Suprasegmental Characteristics of Spontaneous Speech produced in Good and Challenging Communicative Conditions by Talkers aged 9 to 14 years old.[1]

Valerie Hazan[1], Outi Tuomainen[1,] Michèle Pettinato[1,2]
Department of Speech, Hearing and Phonetic Sciences, University College London (UCL), Chandler House, 2 Wakefield Street, London WC1N 1PF, UK
[2] Department of Linguistics, University of Antwerp, Prinsstraat 13, B-2000 Antwerpen, Belgium

Abstract

Purpose: This study investigated the acoustic characteristics of spontaneous speech by 9-14 year olds, and their ability to adapt these characteristics to maintain effective communication when intelligibility was artificially degraded for their interlocutor.

Methods: Recordings were made for 96 children (50 F, 46 M) engaged in a problem-solving task with a same-sex friend; recordings for 20 adults were used as reference. The task was carried out in good listening conditions (normal transmission - NORM) and in a degraded transmission condition (DEG). Articulation rate, median fundamental frequency (f0), f0 range and relative energy in 1-3 kHz range were analyzed.

Results: With increasing age, children significantly reduced their median f0 and f0 range, became faster talkers and reduced their mid-frequency energy in spontaneous speech. Children produced similar clear speech adaptations (in DEG) as adults but only children aged 11-14 increased f0 range, an unhelpful strategy as not transmitted via the vocoder. Changes made by children were consistent with a general increase in vocal effort.

Conclusion: Further developments in speech production take place during later childhood. Children use clear speech strategies to benefit an interlocutor facing intelligibility problems but may not be able to attune these strategies to the same degree as adults.

Keywords: Children, Development, Speech production, Communication

1 Introduction

An important aspect of speech communication is the ability to adapt one's speech, both at the acoustic-phonetic and linguistic levels, to promote understanding whatever the situation in which communication is taking place. Barriers to effective communication can be acoustic, such as the presence of background noise, reverberation or spectral degradation, or can be linguistic, such as the lack of a shared native language or of a similar degree of language

[1] This paper was accepted for publication in the Journal of Speech Hearing and Language Research (http://jslhr.pubs.asha.org/) on 24/2/2016

knowledge. Adults are adept at making appropriate adaptations to their speech to counteract intelligibility-challenging conditions through the use of 'clear' speaking styles (for a review, see Smiljanic & Bradlow, 2009). This research investigates the acoustic characteristics of speech produced in good and challenging communicative conditions by talkers aged 9 to 14 years. In this paper, we focus on the global strategies that young talkers use to clarify their speech in intelligibility-challenging communicative conditions; as there is little data on late stages of speech development for spontaneous speech materials, a subsidiary aim was to investigate age effects in global characteristics of spontaneous speech produced in an interactive task within this age range. Changes that occur in the vowel space of the same group of young talkers in intelligibility-challenging conditions are reported in Pettinato, Tuomainen, Granlund and Hazan (2016).

The focus of the present study is on children's ability to make adaptations to their speech production to meet the specific demands of communicative interactions. Research has shown that the fine adjustments we make to our speech are an important aspect of communication and guide the listener in recognizing speech (e.g., Local, 2003; Hawkins, 2003). Learning to adapt your voice to different conditions is therefore an integral part of speech development. Moreover, it is important to examine this ability in children and adolescents because most of their daily communications, both at school and at home, take place in noisy environments. In spite of this, few studies have examined how children adjust their speech in response to different types of interference. Adults are known to adopt a range of speaking styles that show variations according to the needs of the interlocutor (e.g. infant- and child-directed speech, foreigner-directed speech) or of the environment (e.g. Lombard speech produced to counter the effects of noise of different types of noise, see Cooke et al., 2014, for a review). Adults also appear to modulate the strategies they use depending on the type of interference experienced by their interlocutor even if not directly exposed to this interference (Hazan & Baker, 2011). Sentences produced in 'conversational' and 'clear' styles in a laboratory setting vary across a range of global acoustic-phonetic characteristics with, at the suprasegmental level, 'clear' sentences typically showing reductions in articulation rate and increases in pause frequency and duration, in speech intensity, in mean fundamental frequency (f0) and in f0 range (for a review of the literature on clear speech strategies, see Smiljanic & Bradlow, 2009). Some of these adaptations may be at least partly linked to increased vocal effort when producing clear speech (Garnier et al., 2008).

Children are likely to have greater difficulty communicating in challenging conditions for a number of reasons linked to their abilities to both perceive and produce speech. First, it is known that their perception of speech is more greatly affected than for adults by the presence of noise, reverberation or spectral degradation until early adolescence (e.g., Johnson, 2000). Also, it is increasingly accepted that some aspects of speech production continue to develop well into the second decade of life (e.g. Walsh & Smith, 2002; Jacewicz, Fox & Wei, 2010). Some of these later developments are linked to the significant physical changes that occur in this period, both to the larynx and to the vocal tract (Fitch & Giedd, 1999). There are also ongoing changes to articulatory motor control well into adolescence, with greater trajectory variability and both slower and smaller displacements in teenagers relative to young adults (Walsh & Smith, 2002; Kleinow & Smith, 2006; Sadagopan & Smith, 2008). Some of these differences between adolescent and adult speech production may therefore be linked to ongoing maturation in neural systems controlling language production (Walsh & Smith, 2002) and to adolescents learning to develop motor schemas for different styles of speech (Whiteside, Dobbin & Henry, 2003). Further developments in children's speech in the second decade of life are also linked to the increasing influence of sociophonetic factors (e.g. Jacewicz, Fox & Salmons, 2011; Spencer, Clegg & Stackhouse, 2013).

The focus of analyses in this study is on suprasegmental measures of speech (articulation rate, fundamental frequency characteristics and intensity). The development of conversational articulation rate in children, as measured from tasks such as story retelling or monologues on familiar topics, is of particular interest because it reflects the joint influence of two components: speech motor articulation and linguistic formulation (Flipsen, 2002; Nip & Green, 2013). For example, adults and older children shorten the duration of a word sequence when embedded within a complex structure rather than when presented as a simple sentence, an effect not shown by younger (5 year old) children (Sadagopan & Smith, 2008). Developmental studies of conversational articulation rate typically show evidence of age effects throughout the first and into the second decade of life (e.g. Sturm & Steery, 2007; Flipsen, 2002). In fact, articulation rate has been shown to increase with age until mid-adulthood, with young adults showing slower rates than 35-40 year olds (Jacewicz et al., 2010) as well as adolescents showing slower rates than young adults (Jacewicz et al., 2010; Walsh & Smith, 2002). In a study involving both children and adults, articulation rates have also been shown to be influenced by processing complexity: decreases in articulation rates were seen for more demanding tasks such as story retelling (Nip & Green, 2013). Nip and Green (2013) suggested that, in development, increases in articulation rate are achieved through increases in both motor control and language processing, with less support for a strong contribution of biologically-driven aspects such as speed of movement.

