisons, $p < .007$). There was only a significant difference between NB and VOC conditions for F2 for vowel [ɔ:] in 11–12 year olds and for F2 for vowel [i:] in 13–14 year olds (both comparisons, $p < .001$).

There were no significant main effects or interaction involving sex. This suggests that the effect of the communication barrier on VSA for talker A is similar for female and male speakers across the age bands (Female: $M = 18.81$, $SD = 6.70$; Male: $M = 17.73$, $SD = 7.43$).

In order to better evaluate the development of clear speech adaptations in children, the following analyses focused on children alone and treated age as a continuous variable.

2.2.3. Development of VSA in children

Again, we compared VSA size in the NB and VOC conditions as a function of age (in months) and sex (2: Girls and Boys). The linear mixed effects model's type III Wald Chi square test revealed a significant main effect of Condition, ($\chi^2(1) = 49.36$, $p < .001$) and a significant interaction between Condition and Age ($\chi^2(1) = 4.47$, $p = .034$).

Overall, the VSA were larger in the VOC ($M = 22.43$, $SD = 4.64$) than in the NB condition ($M = 19.31$, $SD = 3.94$), indicating that children use VSA expansion as a clear speech strategy. Children reduced their VSA as a function of age in NB (Age main effect: $F(1,92) = 6.76$, $p = .011$) but not in VOC (Age main effect: $F(1,89) = .12$, $p = .727$, ns).

Again, as in the previous set of analyses, the 'communication barrier' effect on VSA was similar for girls and boys (Girls: $M = 21.13$, $SD = 4.09$; Boys: $M = 20.55$, $SD = 5.02$).

2.2.4. Development of F1 and F2 range in children

Next, we examined change in F1 range and F2 range in order to get a better sense of which dimension was primarily affecting VSA size. For F1 range, there was a significant main effect of Condition ($\chi^2(1) = 45.12$, $p < .001$), Age ($\chi^2(1) = 7.510$, $p = .006$), and Sex ($\chi^2(1) = 5.70$, $p = .017$) (see Fig. 4, and Table 2). For F2 range, there was a significant main effect of Condition ($\chi^2(1) = 16.65$, $p < .001$), and a significant interaction between Condition and Sex ($\chi^2(1) = 8.17$, $p = .004$) (see Fig. 5 and Table 2) but no significant main effect of interaction involving the factor Age (all comparisons, $p > .5$).

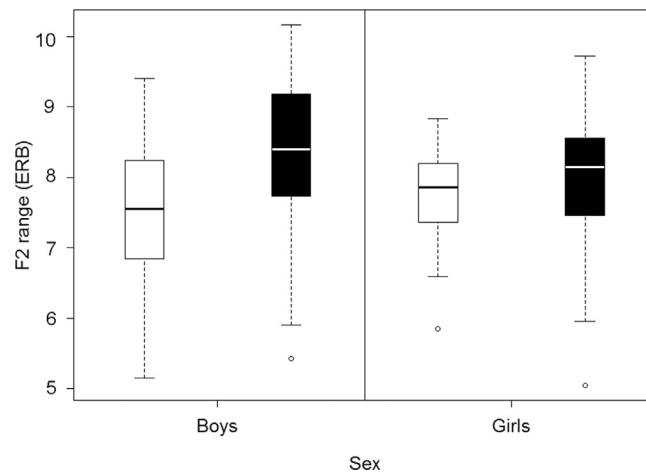


Fig. 5. Box-plots showing the $F2$ range (ERB) for 'no barrier' (NB; white) and communication barrier (VOC; black) conditions in Boys and Girls.

