INVESTIGATING CLEAR SPEECH ADAPTATIONS IN SPONTANEOUS SPEECH PRODUCED IN COMMUNICATIVE SETTINGS

Outi TUOMAINEN & Valerie HAZAN
Department of Speech Hearing and Phonetic Sciences,
University College London (UCL), London, UK.
o.tuomainen@ucl.ac.uk, v.hazan@ucl.ac.uk

Abstract
In order to investigate the clear speech adaptations that individuals make when communicating in intelligibility-challenging conditions, it would seem essential to examine speech that is produced in interaction with a conversational partner. However, much of the literature on clear speech adaptations has been based on the analysis of sentences that talkers were instructed to read clearly. In this chapter, we review methods for eliciting spontaneous speech in interaction for the purpose of investigating clear speech phenomena. We describe in more detail the Diapix task (Van Engen et al., 2010) and DiapixUK picture pairs (Baker & Hazan, 2011) which have been used in the production of large corpora investigating clear speech adaptations. We present an overview of the analysis of spontaneous speech and clear speech adaptations from the LUCID corpora that include spontaneous speech recordings from children, young and older adults.

Keywords: spontaneous speech, clear speech, speech corpus

1 Introduction

The aim of our research is to investigate the acoustic-phonetic adaptations that individuals make to their speech to be able to communicate effectively in challenging environments. These are most often described as talking in a 'clear speaking style' or with 'clear speech' and have been the subject of many investigations over the last 30 or so years. This chapter will review the experimental approaches that have typically been used to investigate clear speech and will argue for investigating this phenomenon using spontaneous speech produced in interaction during a communicative task. We describe techniques for recording speech corpora that have taken this approach. Finally, we present some key findings from three linked studies that have investigated clear speech adaptations in children, young and older adults.

Most speech communication occurs in situations that are less than ideal. These situations have been referred to in the literature as 'adverse' or 'challenging' conditions but in fact represent many of the conditions that we encounter in

DOI: http://doi.org/10.18135/CAPSS.9

our everyday lives. For example, a study that investigated the typical listening environments of a small group of older adults, using an experiential sampling method, found that for 40% of the reported time, individuals were in an environment that was ‘a bit noisy’ and that about 10% of the time was in situations that were noisy or very noisy (Hasan et al., 2014). A recent review paper (Mattys et al., 2012) provided a useful classification of causes of adverse conditions for speech communication. First, there can be environmental/transmission degradation such as the presence of noise or other voices in the environment, or a high degree of reverberation. There can be receiver limitations, such as the presence of a hearing loss, the lack of shared language knowledge or lack of available resources due to a high degree of cognitive load. To this classification can be added speaker limitations, such as the presence of a language impairment or speaking in an unknown accent. All these situations can lead to communication difficulties that can result in disfluencies, multiple requests for repetitions or clarifications, and misunderstandings needing repair.

Young adults are skilled at making adaptations to their conversational speaking style in order to overcome the effects of these types of degradation. The adjustments that they make typically require greater effort on the part of the speaker but are evidence, as suggested by Lindblom in his Hyper-Hypo model of speech production (Lindblom, 1990), that speech produced in interaction is listener-oriented and for the benefit of efficient communication. While talkers typically aim to minimise the degree of articulatory effort that they expend in their conversational speaking style, this will be increased for the sake of efficient communication if the type of communication barriers mentioned above are affecting intelligibility. Talkers continuously assess the level of understanding of their interlocutor via the appropriateness of their responses, the frequency of requests for clarification, pauses, and hesitations. In conditions in which individuals are conversing in a very noisy room, for example, they may adopt a clear speaking style, but if communication is progressing well, talkers might start to reduce the effort that they are making to speak more clearly. However, if there is a breakdown and need for repair, the degree of clear speaking style is likely to increase. This type of speaking style is therefore highly dynamic and listener-focused.