Fundamental frequency characteristics also show substantial change within this age range, primarily linked to physiological changes that are occurring in puberty. Before puberty, between the ages of 7 to 11 years approximately, cross-sectional (Hacki & Heitmueller, 1999) and longitudinal (Bennett, 1983) studies have shown evidence of very gradual reductions in f0 with age. Sex differences typically appear from the age of 10-12 years (Lee, Potamianos & Narayanan 1999), with significant changes within the 12-16 year range due to physical changes to the larynx in puberty in male speakers especially (Hollien, Green & Massey, 1994). Fundamental frequency reduces during adolescence in girls also, but the change is more gradual and less extreme (Lee et al., 1999). There is less available data on f0 range. A study of f0 variation in children aged 6 to 8 years failed to find evidence of an age trend within this range (Gelfer & Denor, 2014); however with a broader age range, of girls aged 8 to 19, a significant reduction in f0 range with age was obtained in continuous speech (Pedersen et al., 1990).

Given these ongoing developmental changes in late childhood, are children able to make the skilled adaptations needed in clear speaking styles? There is evidence that young children can adopt clear speech characteristics: children as young as 3 years old are able to switch perspective and speak differently to a young child than to an adult for example (e.g. Shatz & Gelman, 1977). Children adopt characteristics of child-directed speech such as increased f0 when speaking to babies (Weppelman et al., 2003) and children aged five also produce Lombard speech when speaking in noise (Amazi & Garner, 1982). The few studies carried out to date that involve extensive acoustic analyses have focused on early childhood. When 3 to 5 year olds were asked to name pictures 'using their big girl/boy voice' , vowel formant measures showed that they aimed for more adult-like production in clear than in casual speech, but with little evidence of hyperarticulation (Redford & Gildersleeve-Neumann, 2009). Within the same age range, and when naturally eliciting a clear speaking style in children using an interactive task, adaptations in the clear speaking style included changes in duration, intensity, and fundamental frequency characteristics with a higher f0 and broader f0 range (Syrett & Kawahara, 2014). These studies do show that children in pre-school years already show some ability to adapt and clarify their speech when instructed to do so, even if instructions were indirect. What these studies do not address is how children adapt their speech in spontaneous speech in the course of more complex interactions with an interlocutor, such as those which occur when jointly engaged in a problem-solving task. Also, in these studies, no comparisons were made to the adaptations that would be made by adults in the same communicative situations.

Two aspects of our study are therefore novel and of particular interest. First, we are using an experimental design that naturally elicits clear speaking styles in real communication with another talker, rather than one in which children are instructed to speak clearly. In the 'degraded transmission' (DEG) condition reported in our study, the talker whose speech is being analyzed is hearing normally but interacting with a conversational partner who is hearing speech that is highly spectrally degraded by a vocoder; both talkers had a short (10 minute) period of training with vocoded speech prior to the task. The adaptations produced by the talker that we are analyzing are therefore purely for the benefit of the impaired interlocutor and necessary for both talkers to be able to successfully complete a task. Even though children have been shown to be able to produce clear speaking styles at a pre-school stage, when directly or indirectly instructed to do so, the data collected in this condition will give a more ecologically-valid account of the adaptations they make in challenging communicative conditions. These adaptations reflect both the ability and willingness to make clear speech adaptations for the benefit of the interlocutor, given that these require increased speaker effort. The time taken to complete the task in NORM and DEG conditions (task transaction time) is also informative as a measure of relative communicative difficulty across conditions (Van Engen et al., 2010; Hazan and Baker, 2011). Second, the analyses of the spontaneous speech produced in good listening conditions provide normative data for the 9 to 14 year age range for spontaneous speech produced with communicative intent; these analyses also enable us to examine effects of talker sex and developmental trends, if any, within this age range. The studies reported above on articulation rate and f0 characteristics in the later stages of development were based on read speech or highly-constrained spontaneous speech.

The main research questions are as follows:

1. *How do children and adolescents adapt their speech so that they can maintain good communication in challenging listening environments? Are there differences between how adults and children/adolescents achieve this?*

Given that children aged 3 to 5 years make adaptations when asked to speak clearly, we expect older children to be able to adapt their speaking style when instructed. However, in this task, they are required to infer what clarifications are most useful in counteracting the effects of a noise-excited vocoder without direct exposure to the degraded speech. Given that Nip and Green (2013) found effects of task complexity on characteristics of speech production in children, we predict that older children may be less adept than adults at making task-appropriate adaptations. An effect of condition is still expected, but may be evident also for median f0 and f0 range that were not altered by adults when countering the effects of a noise-excited vocoder (Hazan & Baker, 2011). The global measures that are the focus of the analysis are measures of articulation rate, median f0, f0 range, and mean energy in the 1-3 kHz frequency range (ME1-3kHz). These have been chosen because they were shown to vary between 'casual' and instructed clear speech in children aged 3 to 5 (Syrett & Kawahara, 2014) and are also measures that consistently show adaptations in adult studies of clear speaking styles.

2. Is there a developmental trend in the acoustic-phonetic characteristics of spontaneous speech produced by young people aged 9 to 14 years? Does it differ from spontaneous speech produced by adults?

Given previous studies of articulation rate and f0 characteristics, we expect to see an ongoing developmental trend (i.e. effect of age) in these measures throughout this age range. Given the previous literature, we predict that an effect of talker sex for f0 characteristics will appear from age 11-12 years. Trends for energy are harder to predict given the lack of evidence for this acoustic feature in children.

2 Method

2.1 Participants

Fifty single-sex pairs of children between the ages of 9 and 14 years inclusive were recruited for this study. Children within each pair 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, M=11;8 years, range 9;0 to 15;0 years). Recordings from 20 adults (11M, 9F, M=22;04 years, range 18;0-29;0 years) randomly selected from the LUCID corpus (Baker & Hazan, 2011; Hazan & Baker, 2011), using the same task and recorded in the same conditions, were used as adult reference. Child and adult participants were native Southern British English speakers who reported no history of hearing or language impairments and were not brought up bilingually. All 96 children are schooled in Greater London 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 additional language at the time of the recordings. All participants passed a hearing screen at 25 dB HL or better at octave frequencies between 250 and 8000 Hz in both ears. Ethical approval was obtained from the University College London (UCL) Research Ethics Committee and informed written consent was obtained from the parent/guardian of each child.

2.2 Procedure

During the recording sessions, two participants sat in different rooms and communicated via headsets fitted with a condenser cardioid microphone (Beyerdynamic DT297) whilst playing an interactive 'spot the difference' picture game (diapix: Van Engen et al., 2010). The diapixUK set of picture pairs were used (Baker & Hazan, 2011). The speech of each participant was recorded on a separate channel at a sampling rate of 44100 Hz (16 bit) using an EMU 0404 USB audio interface and Adobe Audition software. For all conditions, one child was designated the 'leader' (Talker A) and instructed to do most of the talking, whereas the other child (Talker B) was mainly required to ask questions and make suggestions. To familiarize 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 sat so that they could not see their friend's picture. They were told the pictures contained 12 differences which they had to find. With the younger age group, the experimenter gave hints to some of the more reticent participants during the training phase. 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 each role in separate recordings, and that they had 10 minutes to find the differences. They were told 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.