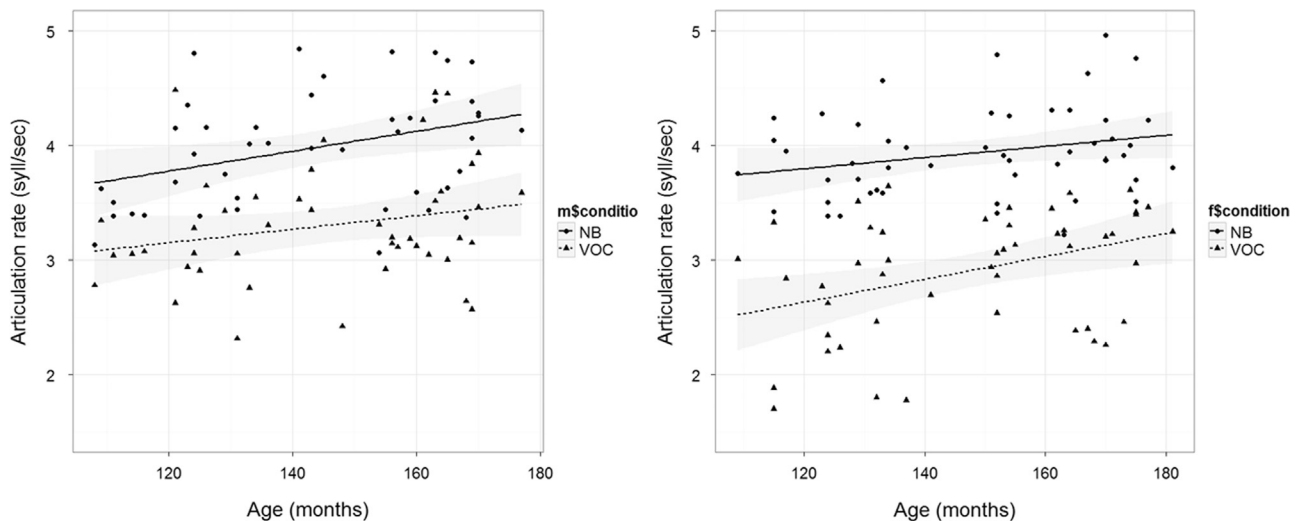


Fig. 6. Articulation rate as a function of age in 'no barrier' (NB) and vocoded (VOC) conditions for boys (left) and girls (right).

However, unlike in the VSA measure, we found that $F1$ range was significantly larger for girls than for boys. Moreover, the significant interaction between Condition and Sex in the $F2$ range showed that boys manipulated $F2$ range to clarify their speech in the VOC condition (NB versus VOC: $b = .78$, $t(44) = 4.83$, $p < .001$) whereas girls did not (NB versus VOC: $b = .16$, $t(44) = 1.15$, $p = .256$, ns; see Fig. 5).

In summary, in accordance with the VSA measure, both $F1$ and $F2$ ranges were larger for the VOC than for the NB condition. Furthermore, $F1$ range (but not $F2$ range) also reduced as a function of age. As with the VSA measure, not all vowels contributed to the increase of $F1/F2$ ranges for VOC in children. The increase in $F1$ range in VOC was due to changes in $F1$ for vowel [æ] (in both girls and boys), and for $F2$ range due to changes in vowel [ɔ:] but only in boys ($p < .001$).

Together these analyses of the acoustic-phonetic characteristics of conversational speech and clear speech adaptations in the presence of a communication barrier suggest that children have significantly larger vowel spaces than adults even up to 14 years of age, even after vowel normalization procedures were applied. In general, all children, except the youngest group of 9–10 year olds, show larger VSA in the VOC condition. Furthermore, when we analysed the VSA data as a function of age in children, we saw age-related articulatory adjustments in the NB condition rather than in the VOC condition. This could indicate that children reduce both VSA and $F1$ range in their conversational speech produced in the 'no barrier' condition, allowing them to show VSA and $F1$ range expansion effects in the condition (VOC) that required them to clarify their speech for the benefit of their interlocutor. Against our predictions, sex differences were only apparent in the $F1$ and $F2$ range measures but not in the VSA measures. In general, girls showed greater $F1$ range than boys suggesting that girls make a greater use of the open-closed dimension in vowel articulation. Boys, however, seemed to utilize the $F2$ (front-back) dimension when they needed to clarify their speech whereas girls did not. These results suggest that children are able to make adjustments in the fine phonetic detail of their speech in response to a communication barrier. However, it is also possible that differences in VSA were simply related to differences in the articulation rate. Therefore, in the next set of analyses, we investigated articulation rate in the same recordings and its relationship with VSA and $F1/F2$ range in children.

2.2.5. Development of articulation rate in children

We investigated the articulation rate in the NB and VOC conditions in children as a function of age (in months) and sex (2: Girls and Boys). The linear mixed effects model's type III Wald Chi square test revealed significant main effects of Condition, ($\chi^2(1)=123.57$, $p<.001$), Age ($\chi^2(1)=12.97$, $p<.001$) and Sex ($\chi^2(1)=11.82$, $p<.001$). We also found a significant interaction between Condition and Sex ($\chi^2(1)=7.41$, $p=.007$) and a marginally significant 3-way interaction between Condition, Age and Sex ($\chi^2(1)=4.15$, $p=.042$). Overall, children slowed down their articulation rate in the VOC condition (NB: $M=3.98$ syllables/s, $SD=.47$; VOC: $M=3.08$ syllables/s, $SD=.58$) and boys spoke faster than girls (Girls: $M=3.41$ syllables/s, $SD=.72$; Boys: $M=3.66$ syllables/s, $SD=.64$).

Overall, the articulation rate increased as a function of age for both girls and boys (both comparisons, $p=.001$). However, despite the significant 3-way interaction including the factor Age, there was no evidence of differences in age-related developmental trajectories between boys and girls (all comparisons, $p>.1$). This significant 3-way interaction is probably caused by young girls slowing down more than boys (see Fig. 6). Therefore, excluding age from the model would provide a better fit for the data. The results show that both girls and boys significantly slowed down their articulation rate in VOC (both comparisons, $p<.001$). This is evidence for a generalized use of reduction in articulation rate as a clear speech strategy, but girls reduced their articulation rate more than boys ($p<.001$).