The adaptations that are made in clear speech can be at the level of acoustic-phonetic or linguistic adjustments. Excellent reviews of the types of adaptations made in clear speech can be found in Smiljanic and Bradlow (2009), Mattys et al. (2012) and Cooke et al. (2014). In summary, acoustic-phonetic adaptations can include reductions in articulation rate, increases in pause frequency and in fundamental frequency, shifts in the energy distribution in the voice and vowel hyper-articulation. Linguistic adaptations can entail the use of more frequent words, reductions in sentence length and complexity or changes in lexical diversity (e.g., Granlund et al., 2018). Adaptations are, to a degree, tailored to best counter the interference that the interlocutor may be experiencing (e.g., Cooke & Lu, 2010; Hazan & Baker, 2011). In summary, speaking style adaptations are a fairly skilled aspect of speech production and are often essential for efficient and effective communication in challenging situations.

Given that clear speaking styles are likely to be strongly dependent on the interaction between talker and listener, and on the degree of difficulty experienced by one or both conversational partners, it would appear of paramount importance to involve interaction and communicative intent in the study of clear speech adaptations. However, the great majority of studies of clear speaking styles which led to the findings listed above have involved an approach where communicative intent was absent. Indeed, a typical approach has been to instruct the participant to read a set of sentences 'normally' or using a conversational style, and then to ask participants to read the sentences again, but this time as if speaking to a person who is hearing impaired or who is a non-native speaker. In terms of experimental control, this approach has a number of advantages over spontaneous speech in that the speech to be analysed is consistent across talkers and across speaking styles. This is a great advantage when measuring the acoustic characteristics of speech which can be affected by coarticulation, lexical content and many other sources of variation. However, such read speech has a number of shortcomings. First, the recorded speech materials lack communicative intent and the dynamic adjustments that talkers make to their speech in natural interactions. Also, in studies involving read materials, the participant is merely a 'talker' whereas true communication involves a participant as both listener and talker, and, more often than not, doing another task while communicating. The added cognitive load involved in such interactions could affect aspects of speech production (e.g., Nip & Green, 2013).

Another important shortcoming of using read speech for investigations into clear speech adaptations is that there is evidence that the clear speech that is recorded when participants are instructed to read clearly differs in some respects from naturally-elicited clear speech. For example, Hazan and Baker (2011) found that, for a same set of talkers, clear speech that was elicited via instruction in read sentences showed more extreme changes in at least certain acoustic-phonetic characteristics than spontaneous speech produced to counteract intelligibility-challenging conditions. In a study comparing different types of instructions to speak clearly with speech directed at an interlocutor, Scarborough and Zellou (2013) found that instructions to speak clearly led to greater changes in vowel duration and more greatly-hyperarticulated vowels than the naturally-elicited clear speech. In the same study, when speech samples were presented in a lexical decision task, listeners responded more quickly to the naturally-elicited clear speech than to the speech spoken 'as if to someone who is hard of hearing'.

Differences were also found between ‘real’ and ‘imagined’ foreigner-directed clear speech and authors argued for the use of ‘communicatively authentic elicitation tasks’ in studies of clear speaking styles (Scarborough et al. 2007). Finally, within ‘instructed’ clear speech, the perceptual benefit of clear speaking styles varies with the type of instruction given (Lam & Tjaden, 2013).

Recently, there has been a move towards investigating clear speech adaptations in corpora of spontaneous speech collected while pairs of participants were involved in a problem-solving task (for a review, see Cooke et al., 2014). These dialogues may still be far from natural communication, as they are recorded in laboratory conditions and involved talkers carrying out a specific problem-solving task in order to maintain some control over the content and duration of the interaction. However, they provide an important half-way house between read speech and totally unstructured spontaneous speech. Another advantage of this approach is that they model the kind of multi-tasking and sharing of cognitive resources that occurs in much natural communication. Using this type of interactive task, clear speech adaptations can be naturally elicited by, for example, adding noise in the background while the task is being carried out, and such speech is then compared to speech recorded when there was no interference affecting participants.

An early example of a collaborative problem-solving task used in the recording of speech corpora is the ‘Map Task’ which was used in the development of the HCRC Map Task corpus (Anderson et al., 1991). The Map Task involves ‘instruction givers’ having to communicate details of a map route and of different key elements on the map to ‘instruction followers’ who have no indication of the route on their map; the two maps can also differ in some key elements. Task success can be measured using a deviation score from the accurate route. The task has been used in a number of studies investigating, for example, word segmentation cues (White et al., 2010) and the role of visual cues in communicating information (Anderson et al., 1991). The speech recorded using this task includes many direction-giving commands and requests. However, the maps are highly simplified and accompanied by labels, so there is little variation in the lexical content produced.