The diapix task was carried out in different conditions varying in the ease of communication for Talker B. In the 'Normal transmission' (NORM) condition[2], the two participants could

[2] Note that in previous publications about the adult (e.g. Hazan & Baker, 2011) and child corpus (Pettinato et al., 2016), the 'Normal transmission' (NORM) condition is referred to as the 'no barrier' (NB) condition, and the 'degraded transmission' (DEG) condition is referred to as the 'Vocoder' (VOC) condition.

hear each other without difficulty. In the 'degraded transmission' condition (DEG), the voice of one of the speakers ('Talker A') was distorted via a three-channel noise-excited vocoder before being channeled in real time to Talker B. The noise vocoder presented speech as three bands of noise using methods described by Shannon et al. (1995): the time-varying speech spectrum over the 100-5000 Hz frequency range was represented by measuring the amplitude envelope within each of the three frequency bands (100 to 548 Hz, 548 to 1755 Hz, 1755 to 5000 Hz) and imposing corresponding amplitude changes on bands of noise with the same frequency ranges. The vocoding was done so that Talker A, who was leading the interactions, had to clarify his or her speech to maintain good communication despite the adverse listening condition affecting their friend. The talkers were not given any instructions as to what strategies they should use or speaking style they should adopt. The talker being vocoded (Talker A) did not hear the effect of the degradation on their voice during the task, so had to rely on their own experience of the effect of vocoding (gleaned from the training task described below or from having previously done the task as Talker B) and from feedback from their interlocutor (for example resulting from miscomprehensions, requests for clarifications, backchanneling) to determine how successful their clarification strategies were. As a significant learning effect is found for vocoded speech, both participants individually completed a 10-minute vocoder training session prior to the recordings using training software described in Faulkner et al. (2012).

Every pair of participants carried out six recordings with different sets of pictures: two in the NORM condition, two in the DEG condition and two in a third condition not reported here. Every pair started out with a recording in NORM, and pairs of participants were counterbalanced between doing two DEG recordings first or the two recordings involving another condition not reported here. Everyone ended with the second NORM recording. Participants switched roles between recordings in each condition, so that every child was Talker A once in NORM and DEG. 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 in the analyses reported here so that the amount of data was comparable to that collected for the child participants. Only Talker A's speech was analyzed as this was the talker leading the interaction and having to clarify their speech for the benefit of 'impaired' interlocutor in the DEG condition.

2.3 Data processing

For all recordings, each audio channel was transcribed and aligned at word and phoneme levels (for further details, see Pettinato et al., 2016). A number of acoustic-phonetic measures were selected; these were the suprasegmental features that were also analyzed for the adult LUCID corpus (Hazan & Baker, 2011). These include measures of median f0 and f0 range, mean energy in the 1-3 kHz range of the long-term average spectrum of speech, and articulation rate. Fundamental frequency analyses were carried out using Praat software on each of the recordings for Talker A in each condition. A Praat script opened each file, extracted the intervals which were not marked as silences, laughter, noise or breath intake, and concatenated the extracted intervals. Then, on the concatenated file, f0 calculations were done using the 'pitch' function in Praat, with a time step of 100 pitch values per second. A formula as described in De Looze and Hirst (2008) was used to calculate ceiling and floor limits specific to each speaker rather than default values for males and females, as this has been shown to be more successful in excluding rogue values. For each speaker in each condition, long term distributional measures, i.e. median fundamental frequency and interquartile range (75th -25th percentile, i.e. 50% span), were calculated over the aggregated speech; semitones relative to 1 Hz were used as a measure to facilitate comparisons of pitch range across male and female talkers. Linguistically-based measures of median f0 and f0 range, such as described in Mennen, Shaeffler and Docherty (2012) would have been preferable, but would require prosodic annotation which was not available in this spontaneous speech corpus. For the energy measure, long-term average spectrum (LTAS) analyses were carried out using a Praat script. First, for each file, the intensity of all labeled speech segments was calculated and those above a set level (88 dB) were excluded for the LTAS calculations, as likely to be instances of shouting. The remaining speech segments were concatenated and the intensity of the resulting waveform scaled to a set level (75 dB). The signal was then band-passed filtered between 1 and 3 kHz and the mean intensity of the resulting waveform calculated to give a measure of the amount of energy present in this frequency range relative to the total energy in the spectrum. This measure was labelled as ‘Mean energy 1-3 kHz’ (ME1-3kHz). This frequency band was chosen because it has been shown that talkers increase their energy in this region when producing clear speech (Krause and Braida, 2004). An increase in the relative energy in this mid-frequency band also reflect a reduction of spectral tilt, documented in speech produced with vocal effort (e.g., Glave and Rietveld, 1975; Sliujter and Van Heuven, 1996). Articulation rate was calculated as the number of syllables produced by Talker A divided by the total duration of speech regions for that talker. Syllable counts were calculated from the orthographic transcriptions of the spontaneous speech using the qdap package in R (Rinker, 2013), after exclusion of segments labelled as unfinished words, hesitations, fillers and agreements (e.g. ‘yeah’, ‘yup’, ‘err’, ‘hmm’).

2.4 Statistical analysis

Statistical analyses of the data were based on linear mixed-effects modelling using the lmer function from lme4 (Bates, Maechler, Bolker & Walker, 2014) package for R (R Development Core Team, 2013). For each dependent variable (median f0, f0 range, articulation rate, ME1-3 kHz) we began with a saturated model which includes interaction terms for all as fixed effects with random intercepts and slopes (Barr, Levy, Scheepers & Tily, 2013). Due to non-convergence we simplified the models hierarchically from most complex to least complex followed by forward entry of random slopes for the fixed effects that were retained in the initial backward elimination. Resulting converged models for all four variables included fixed effects: Condition (2: NORM, DEG), Sex (2: Female, Male) and Age band (4: 9-10 years, 11-12 years, 13-14 years, adults) and Participant as random effect but no random slopes. For the age band comparisons we used treatment coding where the adult group was chosen as the reference level.

We compared model residuals via Chi-square tests (α=.05) from the most complex models (containing largest interaction term) to least complex models (containing only single terms). If an interaction term was significant we included all lower level effects involved in the interaction in the final model. Data from all four measures by Age band and Sex are shown for the NORM and DEG condition in Tables 1 and 2, and by Age band and Condition in Figure 1.

Table 1: Descriptive results for four acoustic-phonetic measures in the normal transmission (NORM) condition: Mean (SD).