Are the adjustments in vowel production (i.e., VSA, $F1$ and $F2$ range) related to the adjustments in articulation rate? Pearson's correlations (for VOC condition only) revealed that faster articulation rate was associated with smaller VSA ($t(183)=-4.04$, $p<.001$, $r^2=.08$) and with smaller $F1$ ranges ($t(187)=-5.53$, $p<.001$, $r^2=.14$) but not with changes in the $F2$ ranges ($t(184)=-1.07$, $p=.285$, $r^2=.01$, ns). However, while fine phonetic adjustments to vowel articulation are associated with more global adjustments such as articulation rate, their shared variance is still fairly low (14% for $F1$ range and 8% for VSA). Two further correlations were investigated. First, we examined whether faster and by implication more skilled speakers have smaller VSA when there is no communication barrier present. Second, we investigated whether faster speakers are also able to implement greater changes in VSA to hyperarticulate. Two Pearson's correlations showed neither to be the case. The lack of a significant correlation between articulation rate and VSA in the NB condition showed that faster speakers do not have smaller, more adult-like vowel space areas by default. The lack of a significant correlation between articulation rate in NB and VSA in VOC revealed that children with a faster articulation rate are not more likely to expand these to a greater degree when faced with a communication barrier (both comparisons, $p>.1$).

2.3. Discussion

Three main research questions motivated this experiment: Firstly, would unscripted, conversational speech in children show the same expanded VSA as reported in previous studies using scripted materials? It was predicted that children's conversational speech may not be typified by such enlarged vowel spaces in comparison to adults; we expected a gradual reduction in VSA with age, and the oldest children to approach adult values. Contrary to our predictions, even in the 'no barrier' speech, children's VSA were considerably larger than adults', and adult targets were not reached by age 14. Whilst previous studies reporting larger VSA in children only compared raw values (Vorperian & Kent, 2007; Flipsen & Lee, 2012), the present study applied a vowel-intrinsic scaling procedure (ERB), and still found considerably larger VSA in child participants. This strengthens the argument for more extreme vowel articulation in children. As predicted, there was a developmental trend for VSA to decrease with age, but this was only in the NB condition. Clearly, further development must be taking place after age 14. These data show that the large VSA reported in previous studies (Flipsen & Lee, 2012; Lee et al., 1999; Vorperian & Kent, 2007) are unlikely to be due to their method of elicitation, as large VSA were also obtained in spontaneous speech. Given that speech rate has often been equated with articulatory maturity (Kent & Forner, 1980; Lee et al., 1999), it is interesting to note that there was no correlation between VSA and speech rate in NB, i.e. faster speakers are not more likely to have smaller, more adult-like vowel space areas. Redford (2014) also reports a lack of correlation between articulation rate and vowel dispersion in younger children.

The second research question concerned the use of vowel hyperarticulation for overcoming the effects of a communication barrier for a conversational partner. We predicted that because of larger default VSA, children may find it difficult to implement additional hyperarticulation, especially in the younger groups. This was partly borne out: before age 11, there was no effect of condition on VSA. However, in children aged 12 and above, VSA varied across the NB and VOC conditions; further, children reduced both VSA and $F1$ distance in the NB condition as a function of age whereas, in the VOC condition, VSA did not show significant age-related changes. This could indicate that children are still learning to reduce VSA in their casual conversational speech. Redford and Gildersleeve-Neumann (2007, 2009) reported a similar process in the speech of preschoolers, and it is remarkable that the speech of adolescents may still be characterized by this mechanism. Therefore we might conclude that an expanded VSA does not prevent children from hyperarticulating vowels per se, but what they are learning to control is the use of (possibly articulatory) shortcuts for casual speech. Whereas speech rate and VSA were not related in NB, they were in VOC, in that slower speech tended to be linked to a greater VSA. In this respect, children's speech also shows adult-like processes: clear speech research has suggested a strong link between a reduction in articulation rate and vowel hyperarticulation (Krause & Braid, 2004; Moon & Lindblom, 1994).

However, not all vowels contributed to the same degree in the VSA expansion in children. We observed changes mostly in the $F1$ value of the low front vowel [æ] for all three groups of children and for boys and girls alike. In the communication barrier condition, vowel [æ] moved to lower location in the vowel space possibly indicating more jaw opening and lower tongue position when hyperarticulating to overcome the signal distortion (Ferguson & Quené, 2014). Gender differences were only apparent in $F2$ of the high back vowel [ɔ:] which shifted to a more back position in boys when hyperarticulating.

The last research question concerned the effect of sex. We were interested in charting the emergence of sex effects given that the age range straddles puberty and given that sex effects have been reported to appear for aspects of fine phonetic details over this age range. Boys and girls differed in terms of speech rate irrespective of age: girls were slower talkers overall, and they slowed down more than boys in VOC. This indicates that sociolinguistic aspects of rate are acquired before the age of nine. However, in terms of VSA, there was no effect of sex. Perhaps somewhat surprisingly, we also did not find strong effects of sex on the raw $F1$ and $F2$ Hz values, even in the older teenage speakers (see Table 1). This may reflect the fact that only a subset of the boys had reached puberty in this age range. However, sex-specific effects were found in the manipulation of the open–closed $F1$ and front–back $F2$ dimension in VOC: whilst boys mostly expanded their VSA via the front–back distinction, girls did not. On the other hand, the open–closed distinction was larger in girls irrespective of condition.