Another approach to elicit spontaneous speech dialogues has been to use popular problem-solving puzzles. In Cooke and Lu (2010), pairs of participants cooperatively completed Sudoku puzzles, which provided many repetitions of number words. Crosswords have also been used (Crawford et al., 1994); these can lead to the use of a wider range of lexical items than Sudoku. However, in both these approaches, both participants see the same information and one participant can dominate the task without requiring much input from the other. It is also the case that individuals vary widely in their skill and interest in completing word- or number-based puzzles and such tasks may be too complex for very young or much older participants. Tangram puzzles (Clark & Wilkes-Gibbs, 1986) have also been used in many studies of speech in interaction. Tangrams are visual puzzles in which sets of shapes (squares, triangles) have to be assembled in a specific way to form a shape. Tangrams have the advantage of involving less skill than word or number puzzles and can also be used in many different permutations, so are less limited than for example Map Tasks which have to be carefully constructed. The type of interactions they elicit is still fairly limited though and would include a high proportion of short commands. Tangrams have been used in studies such as Murfitt and McAllister (2001).

A task that was recently developed for investigations of clear speech adaptations in children is the Grid Task (Granlund et al., 2018). In this task, each participant is given a grid with pictures, an empty grid with squares with coloured numbers and a tray containing five different drawn versions of 16 keywords that formed minimal pairs (e.g., *peach* - *beach*). The aim of the task was for each conversational partner, without being able to see each other's grids, to replicate their partner's grid in their empty grid. In order to do this, they had to converse with their conversational partner to find for each of the 16 boxes on the grid the correct keyword, the correct version of the keyword, and the correct location of the keyword (see Granlund et al., 2018, for an example of the grid and picture materials). This task very much engaged the children and the need to differentiate five different representations of an object led to variation in the lexical content. Multiple iterations of each keyword were produced as well as many repetitions of numbers and colours.

Finally, some spontaneous speech corpora have taken the approach of recording conversations between two interlocutors on everyday topics that are likely to elicit a range of different views and opinions. For example, in the BEA corpus (Gósy, 2012), the conversation module involved the participant, interviewer and a third person discussing topics such as 'marriage vs cohabitation' or 'secondary school final exams'. This approach is likely to elicit more natural spontaneous speech than problem-based tasks, although some individuals might be more reluctant to express personal views and therefore produce less speech.

2 The Diapix task

A recent task, which is becoming widely used for recordings of speech in interaction, is the Diapix task (Van Engen et al., 2010), which was first developed to compare conversational speech interactions between pairs of native and non-native speakers. Diapix involves pairs of participants engaged in a 'spot the difference' picture task. Each participant is presented with a different version of the same cartoon-style picture, and both have to collaborate to find the differences between the two pictures without seeing each other's picture. A set of 12 carefully-designed picture pairs developed by Baker and Hazan, the DiapixUK pictures, have been used in a number of different studies and are available as supplementary materials in Baker and Hazan (2011).

The design of the DiapixUK picture pairs was done with great care (see Baker & Hazan, 2011). There are four picture-pairs for each of three main scenes: beach, street and farm, each containing 12 differences. A number of factors were controlled in these pictures, such as the position of differences within the picture, the type of difference to be found (presence/absence of object vs change in object), which of the two pictures was key to finding the difference (if absence of an object) in order to ensure that both participants had to take an active part in the task. The pictures were also made to be quite humorous to maintain interest and encourage more relaxed conversation. Analyses in Baker and Hazan (2011) revealed that, unless one talker was instructed to take the lead, Diapix led to balanced speech being recorded for both participants (Talker A: 51%, Talker B 49%) which differed from the Map Task where the instruction giver contributed 68% of words. Also, after participants had completed a practice picture, there was no learning effect when several Diapix tasks were run in succession, as shown by task transaction time. The level of difficulty of the pictures was also found to be consistent, as shown by non-significant differences in task transaction time, although there was greater variation for the later-developed pictures (named beach 4, street 4, farm 4).