	Normal transmission (NORM)							
	9-10		**11-12**		**13-14**		**Adults**	
	F (N=14)	**M (N=16)**	**F (N=16)**	**M (N=8)**	**F (N=20)**	**M (N=22)**	**F (N=9)**	**M (N=11)**
F0 median (semitone)	98.01 (2.34)	96.07 (2.55)	96.02 (1.11)	95.71 (1.78)	93.23 (2.14)	88.30 (5.23)	91.54 (1.48)	81.05 (1.48)
F0 range (semitone)	4.29 (0.76)	4.11 (0.79)	3.36 (0.85)	3.88 (1.01)	3.00 (0.79)	3.20 (0.90)	2.66 (0.54)	2.92 (0.70)
Mean energy 1-3 kHz (dB)	67.98[a] (1.70)	67.05 (2.43)	67.03[b] (1.27)	66.46 (1.58	66.92 (2.05)	65.45 (2.17)	61.15 (3.54)	59.66 (2.56)
Articulation rate (syllables/second)	3.81 (0.37)	3.73 (0.46)	3.91 (0.36)	4.25 (0.32)	4.03 (0.44)	4.17 (0.58)	4.26 (0.49)	5.03 (0.43)

[a] N=13 [b] N=15

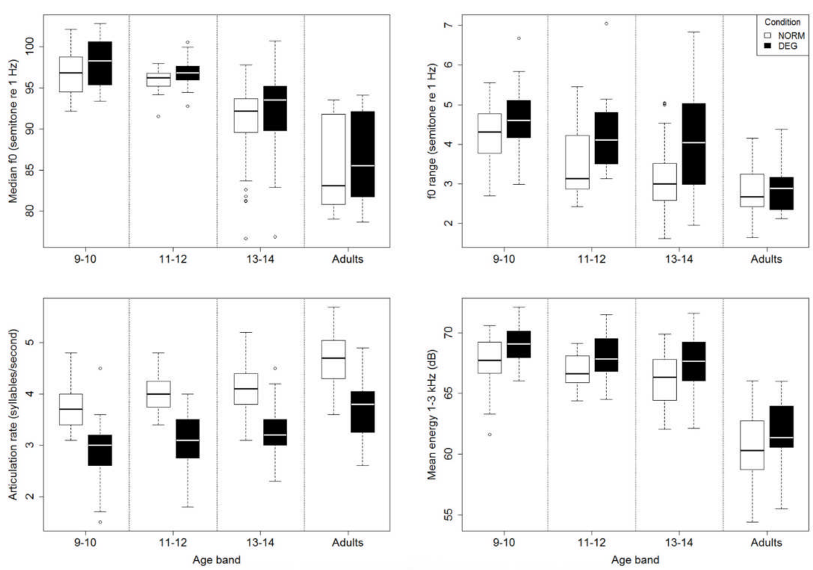

Figure 1. Box-plots showing measures of median f0, f0 range, articulation rate and mean energy in the 1-3 kHz frequency band as a function of Age band and Condition.

Table 2: Descriptive results for four acoustic-phonetic measures in the degraded transmission (DEG) condition: Mean (SD).

	Degraded transmission (DEG)							
	9-10		**11-12**		**13-14**		**Adults**	
	F (N=14)	**M (N=16)**	**F (N=16)**	**M (N=8)**	**F (N=20)**	**M (N=22)**	**F (N=9)**	**M (N=11)**
F0 median (semitone)	99.3 (2.35)	97.49 (2.73)	97.29 (1.60)	95.95 (1.43)	95.46 (2.66)	89.83 (4.98)	92.14 (1.13)	82.38 (2.37)
F0 range (semitone)	4.65 (0.57)	4.64 (0.87)	4.33 (0.93)	4.00 (0.83)	4.05 (1.41)	4.20 (1.19)	2.80 (0.68)	3.00 (0.64)
Mean energy 1-3 kHz (dB)	69.61 (1.49)	68.5 (1.44)	68.52 (1.77)	66.94 (1.13)	68.75 (1.64)	66.39 (1.98)	63.27 (2.39)	60.89 (2.28)
Articulation rate (syllables/ second)	2.56 (0.60)	3.12 (0.48)	2.91 (0.55)	3.35 (0.52)	3.10 (0.47)	3.39 (0.53)	3.40 (0.47)	3.93 (0.68)

3 Results

3.1 Effect of communication barrier on task transaction time

First, recordings for the NORM and DEG conditions were compared to ascertain whether the communication barrier condition made communication more effortful, as intended. As a measure of transaction difficulty, a calculation was made of the average time taken to find a difference in the picture when in 'Talker A' role. Data points were excluded in the few cases where less than five differences were found during the maximum task time of 10 minutes. The total number of differences (out of 12 possible differences) found in the maximum task time was also calculated.

It took significantly longer to find differences in the DEG condition than in the NORM condition for all four groups (9-10 years: 1.2 vs 0.5 minutes; 11-12 years: 1.0 vs 0.5 minutes; 13-14 years: 0.9 vs 0.5 minutes; adults: 0.8 vs 0.6 minutes; for DEG versus NORM; $t=2.61$; $p<.01$). However, the 9-10 and 11-12 years groups showed a greater difference in task transaction between DEG and NORM than the adult group ($t=3.33$, $p<.001$ and $t=2.38$, $p<.05$ respectively) indicating that younger children were more affected than adults in terms of how long it took them to complete the task. In addition to the longer task transaction time, overall, participants found less differences in the DEG (M=9) than in the NORM condition (M=11; t=-

2.73, p<.01). Together these findings of longer transaction times and fewer differences found in the DEG than in the NORM condition indicate greater task difficulty, and thus an increased need for participants in 'Talker A' role to clarify their speech.

3.2 Statistical analysis

Median f0: The final model for median f0 included the fixed effects of Condition, Sex and Age band, the interaction between Sex & Age band, and a random effect of Participant. The model is summarized in Table 1 of the supplementary materials.

All groups increased their median f0 in the DEG condition compared to the NORM condition (p<.001) showing that, on average, talker groups raised their fundamental frequency by approximately 2 semitones when attempting to clarify their speech for their interlocutor. However, despite the fact that all three groups of children showed an increase in median f0 in the DEG condition, there were several age-related developmental trends in median f0 of spontaneous speech. As expected, child groups produced speech with a higher median f0 than adults (p<.001). Across conditions, median f0 values decreased gradually from 98 semitones to 92 semitones in children between 9 and 14 years of age, and had not yet reached adult levels (86 semitones) by 14 years of age (all child groups differed from adults, p<.001). Contrary to previous findings, however, adult-like sex differences in median f0 did not emerge by 13-14 years of age as all child groups differed from adults (see Figure 2; all comparisons, p<.01, see Table 1 in the supplementary material). In the three groups of children, the difference in median f0 between boys and girls varied between 1 semitone (11-12 years) to 5 semitones (13-14 years) whereas in the adult groups the difference between male and female speakers was 10 semitones.