3. Experiment 2

The results from acoustic–phonetic adaptations in vowel articulation suggest that children have significantly larger VSA than adults even up to age 14. Why do we not see more marked VSA reduction in the older children as reported in some previous studies? One possibility lies in the experimental set-up for this study. The recordings were conducted in formal laboratory settings that could already elicit a form of clear or hyperarticulated speech. Moreover, during the recordings, children were placed in different rooms without sight of each other, and this lack of visual cues during the task may have also led children to hyperarticulate already in the baseline ‘no barrier’ condition.

Previous research has shown that, when engaged in a problem-solving task, adult participants use more formal language and more words and turns to complete the task in audio-only conditions than in face-to-face conditions (Boyle, Anderson, & Newlands, 1994; Doherty-Sneddon, Anderson, O’Malley, Langton, Garrod, & Bruce, 1997). Similarly, Hazan and Kim (2013) showed a trend towards less VSA expansion in an audio-only condition than in a face-to-face setting in adults. This implies that, for adult speakers, completing a task without seeing the interlocutor may be more effortful. There is some evidence that audiovisual information facilitates communication in a problem-solving task in 4–6-year-old children, although 10-year-olds’ performance on the task was found to be equal in audiovisual and visual conditions (Doherty-Sneddon & Kent, 1996; although see Doherty-Sneddon, McAuley, Bruce, Langton, Blokland, & Anderson, 2000, for a finding of negative influence of visual modality in task outcomes for children). Children’s processing of audiovisual information has been found to develop even through adolescence (e.g. Barutchu, Danaher, Crewther, Innes-Brown, Shivdasani, & Paolini, 2010; Hillock-Dunn & Wallace, 2012) – therefore it is possible that the lack of face-to-face interaction in experiment 1 may have influenced children’s speech differently to adults’, especially in the cognitively demanding diapix task. To investigate the effect of recording conditions and presence/absence of face-to-face interaction on VSA size, we compared our NB condition results with those from a linked study using the same diapix task, where the recordings were made when children were in familiar surroundings and in face-to-face interaction with their conversational partner.

3.1. Method

3.1.1. Participants

Thirteen monolingual Southern British English children with normal hearing (4M, 9F, mean age: 11;6 years, range: 9;0–14;5 years) were recorded completing the diapixUK (Baker & Hazan, 2011) task in pairs in a classroom environment. The children participated in primary and secondary schools in London and the South East, and were not reported to have any neurological, medical or learning difficulties. Both mixed and same-sex pairs were used, and as in the main study, the two conversational partners were friends. These 13 participants (‘classroom face to face condition’) were matched (on sex and age) with quasi-randomly selected 13 participants (4M, 9F, age: 11;5, range: 9;0–14;5; Age: $t(24) = .04, p = .965$) from Experiment 1 (‘laboratory audio condition’).

The full set of recordings involves each participant completing tasks in two sessions: once together with a friend with normal hearing and once with a friend who has a hearing loss². However, in this paper, only the recordings of pairs of children with normal hearing are used to compare to the NB condition in Experiment 1. Each pair of participants was recorded in a quiet room in their school using a laptop running Audacity with a sample rate of 44,100 Hz (16 bit) using a Scarlett 2i2 USB audio set-up. The participants wore Audiotechnica AT8531 lapel microphones and were sitting at a table facing each other, approximately 1–1.5 m away from each other³. Each participant was given a board with the diapix picture which stood upright between the participants – the participants could see each other’s faces and gestures but not the front of each other’s boards. Note that the participants were not given any instructions as to how to speak to each other; they were merely told that the aim of the game was to talk to each other to find the 12 differences between their two pictures. In each participant’s first session, before the start of the task, the experimenter showed the participants a pair of example diapix pictures, and pointed out the types of differences found in the pictures. As in the main experiment, the participants were also given a practice picture task to familiarize themselves with the task; the participants were given time to find a few differences in the practice picture before starting the main diapix picture task. Participants were asked to start the description of their pictures from the top left-hand corner and work clockwise, and to circle any differences found; unlike in the main experiment, they were asked to contribute equally to the conversation. The participants were given a 10-min time-limit for

² Half of the participants completed the first session with a hearing friend, and the rest completed it initially with a friend with hearing loss.

³ Additionally, a video camera was positioned behind each participant, which enabled video recordings to be made of the participant sitting opposite. The video recordings were not analysed in the current study.

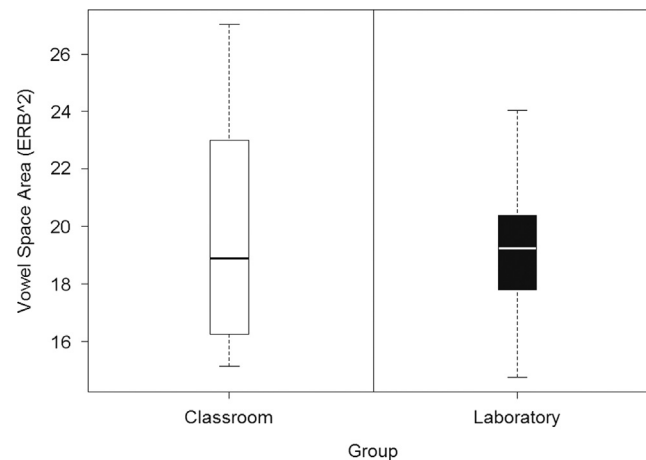


Fig. 7. Box-plots showing the size of the vowel space (ERB²) for two groups of age- and gender-matched children ('Classroom face to face' and 'Laboratory audio').

finding the differences and, depending on time limitations, were given one or two picture-scenes to complete. As part of the full procedure, the participants also completed another communicative task both before and after the diapix tasks – however, the results are not reported here.