The DiapixUK picture pairs (Baker & Hazan, 2011) were developed with reusability in mind: they were designed using Adobe Photoshop software with each object placed on a different layer so that the pictures could easily be edited if further changes were needed. For example, these pictures have been adapted for use with Finnish (Granlund et al., 2012), Spanish (Lecumberri et al., 2017) and Swedish participants (Sørensen et al., 2017) by changing some of the written elements (such as shop names) in the pictures. The DiapixUK picture pairs have also been adapted to investigate regional dialectal differences in British Sign Language (Stamp et al., 2016). It should be noted that certain visual elements, such as the fact that men were wearing socks with their sandals on the beach, are rather culturally-biased, but these objects can be edited or removed. Unlike word- or number-based puzzles that have different demands across groups varying in age and ability, Diapix is well suited for a wide variety of participants, including clinical populations, as the task can be solved using simple vocabulary and grammar. The DiapixUK picture pairs have been used, without alterations, with participants aged 8 to 85 years.

As with the Map Task, the differences in Diapix were designed to encourage the repetition of specific keywords, in this case words from /p/-/b/ and /s/-/ʃ/ minimal pairs, so that segmental contrasts could be analysed. It is also possible to obtain measures of vowel space by accumulating vowel formant measures for point vowels that occur frequently in content words throughout the spontaneous speech interactions (e.g., Pettinato et al., 2016). However, rather than detailed segmental analyses, Diapix recordings are more commonly used to investigate more global characteristics of speech, such as measures of articulation rate, fundamental frequency and long-term average spectrum. They can also be used to obtain measures related to the conversational interaction, such as rate and type of disfluencies or of repairs.

Another type of measure that can be obtained from Diapix is a measure of communication 'success' or efficiency. Indeed, difficulties in communication result in increased pausing, disfluencies, repetitions, elaborations that all lengthen the time needed to complete the task successfully. Measures of communication efficiency include task transaction time (e.g., Van Engen et al., 2010), the number of differences found in a set time, and the frequency of communication breakdowns (McInerney & Walden, 2013). Another measure which can be of use in evaluating participant dominance is the time spent by each holding the floor in the interactions while they are completing the task (Sørensen et al., 2017).

In order to use the Diapix task to naturally elicit clear speech adaptations, it is necessary to make communication difficult for one or both participants in the interaction. One means of achieving this while also maintaining high quality and 'clean' recordings for each of the participants in the interaction involves seating participants in separate sound-treated booths and having them communicate via headsets. Recordings are controlled via a computer in a separate control room. One or both audio channels can be manipulated to degrade the signal being transmitted from one participant to the other in real time. This can be done using a vocoder, for example, by adding noise or babble to the channel or by using software such as HELPS (Zurek & Desloge, 2007) to simulate a sensorineural hearing loss. The aim is to naturally elicit clear speech adaptations in the 'unimpaired' participant who has to make him or herself clear for their conversational partner who has difficulty hearing them. The advantage of using this approach is that the degree of degradation can be carefully controlled, with no headphone 'leakage' heard by the other participant, and also that the speech of each participant is recorded on a separate channel with no audible interference. This is particularly important if the speech is to be used for acoustic analyses. A simpler recording set-up with both participants seated in the same room and recorded on a single channel can work well if the aim is to collect more general measures of task duration or measures of disfluencies, for example, and if recording quality is less paramount.

To date, Diapix has been used in studies of clear speech adaptations in young adults (Hazan & Baker, 2011), children with typical hearing (Hazan et al., 2016) and hearing loss (Granlund et al., 2018), older adults (Tuomainen & Hazan, 2016), native and nonnative talkers (Van Engen et al., 2010). It has also been used to examine how speech characteristics vary when speaking in one's first and second language (Lecumberri et al., 2017) and to investigate phenomena of talker convergence (Kim et al., 2011; Solanski et al., 2015; Stamp et al., 2016). Recently, the Spanish version of the Diapix pictures was used for sociolinguistic purposes, in a language contact study in Colombia. The value of the Diapix task for investigating speech interactions in more clinical settings is also being recognised. A pilot study (McInerney & Walden, 2013) used Diapix interactions to evaluate the effect of assistive listening devices (ALD) on communication efficiency in older adults with hearing loss, using the frequency of communication breakdowns as efficiency measure.