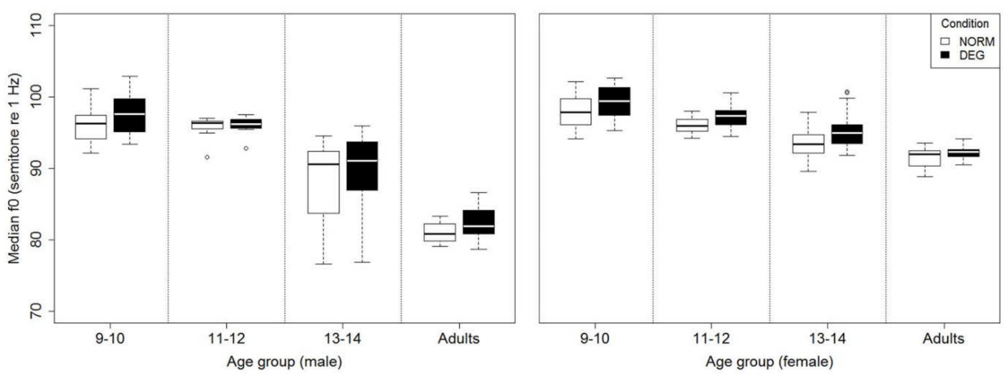

Figure 2. Box plot showing measures of median f0 as a function of Age band and Condition. The data for male talkers is displayed in the left panel, and for female talkers in the right panel.

f0 range: The final model for f0 range included the fixed effects of Condition and Age band, the interaction between Condition & Age band, and a random effect of Participant. The model is summarised in Table 2 of the supplementary materials. In the DEG condition, 11-12 and 13-14 year old children increased their f0 range on average by 1.4 semitones (11-12 and 13-14 years versus adults, p<.05) whereas adults and the youngest group of children only raised f0 range on average by 0.3 semitones (see Table 2 of the supplementary materials, and Figure 1; 9-10 years versus adult, p>.05). This indicates that child groups (except the youngest 9-10 years group) are showing a significantly greater increase in f0 range than adults when clarifying their speech. Overall, across conditions, f0 range decreased gradually with age from 4.4 semitones in the 9-10 years group to 3.9 semitones in the 11-12 years group, reaching adult-like values (Adult M=2.9 semitones) in the older 13-14 years group (M=3.6 semitones).

Articulation rate: The final model for articulation rate included the fixed effects of Condition, Sex and Age band, as well as the interactions between Condition & Sex, Sex & Age band, Condition & Age band, Condition, Sex & Age band, and a random effect of Participant. The model is summarised in Table 3 of the supplementary materials. Articulation rate was slower in the DEG condition than in NORM ($p<.001$). Despite the significant Condition and Age band interaction, all four groups of participants slowed down their articulation rate for the DEG condition compared to the NORM condition (DEG - NORM: 9-10 years: 1.0 syllables/second; 11-12 years: 1.0; 13-14 years: 0.9 and adults: 1.0; $p<.001$). However, the difference between NORM and DEG condition varied significantly from adults in the 9-10 years group possibly because their articulation rate was overall slower in both NORM and DEG conditions (see Table 3 of supplementary materials, and Figure 1).

The significant 3-way interaction between Condition, Sex and Age band revealed that whilst all four groups slowed down their articulation rate for the DEG condition, in the adult group male participants slowed down more than female participants whereas in the three groups of children, girls slowed down more than boys, though this comparison was only significant in the youngest (9-10 year) group (see Tables 1 and 2).

However, despite the fact that all three groups of children adapted the articulation rate to counter the effect of vocoding in an adult-like manner, there were significant age-related changes in articulation rate in children. Overall, the articulation rate increased with increasing age (9-10 years: 3.3 syllables/second; 11-12 years: 3.6; 13-14 years: 3.7 and adults: 4.2) and all three groups of children were significantly slower than adults (all three groups of children versus adults, $p<.001$).

Mean energy (1-3 kHz): The final model for ME1-3kHz included the fixed effects of Condition, Sex and Age band, the interaction between Condition & Sex, and a random effect of Participant. The model is summarized in Table 4 of the supplementary materials. All three groups of children had higher mid-frequency energy than adults ($p<.001$). Note that this may partly be due to differences in the distribution of spectral energy between child and adult speech. All participants (children and adults) increased their mid-frequency energy in the DEG condition ($p<.001$). However, despite the significant 2-way interaction between Condition and Sex, female participants only showed marginally greater mid-frequency energy in DEG condition than male participants (Female: NORM: 65.8, DEG: 67.5; Male: NORM: 64.7, DEG: 65.7 dB; see Table 4 in the supplementary material).

3.3 Correlations in the degree of change in acoustic measures between the NORM and DEG conditions

When vocal effort is increased, as occurs when communicating with a talker who is a distance away or shouting, increases in intensity are typically coupled with increases in f0 and reductions in speaking rate (e.g., Traunmüller & Ericsson, 2000). If the adaptations found in the DEG condition are generally linked to an increase in vocal effort, then we would expect that increases in mid-frequency energy (ME1-3 kHz), reflecting a reduction in spectral tilt, would be correlated with f0 and articulation rate measures. For each measure, we first calculated the percentage of relative change across the DEG and NORM conditions: e.g., ((f0_DEG-f0_NORM)/f0_NORM)*100. Pearson's correlations (See Table 3) were then carried out to look at the relation between changes in ME1-3 kHz and changes in these other two measures (median f0 and articulation rate). F0 range was not included as it is typically correlated with median f0. Across all children, increases in ME1-3 kHz were significantly correlated with increases in median f0 ($r=0.594$, $N=94$, $p<.0001$) and decreases in articulation rate ($r=-.273$, $N=94$, $p=.008$). However, this relationship did not hold in the adult group, where there was no significant relation between ME1-3 kHz and either median f0 ($r=0.301$, N=20, p=.197) or articulation rate (r=-0.299, N=20, p=.20). Due to the difference in group size between adults and children, it is relevant to look at whether correlations remain significant within each of the child age bands. In the 9-10 years and 13-14 years groups, increases in ME1-3 kHz were correlated with increases in median f0 (r=0.774, N=29, p<.0001 for 9-10 year olds; r=0.584, N=42, p<.0001 for 13-14 year olds) but not with decreases in articulation rate (r=-0.330, N=29, p=.081 for 9-10 year olds; r=-0.129, N=42, p=.414 for 13-14 year olds). For 11-12 year olds, increases in ME 1-3 kHz were correlated with increases in median f0 (r=0.423, N=23, p=.026) and with decreases in articulation rate (r=-0.485, N=23, p=.019). Together these results indicate that when children increased their energy in the mid-frequency range, they also increased their median f0 and some children slowed down their speech, which is compatible with a general increase in vocal effort. The absence of this significant correlation in adults suggests that they may be using more modulated strategies in the face of their interlocutor's communication difficulties.

Table 3. Pearson's correlation coefficients between median f0, f0 range, mean energy (1-3kHz) and articulation rate in DEG condition across all children (N=96).

	% relative change in median f0 (semitone)	% relative change in mean energy (1-3 kHz)	% relative change in articulation rate
% relative change in median f0(semitone)	1 N=96		
% relative change in mean energy (1-3 kHz)	.594 p=<.0001 N=94	1 N=94	
% relative change in articulation rate	-.282 p=.005 N=96	-.273 p=.008 N=94	1 N=96

4 Discussion

This study investigated suprasegmental aspects of spontaneous speech produced by 96 young talkers aged 9 to 14 years. This speech was produced within the context of a problem-solving task with a friend of the same sex in good communicative conditions and in a condition when intelligibility was challenging for the child's interlocutor.