3.1.2. Procedure and analyses

A comparison between the “classroom face to face condition” and ‘laboratory audio’ participants was done only in terms of VSA in the condition between the two participants with normal hearing conversing without a communication barrier. The recordings were transcribed in stereo format using Praat; in all other aspects, the data processing and analysis was the same as for the main experiment. The number of vowels included in the averages for vowel space area measurement did not differ between studies (Classroom face-to-face: 24 [i:], 14 [æ], 14 [ɔ:]; Laboratory audio: 25 [i:], 18 [æ], 13 [ɔ:], all comparisons, $p > .1$).

3.2. Results and discussion

Fig. 7 illustrates the size of the vowel space area in the two different recording settings (classroom face-to-face and laboratory audio). A *t*-test comparing the size of the vowel space area in age- and sex-matched children between two different recording settings suggest that large VSA already in the NB condition (when there is no communication barrier present) is not due to lack of visual cues between the two talkers $t(24) = .48$, $p = .639$ (Laboratory audio: $M = 19.12$, $SD = 2.71$; Classroom face to face: $M = 19.76$, $SD = 4.06$).

Overall, the finding that VSA did not differ between the ‘classroom face-to-face condition’ and ‘laboratory audio’ lends further support to the conclusion that large VSA are indeed part of the acoustic signature of children's spontaneous speech, as late as age 14. The ratio of children's to adults' spaces appears very large (approximately 20:6 ERB²), but it is not possible to ascertain how this compares to previous studies, since these included different dialects and elicitation methods.

Presently we can only speculate about the causes behind these large VSA. In addition to articulatory factors, developmental changes to language processing may also play a role. In adult discourse, repetitions of given or accessible referents are produced with less acoustic prominence than new or non-accessible referents (Prince, 1981; Watson, Arnold, & Tanenhaus, 2010). Second-mentioned forms are therefore also likely to be produced with less extreme vowels. The diapix task gives rise to a number of repetitions as participants attempt to determine which referents are being described. If child and adolescent participants realise given and new referents with similar levels of acoustic prominence during this particular task, this may well lead to bigger VSA than for adults (although Wonnacott and Watson (2008), using a different task, showed differentiation of given and new referents in 4 year olds). Another factor may be differences in the size of articulatory units for adults and children. Acoustic and kinematic studies have suggested that whilst adults are able to plan longer sequences, children seem to be restricted to word-level or even syllable-level articulatory plans up to the age of 16 (Nip & Green, 2013; Sadagopan & Smith, 2008). Smaller articulatory units may engender fewer compression phenomena, which in turn may leave space for more extreme articulatory targets. Furthermore, if children's speech also contains more and smaller prosodic chunks than adult's speech (Sabin, Clemmer, O'Connell, & Kowal, 1979), a greater number of their vowels may have received sentential accent, which entails articulation at the periphery of the vowel space (de Jong, 1992).

4. Conclusions

To our knowledge this is the first study to examine the development of VSA using unscripted spontaneous speech. We found that, as late as 14 years, children's VSA were significantly larger than adults'. In accord with previous studies of clear speech, we found that, from age 11, vowel hyperarticulation was evident in response to a communication barrier. Interestingly, the manner in which this was achieved differed from adults: children did not learn to expand the VSA to overcome the barrier, rather they learnt to have a more compact area in what might be deemed casual speech. This difference in the manner in which hyperarticulation was achieved only

became apparent when considering the developmental trajectories of VSA in the two conditions. Redford (Redford, 2014; Redford & Gildersleeve-Neumann, 2009) has suggested that the Hyper–Hypo approach (Lindblom, 1990) may not be directly applicable to child speech, in that rather than struggling to reach full adult targets, children may instead need to learn the efficient shortcuts characteristic of mature speech, i.e. to hypoarticulate when appropriate.

Acknowledgements

This work was supported by the Economic and Social Research Council [Grant no. RES-062-23-3106]. The full dataset is accessible at the following website: <http://reshare.ukdataservice.ac.uk/851525/>. We thank four anonymous reviewers and the associate editor for their helpful comments.