3 LUCID corpora

Three major corpora were collected at UCL consecutively over a nine year period using the Diapix task and DiapixUK picture sets. The first corpus (LUCID: London UCL Clear speech in Interaction Database) includes extensive speech recordings for 40 native Southern British English adults (20 female) aged between 18 and 29 years old (mean age: 23 years). Participants were monolingual and had normal hearing thresholds. They were recorded in a number of easy and challenging communicative conditions, with three Diapix picture tasks per condition. In the easy NORM condition, both participants could hear each other without interference. Challenging conditions included the vocoder condition (VOC: talker B heard talker A via a three-channel noise-excited vocoder) which was done by all participants. For this condition and the NORM condition, the conversational partners were known to each other. There were two further conditions, each carried out by half of the participants: in the Babble condition (BAB) talker B heard talker A's voice mixed with 8-talker babble at approximately 0 dB SNR and in the L2 condition, talker B was a low-proficiency L2 speaker. For these two conditions, Talker B was a confederate not known to the key participant. Further details about the design of the challenging conditions and of the resulting corpus can be found in Hazan and Baker (2011). The LUCID corpus also included read sentences and picture naming in two speaking styles. In the casual style, participants were instructed to read 'casually as if talking to a friend' and in the clear speaking style, they were instructed to read 'clearly as if talking to someone who is hearing impaired'. The corpus includes stereo audio files for two-way dialog (wav format) and individual wav files for each speaker as well as word-aligned orthographic transcriptions (in Praat TextGrid format). Note that annotations are not available for the speech produced by the L2 confederates. This corpus is available online (following password request) and stored within the OSCAAR archive based at Northwestern University (https://oscaar.ci.northwestern.edu/).

The kidLUCID corpus includes Diapix recordings from 96 children and adolescents aged between 9 and 14 years inclusive (50 F, 46 M, mean: 11;8 years). Participants were non-bilingual native Southern British English speakers who reported no history of hearing or language impairments. In this corpus, only one Diapix task was carried out per condition. Diapix was carried out in three of the conditions also included in the LUCID corpus for comparability: the NORM, BAB and VOC conditions. Further details about corpus design are available in Hazan et al. (2016). This corpus, together with word-level annotations in Praat Textgrids, is also available within the OSCAAR archive.

The most-recently collected elderLUCID corpus includes speech from 83 single-sex pairs of native Southern British English adult talkers between the ages of 19 and 84 years. Talker A participants were from two distinct age groups: 'younger adults' (YA) aged 19-26 years (15 F, 11 M; Mean: 21.5 yrs) and 'older adults' (OA) aged 65-84 years (30 F, 27 M, Mean: 72.5 yrs). Participants in Talker B role were always younger adults (N = 83, between 18-30 years of age) of the same sex as the Talker A. Participants reported no history of speech or language impairments. YA participants all had normal hearing thresholds. OA participants had either normal hearing (OANH: 14 F, 13M), i.e., hearing threshold of < 20 dB between 250-4000 Hz, or a mild hearing loss (OAHL: 16 F, 14 M), with hearing threshold of < 45 dB between 250-4000 Hz, typical of early stages of age-related hearing loss or presbycusis. In addition to the normal (NORM) condition, there were three challenging conditions carried out by all participants. In the hearing loss simulation condition (HLS), the voice of Talker A was processed in real time through the HELPS software (Zurek & Desloge, 2007) mimicking the effect of severe-to-profound age-related hearing loss before being transmitted to Talker B. In the BABBLE (BAB-1) condition, the speech of talker A was mixed with the same 8-talker babble as used in the previous LUCID corpora before being channelled through to the confederate's headphones, at a difficulty level equated to the HLS conditions via a Modified Rhyme Task (MRT). In the other BABBLE (BAB-2) condition, both talkers heard the same babble as in BAB-1 but at 0 dB SNR. One Diapix task was carried out per condition. For the same participants, further speech is available for the same conditions using a sentence repetition task where participants had to read sentences to Talker B who had to repeat them back. The elderLUCID corpus will be made available on request from the authors.