The first research question addressed whether children and adolescents adapt their speech to maintain good communication in challenging listening environments, and whether their strategies differed from adults. We found that children between 9-14 years of age showed adaptations in the suprasegmental measures of fundamental frequency median and range, articulation rate and mid-frequency energy when clarifying their speech to counter the effects of vocoding. More specifically, our results show that from 9 years of age, children use adult-like adaptations by slowing down their articulation rate and increasing the mid-frequency energy and median fundamental frequency of their speech for the benefit of their interlocutor. However, unlike adults, 11-14 year olds also increased their fundamental frequency range to counter the effects of vocoding. This increase in f0 range is unlikely to be due to an overall increase in the proportion of question utterances in this condition as it is Talker B who was having difficulty understanding and was therefore asking more questions. We have no clear explanation for the fact that younger children did not show the same increase in f0 range in the DEG condition as older children. The lack of expansion in f0 range in younger children may also be linked to the fact that they already showed an expanded f0 range in the NORM condition. Another possibility may be that the 11-14 year olds showed more expanded prosody in the DEG condition in response to their interlocutor's perceptual difficulties (e.g., due to increased frustration) but this explanation is difficult to substantiate.

A second research question concerned evidence for a developmental trend in the acoustic-phonetic characteristics of spontaneous speech produced by young people aged 9 to 14 years. In addition to speech adaptations, across both conditions we found significant age-related changes in all four acoustic-phonetic measures within the 9 to 14 year age range. As children got older, they reduced their median f0 and f0 range, they became faster talkers and showed reduced mid-frequency energy in spontaneous speech produced in good listening conditions. Measures showed different maturational patterns: children were still showing higher median f0, greater mid-frequency energy and slower articulation rate than adults in our oldest group of 13-14 year olds. However, by the age of 14, children had reached adult-like values for f0 range.

Together these results indicate that despite ongoing developmental changes in their acoustic-phonetic aspects of their speech in late childhood, children can produce adult-like clear speech adaptations for the benefit of their interlocutor. However, the changes that they are making are fairly consistent with a general increase in vocal effort. Indeed, increases in mid-frequency energy, reflecting a decrease in spectral tilt, were coupled with increases in f0 characteristics and, for some groups, in decreases in articulation rate. Increases in f0 and mid-frequency energy and reduction in duration, are characteristics of Lombard speech, i.e. the speech style produced when speaking in background noise (e.g. Junqua, 1993; Van Summers et al., 1988) which involves an increase in vocal effort. We suggest that children used clear speech strategies learnt through their more usual experience of communicating in noisy environments. Given that this strategy was still present in 13-14 year olds, an intriguing question is the age at which young speakers do begin to show greater differentiation in clear speaking styles in response to different types of environmental or linguistic barriers. As youngsters typically become exposed to much greater variability in accents and communicative conditions in late teenage years, as they move from school to higher education or employment, this further experience may lead to ongoing changes in such attunements to different communicative conditions. Adults did not show any increases in f0 range, and tended to show less of an increase in median f0 than children, although this effect failed to reach significance. Also, their increase in mid-frequency energy was not significantly correlated with changes in fundamental frequency characteristics or articulation rate. In Hazan and Baker (2011), it was argued that fundamental frequency changes are not helpful to the interlocutor as not transmitted via the noise-excited vocoder, although it should be stressed that not all prosodic information is absent as it may still be transmitted within the noise-excited signal via changes in spectral characteristics (Niebuhr, 2012). Hazan and Baker (2011) found that adult talkers used different adaptation strategies when countering the effects of vocoding and the effects of babble-noise for their interlocutor. Here, the differences between the acoustic characteristics of child and adult speech in the DEG condition suggest that, while children are making changes consistent with a general increase in vocal effort in the DEG condition, adults are using more specific strategies. An alternative explanation, which cannot be discounted, is that children simply show a slower learning effect in terms of developing appropriate strategies, and that they may have attuned their strategies given a longer task.

The DEG condition in which a communication barrier is introduced that only affects the interlocutor, is informative as regards the degree to which speech production is guided by communicative needs rather than the talker's articulatory ease. Note that the communicative aspect is a key element of this study: studies that instruct children to speak clearly through direct or indirect means (e.g. Redford & Gildersleeve-Neuman, 2009) may show that children are competent at hyperarticulating their speech when requested to do so but this study investigates whether children do indeed use these strategies when the demands of the interaction require it, even if they themselves are not experiencing any communication deficit. According to the Hyper-Hypo model of speech production (Lindblom, 1990), environmental difficulties that affect communication should induce the ‘unimpaired’ talker to put greater weight on ‘system-oriented controls’ and use greater articulatory effort and thus a clear speaking style (‘Hyperspeech’). Indeed, according to the H&H model, speech production is goal-directed and aimed at maximizing communication; talkers have to dynamically adjust their speech production along a hyper- to hypospeech continuum, as communicative demands change, to maintain communication at the least cost in terms of articulatory effort. Within the H&H model, there are no predictions as to when these processes may develop during language acquisition. There is some evidence that even young children below the age of five are sensitive to the needs of their interlocutor: they use short utterances and simpler syntactic structures when talking to children than to adults (Shatz & Gelman, 1977) and, as discussed in the introduction, they make some adaptations to their speech production. Generally though, children may have greater difficulty in adopting another person’s perspective and a level of egocentrism is still evident in adolescents (e.g. Blakemore & Choudhry, 2006). Furthermore, in this task, they are interacting with a friend, with whom they would usually adopt a casual speaking style. Our data show that children within this age range did adopt a type of hyperspeech in the DEG condition, even in this situation of one-way communicative difficulty and when facing an unknown type of degradation.

Differences between the clear speech characteristics of adults and children may be at least partly related to more general differences in their speech. It has been suggested that the casual speaking style of 3-5 year old children cannot really be characterized as ‘hypospeech’ in that it does not show the type of articulatory reduction that is found in adult casual speech (Redford & Gildersleeve-Neuman, 2009). This may be due to immature articulatory skills resulting in inefficient articulatory movements (Walsh & Smith, 2002). Although the children recorded in our study were much older, Walsh and Smith (2002), when directly measuring motor processes in terms of their displacement, velocity and variability, showed that full maturity was not achieved until late adolescence. If children produce less reduced speech than adults, they may have less scope for further clarifying their speech. However, these differences are more likely to affect segmental measures such as the measurement of vowel space where there are clearer absolute limits as to the degree of further expansion that may be possible. Indeed, the same group of young talkers as reported here produced a much more expanded vowel space area than adults in the NORM condition; the increased differentiation with age between their vowel space area in the NORM and DEG conditions was due to the fact that older children showed greater reduction in vowel space for the NORM condition, while maintaining similar rates of hyperarticulation in the DEG condition (Pettinato et al., 2016). Here, in terms of articulation rate for example, children were slower than adults overall, but still showed similar rates of further reductions as a clear speech strategy.