References

- Adank, P., Smits, R., & van Hout, R. (2004). A comparison of vowel normalization procedures for language variation research. *Journal of the Acoustical Society of America*, 116(5), 3099–3107.
- Baker, R. C., & Hazan, V. (2011). DiapixUK: A task for the elicitation of spontaneous speech dialogs. *Behavioural Research Methods*, 43(3), 761–770.
- Barutchu, A., Danaher, J., Crewther, S. G., Innes-Brown, H., Shivasani, M. N., & Paolini, A. G. (2010). Audiovisual integration in noise by children and adults. *Journal of Experimental Child Psychology*, 105(1–2), 38–50.
- Bent, T., Buchwald, A., & Pisoni, D. B. (2009). Perceptual adaptation and intelligibility of multiple talkers for two types of degraded speech. *Journal of the Acoustical Society of America*, 126(5), 2660–2669.
- Boyle, E. A., Anderson, A. H., & Newlands, A. (1994). The effects of visibility on dialogue and performance in a cooperative problem solving task. *Language and Speech*, 37(1), 1–20.
- Bradlow, A. R., Torretta, G. M., & Pisoni, D. B. (1996). Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication*, 20(3), 255–272.
- Busby, P. A., & Plant, G. L. (1995). Formant frequency values of vowels produced by preadolescent boys and girls. *Journal of the Acoustical Society of America*, 97(4), 2603–2606.
- Calamai, S. (2006). Intrinsic methods in vowel normalization: Data from Pisa and Florence Italian. In E. Miyares Bermúdez, & L. Ruiz Miyares (Eds.), *Linguistics in the twenty first century*, 2006 (pp. 113–124). Newcastle: Cambridge Scholars Press.
- Clopper, C. (2009). Computational methods for normalization acoustic vowel data for talker differences. *Language and Linguistics Compass*, 3(6), 1430–1442.
- Cooke, M., King, S., Garnier, M., & Aubanel, V. (2014). The listening talker: A review of human and algorithmic context-induced modifications of speech. *Computer Speech and Language*, 28(2), 543–571.
- Davis, M. H., Johnsruide, I. S., Hervais-Adelman, A. G., Taylor, K. J., & McGettigan, C. (2005). Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise-vocoded sentences. *Journal of Experimental Psychology*, 134(2), 222–241.
- Doherty-Sneddon, G., Anderson, A., O'Malley, C., Langton, S., Garrod, S., & Bruce, V. (1997). Face-to-face and video-mediated communication: A comparison of dialogue structure and task performance. *Journal of Experimental Psychology: Applied*, 3(2), 105–125.
- Doherty-Sneddon, G., & Kent, G. (1996). Visual signals in children's communication. *Journal of Child Psychology and Psychiatry*, 37(8), 949–959.
- Doherty-Sneddon, G., McAuley, S., Bruce, V., Langton, S., Blokland, A., & Anderson, A. H. (2000). Visual signals and children's communication: Negative effects on task outcome. *British Journal of Developmental Psychology*, 18(4), 595–608.
- de Jong, K. J. (1992). The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *Journal of the Acoustical Society of America*, 97(1), 491–504.
- Donegan, P. (2002). Normal vowel development. In M. Ball, & F. E. Gibbon (Eds.), *Vowel disorders* (pp. 1–35). Boston, MA: Butterworth Heinemann.
- Ecuchi, S., & Hirsh, I. J. (1969). Development of speech sounds in children. *Acta Oto-Laryngologica, Supplementum*, 257, 1–51.
- Fant, G. (1975). A note on vocal tract size factors and non-uniform F-pattern scalings. *Speech Transmission Laboratory Quarterly Progress & Status Reports*, 4, 22–30.
- Faulkner, A., Rosen, S., & Green, T. (2012). Comparing live to recorded speech in training the perception of spectrally shifted noise-vocoded speech. *The Journal of the Acoustical Society of America*, 132(4), EL336–EL342.
- Ferguson, S. H. (2004). Talker differences in clear and conversational speech: Vowel intelligibility for normal-hearing listeners. *Journal of the Acoustical Society of America*, 116(4), 2365–2373.
- Ferguson, S. H., & Kewley-Port, D. (2007). Talker differences in clear and conversational speech: Acoustic characteristics of vowels. *Journal of Speech, Language, and Hearing Research*, 50, 1241–1255.