As the three corpora were collected in separate studies, they are not fully comparable in terms of the methodology used in their collection. This reflects the difficult decision to be made between maintaining full compatibility across related corpora collected over a period of years, and making necessary improvements or adjustments, or simply practical changes to aid recruitment. For example, in the LUCID and kidLUCID corpora, participants carried out the

Diapix task with a friend (for the NORM and VOC conditions only in LUCID) whereas in the elderLUCID corpus, participants were paired with a young adult conversational partner they had just met. This difference in degree of familiarity is likely to have an effect in the level of alignment between speakers during their interactions, for example.

4 Summary of findings on spontaneous speech across the lifespan

In this section, we summarise the main findings resulting from the analyses of the suprasegmental features of articulation rate, fundamental frequency and long-term spectrum characteristics in the three LUCID corpora. A description of the post-processing stages used in analysing these acoustic features is in Hazan et al. (2016).

First, the availability of spontaneous speech data collected using a common task in related studies with children aged 9 to 14, young adults and older adults aged 65 to 85 enables us to examine age trends for conversational speech produced in good communicative conditions but without face-to-face visual cues. Trends for articulation rate (syllables produced per second) in conversational speech showed an inverted U shape with children up to the age of 11 speaking at a slower speech rate than young adults (Hazan et al., 2016), but older adults in the 65-85 year age range also speaking at a lower articulation rate than young adults (Tuomainen & Hazan, 2016a), see Figure 1. This is consistent with the findings of Jacewicz et al. (2010), for example, for spontaneous speech monologues and Bóna (2014) for a variety of speaking styles.

Fundamental frequency measures were calculated using the de Looze and Hirst formula described in Hazan et al. (2016). Changes in mean fundamental frequency (see Figure 2) followed trends in terms of talker sex and age that were expected from the literature with differentiation on the basis of talker sex appearing around the age of 13-14 years followed by a steep reduction in fundamental frequency for male speakers in young adulthood and a more gradual reduction for female speakers (Hazan, 2017). Some studies have identified increases in mean fundamental frequency in older males and decreases in post-menopausal female talkers but, although such a trend was present, this effect was not statistically significant in our elderLUCID corpus.

Normalized pitch range (75th -25th percentile range of fundamental frequency values calculated over the aggregated speech per condition and converted to semitones relative to 1 Hz) also showed a U-shape but in the opposite direction than articulation rate: both 9-12 year olds and older adults used a wider pitch range in their conversational speech than 13-14 year olds and young adults. It is possible that the increased pitch range in some participant groups reflects increased engagement with the task resulting in greater animation rather than physiological effects. There was an age effect also in terms of the energy distribution in the long-term average spectrum of speech. More specifically, we analysed the relative amount of energy in the mid-frequency region of the speech (1-3 kHz) which has been identified as an important predictor of speech intelligibility in background noise. The speech of children had higher mid-frequency energy than young adults (Hazan et al., 2016); this may partly be due to differences in the distribution of spectral energy between child and adult speech. There was also a significant difference between the speech of younger and older adults, with older adults having less energy in the mid-frequency region (Tuomainen & Hazan, 2016b). This could be linked to weaker and more irregular vocal fold excitation in older adults. Speech which has less energy in the mid-frequency region of speech is likely to be more difficult to understand in the presence of noise. This is because this region of the speech spectrum, which contains many acoustic-phonetic cues, is more likely to be masked by noise if it is of lower intensity.