A subsidiary aim of this study was to provide normative data for articulation rate and fundamental frequency characteristics in children for spontaneous speech produced in interaction with another talker, rather than from monologues or sentence reading. Our task involved greater degree of cognitive load than in a simple sentence imitation task, for example, as the children had to plan and formulate their utterances at the same time as completing a game-like task; also they had no face-to-face contact with their interlocutor which led to reduced articulatory and interactional cues. Given that speaking rates have been shown to decrease with increased task complexity (Nip & Green, 2013), there could be an expectation of slower articulation rates than those cited in previous studies with a similar age range. The articulation rates obtained here (*M*=3.8 syllables/second for 9-10 year olds, *M*=4.0

for 11-12 year olds, *M*=4.1 for 13-14 year olds) were indeed lower than those cited for conversational speech produced in answer to open-ended questions (e.g., 'tell me about your family') in Sturm and Seery (2007)'s study (*M*=5.6 for 9 year olds and *M*=5.jm5 in 11 year olds), although note that those measures were based on perceptually-fluent sentences only. Our articulation rates compared well with those reported by Nip and Green (2013) for 9 to 13 year olds for a narrative retell task which required participants to recall a story and formulate a set of sentences to describe it (*M*=3.6 to 4.1 approximately, as derived from Figure 2), also a cognitively demanding task. The fact that all child groups had a slower articulation rate than adults is consistent with the finding of differences between 12 year olds and adults in Jacewicz et al. (2010).

Our findings of ongoing changes in f0 characteristics within this age range were also very much as expected from the previous literature and likely to be primarily due to physiological changes in puberty, although the differentiation between male and female talkers, apparent from around 12, precedes the age at which the greatest effects of puberty are typically seen (Ferrand & Bloom, 1996). The greater change in f0 in boys also reflects the much greater increase in vocal fold size from pre-puberty to puberty than for girls, a change typically occurring between 7 and 13 years. The fact that children spoke with higher mid-frequency energy than adults replicates previous findings with younger children (e.g. Amazi & Garner, 1982). Note however that differences in energy measures between adults and children might at least partly be due to the fact that energy distribution across the speech spectrum varies across these two participant groups. Overall, therefore, the differences in suprasegmental features found in the spontaneous speech of older children and adults are in great part due to physiological differences related to the voice source.

In summary, in accord with listener-oriented models of speech production, older children are able to make adult-like adaptations to the suprasegmental aspects of their speech to benefit their interlocutors in challenging communicative conditions, even when not experiencing intelligibility difficulties themselves. Further developments are likely in later adolescence which may lead to adaptations that are more closely modulated to specific communicative situations.

5 Acknowledgments

This work was supported by the Economic and Social Research Council [grant number: RES-062-23-3106]. The full dataset is accessible at the following website: http://reshare.ukdataservice.ac.uk/851525/

6 References

Amazi, D.K., & Garber, S.R. (1982). The Lombard sign as a function of age and task. Journal of Speech and Hearing Research, 25 (4), 581-585.

Baker, R., & Hazan, V. (2011). DiapixUK: task materials for the elicitation of multiple spontaneous speech dialogs. Behavior Research Methods, 43(3), 761-770.

Barr, D.J., Levy, R., Scheepers, C., & Tily, H.J. (2013). Random effects structure for confirmatory hypothesis testing: keep it maximal. Journal of Memory and Language, 68, 255-278.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and s4. R package version 1.0-6. http://CRAN.R-project.org/package=lme4.

Bennett, S. (1983). A 3-year longitudinal study of school-aged children's fundamental frequencies. Journal of Speech and Hearing Research, 26, 137-142.

Blakemore, S-J., & Choudhury, S. (2006). Development of the adolescent brain: implications for executive function and social cognition. Journal of Child Psychiatry and Psychology, 47, 296-312.

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, 543–571.

De Looze, C., & Hirst, D.J. (2008). Detecting Key and Range for the Automatic Modelling and Coding of Intonation, Proceedings of the Speech Prosody 2008 Conference, Campinas, Brazil, 135-138.

Faulkner, A., Rosen, S., & Green, T. (2012). Comparing live to recorded speech in training the perception of spectrally-shifted noise-vocoded speech. Journal of the Acoustical Society of America – Electronic letters, 132 (4), EL336-EL342.

Ferrand, C. T., & Bloom, R. L. (1996). Sex differences in children's intonational patterns. Journal of Voice, 10(3), 284–291.

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. (2002). Longitudinal changes in articulation rate and phonetic phrase length in children with speech delay. Journal of Speech Hearing and Language Research, 45, 100-110.

Garnier, M., Wolfe, J., Henrich, N. & Smith, J. (2008). Interrelationship between vocal effort and vocal tract acoustics: a pilot study. Proceedings of INTERSPEECH, Brisbane, Australia, 2302-2305.

Gelfer, M. P., & Denor, S.L. (2014). Speaking fundamental frequency and individual variability in Caucasian and African American school-age children. American Journal of Speech-Language Pathology, 23, 395–406.

Glave, R. D., and Rietveld, A. C. M. (1975). Is the effort dependence of speech loudness explicable on the basis of acoustical cues. Journal of the Acoustical Society of America, 58, 875–879.

Hacki, T., & Heitmueller, S. (1999). Development of the child's voice: permutation, mutation. International Journal of Pediatric Otorhinolaryngology, 49, Suppl. 1, S141-S144.

Hawkins, S. (2003) Roles and representations of systematic fine phonetic detail in speech understanding. Journal of Phonetics, 31, 373-405.

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.

Hollien H., Green R., & Massey K. (1994). Longitudinal research on adolescent voice change in males. Journal of the Acoustical Society of America, 96, 2646–2654.

Jacewicz, E., Fox, R. A., & Salmons, J. (2011). Regional dialect variation in the vowel systems of typically developing children. Journal of Speech, Language, and Hearing Research, 54, 448-470.

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, 839-850.

Johnson, C.E. (2000) Children's phoneme identification in reverberation and noise. Journal of Speech Language and Hearing Research, 43 (1), 144-157.

Junqua, J.C. (1993).The Lombard reflex and its role on human listeners and automatic speech recognizers. Journal of the Acoustical Society of America, 93(1), 510-524.

Kleinow, J., & Smith, A. (2006). Potential interactions among linguistic, autonomic, and motor factors in speech. Developmental Psychobiology, 48, 1098-2302.

Krause, J. C., and Braida, L. D. (2004). Acoustic properties of naturally produced clear speech at normal speaking rates. Journal of the Acoustical Society of America, 115, 362–378.

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 Hardcastle, W.J. and Marchal, A. (eds) Speech Production and Speech Modelling, 403-439, The Netherlands: Kluwer Academic.

Local, J.K. (2003). Variable domains and variable relevance: Interpreting phonetic exponents. Journal of Phonetics, 31, 321-339.

Mennen, I., Schaeffler, F., & Docherty, G. (2012). Cross-language difference in f0 range: a comparative study of English and German. Journal of the Acoustical Society of America, 131 (3), 2249-2260.