- Ferguson, S. H., & Quené, H. (2014). Acoustic correlates of vowel intelligibility in clear and conversational speech for young normal-hearing and elderly hearing-impaired listeners. *Journal of the Acoustical Society of America*, 135(6), 3570–3584.
- Fitch, W. T., & Giedd, J. (1999). Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *Journal of the Acoustical Society of America*, 106(3), 1511–1522.
- Flipsen, P., Jr., & Lee, S. (2012). Research note: Reference data for the American English acoustic vowel space. *Clinical Linguistics and Phonetics*, 26(11–12), 926–933.
- Foulkes, P., & Docherty, G. (2006). The social life of phonetics and phonology. *Journal of Phonetics*, 34(4), 409–438.
- Glasberg, B. R., & Moore, B. C. (1990). Derivation of auditory filter shapes from notched-noise data. *Hearing Research*, 47(1–2), 103–138.
- Goldfield, E. C., Kay, B. A., & Warren, W. H. (1993). Infant bouncing: The assembly and tuning of action systems. *Child Development*, 64(4), 1128–1142.
- Green, J. R., Moore, C. A., & Reilly, K. U. (2002). The sequential development of jaw and lip control in speech. *Journal of Speech, Language, and Hearing Research*, 45(1), 66–79.
- Green, J. R., & Nip, I. S. (2010). Some organization principles in early speech development. In B. Maassen, & P. H. H. M. van Lieshout (Eds.), *Speech motor control: New developments in basic and applied research* (pp. 171–188). NC: Oxford University Press.
- Hazan, V., & Kim, J., (2013). Acoustic and visual adaptations in speech produced to counter adverse listening conditions. In *Proceedings of Auditory-Visual Speech Processing 2013*. (pp. 93–98). Inria: Rocquencourt, France.
- Hazan, V., & Markham, D. (2004). Acoustic-phonetic correlates of talker intelligibility for adults and children. *Journal of the Acoustical Society of America*, 116(5), 3108–3118.
- Hazan, V., & Baker, R. (2011). Acoustic-phonetic characteristics of speech produced with communicative intent to counter adverse listening conditions. *Journal of the Acoustical Society of America*, 130(4), 2139–2152.
- Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America*, 97(5), 3099–3111.
- Hillock-Dunn, A., & Wallace, M. T. (2012). Developmental changes in the multisensory temporal binding window persist into adolescence. *Developmental Science*, 15(5), 688–696.
- Holliday, J. J., Beckman, M. E., & Mays, C. (2010). Did you say susi or shushi? Measuring the emergence of robust fricative contrasts in English- and Japanese-acquiring children. *Proceedings of Interspeech 2010*, 1886–1889.
- HTK Team (2012). The Hidden Markov Model Toolkit. URL: (<http://htk.eng.cam.ac.uk/>).
- Jacewicz, E., Fox, R. A., & Wei, L. (2010). Between-speaker and within-speaker variation in speech tempo of American English. *Journal of the Acoustical Society of America*, 128(2), 839–850.
- Jeannerod, M. (1988). *The neural and behavioural organization of goal-directed movements*. Oxford: Oxford University Press.
- Kent, R. D., & Vorperian, H. K. (1995). *Development of the craniofacial-oral-laryngeal anatomy*. San Diego, CA: Singular Publishing Group Inc.
- Kent, R. D., & Forner, L. L. (1980). Speech segment durations in sentence recitations by children and adults. *Journal of Phonetics*, 8(2), 157–168.
- Krause, J. C., & Braid, L. D. (2004). Acoustic properties of naturally produced clear speech at normal speaking rates. *Journal of the Acoustical Society of America*, 115(1), 362–378.
- Labov, W. (1986). Sources of inherent variation in the speech process. In J. S. Perkell, & D. H. Klatt (Eds.), *Invariance and variability in the speech processes* (pp. 402–425). New Jersey: Lawrence Erlbaum Associates.
- Labov, W. (1994). *Internal factors. Principles of linguistic changes*, Vol. 1. Oxford: Blackwell.
- Lam, J., & Tjaden, K. (2013). Intelligibility of clear speech: Effect of instruction. *Journal of Speech, Language, and Hearing Research*, 56(5), 1429–1440.
- Lee, S., Potamianos, A., & Narayanan, S. (1999). Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *Journal of the Acoustical Society of America*, 105(3), 1455–1468.

- Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In W. J. Hardcastle, & A. Marchal (Eds.), *Speech production and speech modelling* (pp. 403–439). The Netherlands: Kluwer Academic.
- Logan, K. J., Byrd, C. T., Mazzocchi, E. M., & Gillam, R. B. (2011). Speaking rate characteristics of elementary-school-aged children who do and do not stutter. *Journal of Communication Disorders, 44*(1), 130–147.
- McCloy, Daniel (2015). "phonR: Tools for phoneticians and phonologists." R package version 1.0-0. (<http://drammock.github.io/phonR/>).
- Moon, S. J., & Lindblom, B. (1994). Interaction between duration, context, and speaking style in English stressed vowels. *Journal of the Acoustical Society of America, 96*(1), 40–55.
- Nip, I. S., & Green, J. R. (2013). Increases in cognitive and linguistic processing primarily account for increases in speaking rate with age. *Child Development, 84*(4), 1324–1337.
- Perry, T. L., Ohde, R. N., & Ashmead, D. H. (2001). The acoustic bases for gender identification from children's voices. *Journal of the Acoustical Society of America, 109*(6), 2988–2998.
- Prince, E. (1981). Toward a taxonomy of given-new information. In P. Cole (Ed.), *Radical pragmatics* (pp. 223–255). New York, NY: Academic Press.
- Quené, H. (2008). Multilevel modeling of between-speaker and within-speaker variation in spontaneous speech tempo. *Journal of the Acoustical Society of America, 123*(2), 1104–1113. R development core team (2007). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, <http://www.R-project.org>.
- Redford, M. A. (2014). The perceived clarity of children's speech varies as a function of their default speech rate. *Journal of the Acoustical Society of America, 135*(5), 2952–2963.
- Redford, M.A. & Gildersleeve-Neumann, C.E. (2007). Children's clear speech suggests word-level targets: Preliminary evidence. In J. Trouvain, W.J. Barry (Eds.), *Proceedings from the 16th International Congress of Phonetic Sciences*, pp. 365–368.
- Redford, M. A., & Gildersleeve-Neumann, C. E. (2009). The development of distinct speaking styles in preschool children. *Journal of Speech, Language, and Hearing Research, 52*(6), 1434–1448.
- Rinker, T. W. (2013). *qdap: Quantitative Discourse Analysis Package* version 1.3.1. Buffalo, New York: University at Buffalo. (<http://github.com/trinker/qdap>).
- Romeo, R., Hazan, V., & Pettinato, M. (2013). Developmental and gender-related trends of intra-talker variability in consonant production. *Journal of the Acoustical Society of America, 134*(5), 3781–3792.
- Rosen, S., Faulkner, A., & Wilkinson, L. (1999). Adaptation by normal listeners to upward spectral shifts of speech: Implications for cochlear implants. *Journal of the Acoustical Society of America, 106*(6), 3629–3636.
- Sabin, E., Clemmer, E., O'Connell, D., & Kowal, S. (1979). A pausal approach to speech development. In A. Siegman, & S. Feldstein (Eds.), *Of speech and time: Temporal speech patterns in interpersonal contexts* (pp. 35–55). Hillsdale, NJ: Lawrence Erlbaum.
- Sadagopan, N., & Smith, A. (2008). Developmental changes in the effects of utterance length and complexity on speech movement variability. *Journal of Speech, Language and Hearing Research, 51*(5), 1138–1151.
- Sapir, S., Ramig, L. O., Spielman, J. L., & Fox, C. (2010). Formant Centralization Ratio (FCR): A proposal for a new acoustic measure of dysarthric speech. *Journal of Speech, Language, and Hearing Research, 53*(1), 114.
- Scarborough, R., Brenier, J., Zhao, Y., Hall-Lew, L., & Dmitrieva, O. (2007). An acoustic study of real and imagined foreigner-directed speech. *Proceedings of the 16th International Congress of Phonetic Sciences, Saarbrücken*, 2165–2168.
- Simpson, A. P. (2009). Phonetic differences between male and female speech. *Language and Linguistics Compass, 3*(2), 621–640.
- Smiljanic, R., & Bradlow, A. R. (2009). Speaking and hearing clearly: Talker and listener factors in speaking style changes. *Linguistics and Language Compass, 3*(1), 236–264.
- Smiljanic, R., & Bradlow, A. R. (2008). Stability of temporal contrasts across speaking styles in English and Croatian. *Journal of Phonetics, 36*(1), 91–113.
- Smiljanic, R., & Bradlow, A. R. (2005). Production and perception of clear speech in Croatian and English. *Journal of the Acoustical Society of America, 118*(3), 1677–1688.
- Stoel-Gammon, C., & Herrington, P. B. (1990). Vowel systems of normally developing and phonologically disordered children. *Clinical Linguistics and Phonetics, 4*(2), 145–160.
- Strange, W., Weber, A., Levy, E. S., Shafiro, V., Hisagi, M., & Nishi, K. (2007). Acoustic variability within and across German, French, and American English vowels: Phonetic context effects. *The Journal of the Acoustical Society of America, 122*(2), 1111–1129.
- Sturm, J. A., & Seery, C. H. (2007). Speech and articulatory rates of school-age children in conversation and narrative contexts. *Language, Speech, and Hearing Services in Schools, 38*(1), 47–59.
- Sundberg, U. (2001). Consonant specification in infant-directed speech. Some preliminary results from a study of Voice Onset Time in speech to one-year-olds. *Lund Working Papers in Linguistics, 148–151*.
- Synnestvedt, A., Bernstein-Ratner, N., & Newman, R. (2010). Voice onset time in infant-directed speech at 7.5 and 11 months. *Journal of the Acoustical Society of America, 127*(3), 1853.
- Syrett, K., & Kawahara, S. (2013). Production and perception of listener-oriented clear speech in child language. *Journal of Child Language, 14*(6), 1–17.
- Uchanski, R. M. (2005). Clear speech. In D. B. Pisoni, & R. Remez (Eds.), *The handbook of speech perception* (pp. 207–235). Malden, MA/Oxford, UK: Blackwell.
- Van Engen, K. J., Baese-Berk, M., Baker, R. E., Choi, A., Kim, M., & Bradlow, A. R. (2010). The wildcat corpus of native and foreign-accented English: Communicative efficiency across conversational dyads with varying language alignment profiles. *Language and Speech, 53*(4), 510–540.
- Vorperian, H. K., & Kent, R. D. (2007). Vowel acoustic space development in children: A synthesis of acoustic and anatomic data. *Journal of Speech, Language, and Hearing Research, 50*(6), 1510–1545.
- Walsh, B., & Smith, A. (2002). Articulatory movements in adolescents: Evidence for protracted development of speech motor control processes. *Journal of Speech, Language, and Hearing Research, 45*(6), 1119–1133.
- Watson, D., Arnold, J. E., & Tanenhaus, M. K. (2010). Tic Tac TOE: Effects of predictability and importance on acoustic prominence in language production. *Cognition, 114*(3), 462–463.
- Whiteside, S. P. (2001). Sex-specific fundamental and formant frequency patterns in a cross-sectional study. *Journal of the Acoustical Society of America, 110*(1), 464–478.
- Whiteside, S. P., & Hodgson, C. (2000). Speech patterns of children and adults elicited via a picture-naming task: An acoustic study. *Speech Communication, 32*(4), 267–285.
- Wonnacott, E., & Watson, D. (2008). Acoustic emphasis in four year olds. *Cognition, 107*(3), 1092–1101.