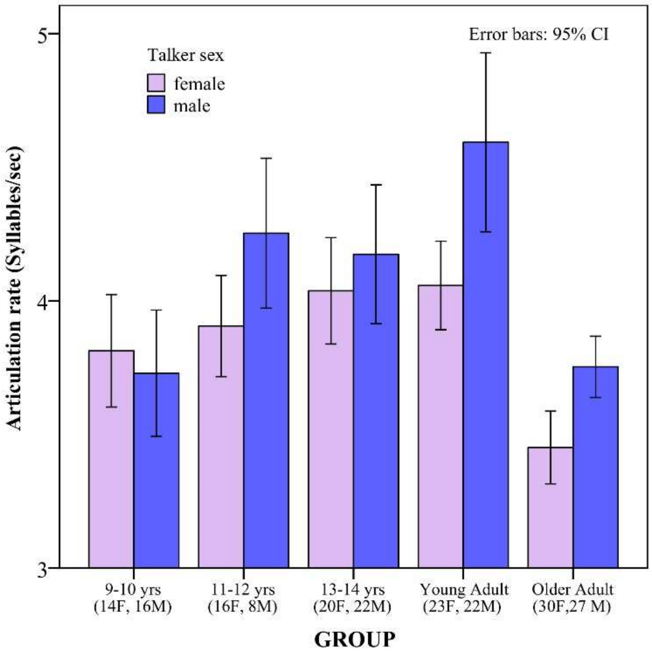

Figure 1.
Conversational articulation rate based on data collected from studies carried out with children (reported in Hazan et al., 2016) and with young and older adults (reported in Tuomainen & Hazan, 2016).

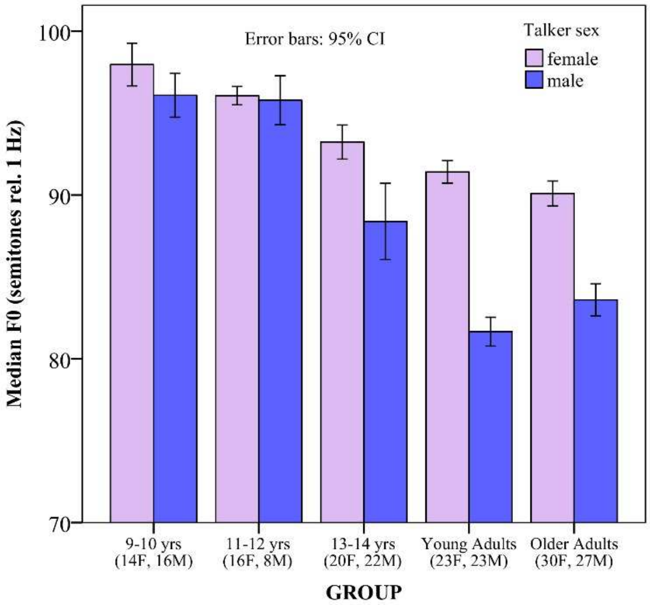

Figure 2.
Median fundamental frequency in conversational speech,
based on data collected from studies carried out
with children (reported in Hazan et al., 2016) and
with young and older adults (reported in Tuomainen & Hazan, 2016).

5 Summary of findings on listener-oriented adaptations in speech production

As the challenging conditions were not identical across the three Diapix studies, and as adaptation strategies can vary as a function of the type of communication barrier to be overcome (e.g., Hazan & Baker, 2011), it is not possible to directly compare the three age groups on a given condition. However, it is possible to compare the adaptation strategies of children to those of young adults for the VOC condition and the strategies of young adults to older adults for the hearing loss simulation condition. Both these conditions proved to cause significant communicative difficulties for participant pairs resulting in longer task transaction times (Hazan et al., 2016). It should be noted also that both involved an interference affecting Talker B only and that the adaptation strategies of Talker A were under scrutiny. These were therefore strategies that were purely listener-oriented and for the benefit of efficient communication and should be differentiated from the type of adaptations (e.g., Lombard speech) that talkers make when directly exposed to interference such as loud background noise.

When comparing the strategies used by children to those used by young adults, it was found that from 9 years of age, children used some adult-like adaptations: they slowed down their articulation rate and increased 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, even though this would not have been transmitted to the conversational partner. This was the case because a noise rather than periodic source was used to excite the vocoder and information about changes in periodicity would therefore not have been present. For child speech only, in the clear speaking style, significant correlations were obtained between increases in mid-frequency energy, reflecting a decrease in spectral tilt, and increases in f0 median and range and, for some child groups, in decreases in articulation rate. Such correlations suggest an increase in vocal effort as would be seen when speaking in a very loud voice or shouting. Children were perhaps using clear speech strategies learnt through their more usual experience of communicating in noisy environments, which would suggest less attunement to the specific characteristics of the interference (Hazan et al., 2016).