Niebuhr, O. (2012). At the edge of intonation: The interplay of utterance final F0 movements and voiceless fricative sounds. Phonetica, 69, 7-27.

Nip, I.S.B., & Green, J.R. (2013). Cognitive and linguistic processing primarily account for increases in speaking rate with age. Child Development, 84 (4), 1324–1337.

Pedersen, M.F., Møller, S., Krabbe, S., Bennett, P., & Svenstrup, B. (1990). Fundamental frequency in female puberty measured with electroglottography during continuous speech as a secondary sex characteristic. A comparison between voice, pubertal stage, estrogens and androgens. International Journal of Pediatric Otolaryngology, 20, 17-24.

Pettinato, M., Tuomainen, O., Granlund, S., & Hazan, V. (2016) Vowel space area in later childhood and adolescence: effects of age, sex and ease of communication. Journal of Phonetics, 54, 1-14.

Redford, M.A., & Gildersleeve-Neumann, C.E. (2009). The development of distinct speaking styles in preschool children. Journal of Speech, Language, and Hearing Research, 52,1434 –1448.

Rinker, T. W. (2013). qdap: Quantitative Discourse Analysis Package. version 1.3.1.University at Buffalo. Buffalo, New York. http://github.com/trinker/qdap

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.

Shannon, R.V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (1995). Speech recognition with primarily temporal cues. Science, 270, 303–304.

Shatz, M., & Gelman, R. (1977). Beyond syntax: The influence of conversational constraints on speech modifications. In C. Ferguson and C. Snow (Eds.), Talking to children: Language input and acquisition. Cambridge, England: Cambridge University Press, 7, 189-198

Sluijter, A.C.M., Heuven, V.J. van (1996). Spectral balance as an acoustic correlate of linguistic stress. Journal of the Acoustical Society of America, 100, 2471-2485.

Smiljanic, R., & Bradlow, A.R. (2009). Speaking and hearing clearly: Talker and listener factors in speaking style changes. Linguistics and Language Compass, 3, 236–264.

Spencer, S., Clegg, J., & Stackhouse, J. (2013). Language, social class and education: listening to adolescents’ perceptions. Language and Education, 27, 129-143.

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, 47–59.

Syrett, K., & Kawahara, S. (2014). Production and perception of listener-oriented clear speech in child language. Journal of Child Language, 41, 1373 - 1389.

Traunmüller, T., & Eriksson, A. (2000).Acoustic effects of variation in vocal effort by men, women and children. Journal of the Acoustical Society of America, 107 (6), 3438-3451.

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, 510-540.

Van Summers W. V., Pison,i D. B., Bernacki, R. H., Pedlow, R. I., & Stokes, M. A. (1988). Effects of noise on speech production: acoustic and perceptual analyses Journal of the Acoustical Society of America, 84, 917-28.

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, 1119–1133.

Weppelman, T.L., Bostow, A., Schiffer, R., Elbert - Perez, E., & Newman, R.S. (2003) Children's Use of the Prosodic Characteristics of Infant-Directed Speech, Language & Communication, 23(1), 63-80.

Whiteside, S. P., Henry, L., & Dobbin, R. (2003). Patterns of variability in voice onset time: a developmental study of motor speech skills. Neuroscience Letters, 347, 29-32.

Supplemental materials

Table 1. Fixed and Random effects in a mixed-effects model of median f0.

Fixed effects		Estimate	SE	t	estimated p-value
(intercept)		81.02	0.85	**95.27**	<.001
Condition	DEG	1.39	0.16	**8.46**	<.001
Sex	Female	10.13	1.26	**8.03**	<.001
Age band	9-10	15.06	1.10	**13.70**	<.001
	11-12	14.11	1.30	**10.82**	<.001
	13-14	7.35	1.04	**7.09**	<.001
Age band * Sex	9-10, Female	-8.26	1.63	**-5.07**	<.001
	11-12, Female	-9.30	1.75	**-5.31**	<.001
	13-14, Female	-4.85	1.53	**-3.17**	<.01
Random effects		**Variance**			
Participant	(intercept)	7.10			
Residual		1.57			

Table 2. Fixed and Random effects in a mixed-effects model of f0 range.

Fixed effects		Estimate	SE	t	estimated p-value
(intercept)		2.81	0.20	**13.86**	<.001
Condition	DEG	0.10	0.19	0.55	>.05
Age band	9-10	1.39	0.26	**5.32**	<.001
	11-12	0.73	0.27	**2.65**	<.01
	13-14	0.30	0.25	1.23	>.05
Condition *Age band	DEG, 9-10	0.35	0.24	1.42	>.05
	DEG, 11-12	0.58	0.25	**2.29**	<.05
	DEG, 13-14	0.92	0.23	**4.03**	<.001
Random effects		**Variance**			
Participant	(intercept)	0.47			
Residual		0.35			

Number of observations=232, Participants=116.

Table 3. Fixed and Random effects in a mixed-effects model of Articulation rate.

Fixed effects		**Estimate**	**SE**	**t**	**estimated p-value**
(intercept)		5.03	0.15	**33.60**	<.001
Condition	DEG	-1.10	0.16	**-6.76**	<.001
Sex	Female	-0.77	0.22	**-3.46**	<.001
Age band	9-10	-1.30	0.19	**-6.67**	<.001
	11-12	-0.78	0.23	**-3.37**	<.001
	13-14	-0.86	0.18	**-4.69**	<.001
Condition * Sex	DEG, Female	0.24	0.24	1.01	>.05
Sex * Age band	Female, 9-10	0.85	0.29	**2.97**	<.01
	Female, 11-12	0.43	0.31	1.38	>.05
	Female, 13-14	0.63	0.27	**2.34**	<.05
Condition *Age band	DEG, 9-10	0.49	0.21	**2.34**	<.05
	DEG, 11-12	0.20	0.25	0.80	>.05
	DEG, 13-14	0.32	0.20	1.60	>.05
Condition * Sex * Age band	DEG, Female, 9-10	-0.90	0.31	**-2.86**	<.01
	DEG, Female, 11-12	-0.34	0.34	-1.02	>.05
	DEG, Female, 13-14	-0.39	0.29	-1.33	>.05
Random effects		**Variance**			
Participant	(intercept)	0.10			
Residual		0.15			

Number of observations=232, Participants=116.

Table 4. Fixed and Random effects in a mixed-effects model of Mean energy (1-3 kHz).

Fixed effects		**Estimate**	**SE**	**t**	**estimated p-value**
(intercept)		59.92	0.44	**135.93**	<.001
Condition	DEG	1.08	0.23	**4.67**	<.001
Sex	Female	1.23	0.37	**3.29**	<.001
Age band	9-10	7.10	0.51	**13.81**	<.001
	11-12	5.95	0.54	**10.93**	<.001
	3-14	5.64	0.48	**11.67**	<.001
Condition * Sex	DEG, Female	0.63	0.33	1.92	>.05
Random effects		**Variance**			
Participant	(intercept)	2.41			
Residual		1.51			

Number of observations=230, Participants=116.