When comparing younger and older adults in the hearing loss simulation condition (affecting Talker B), both groups reduced their articulation rate in this condition and both groups also showed increased energy in the mid-frequency region of the long-term average spectrum relative to their spontaneous speech produced in good communicative conditions. Interestingly, for the older adult group with age-related hearing loss only, just as had been found for children, a strong correlation was obtained between increases in fundamental frequency and increases in mid-frequency energy (Hazan & Tuomainen, 2017) as shown in Figure 3. Again, this correlation was totally absent for the young adult group and for older adults with normal hearing thresholds and suggested that some older adults at least were using a strategy of strongly increasing their vocal effort as a clear speech strategy. An intriguing question is why older adults and young children have a greater tendency of strongly raising their voice or shouting when trying to speak clearly for an ‘impaired’ interlocutor. This could be due to a greater degree of frustration experienced when communication problems arise, although there is no objective data to support this hypothesis. It could also be due to a lower degree of inhibition as shouting to an interlocutor may be considered by young adults as inappropriate.

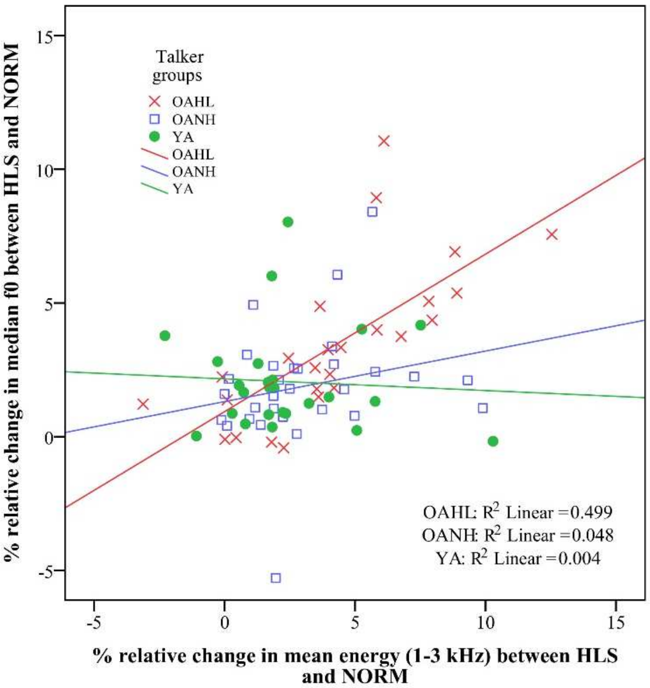

Figure 3.
Scatterplot showing the relation between changes in median f0 and changes in mean energy in the mid-frequency region of the spectrum in the hearing loss simulation (HLS) condition relative to the NORM condition in the study reported in Hazan and Tuomainen (2017) involving young adults (YA), older adults with normal hearing (OANH) and older adults with hearing loss (OAHL).

6 Conclusions

In conclusion, we argue that, despite the increased variability that comes from using spontaneous speech in the analysis of clear speech adaptations, there are benefits in using speech in which these adaptations are naturally elicited due to communicative demands. Further work is needed in order to develop analysis methods that can better represent the dynamic aspects of these adaptations in relation with the degree of communicative success.

References

Anderson, A. H., Bader, M., Bard, E. G., Boyle, E. H., Doherty, G. M., Garrod, S. C., Isard, S. D., Kowtko, J. C., McAllister, J. M., Miller, J., Sotillo, C. F., Thompson, H. S., & Weinert, R. (1991). The HCRC Map Task Corpus. *Language and Speech, 34,* 351-366.

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

Bóna, J. (2014). Temporal characteristics of speech: The effect of age and speech style. *Journal of the Acoustical Society of America Express Letters, 136,* EL116-EL121.

Clark, H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition, 22,* 1-39.

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.

Cooke, M., & Lu, Y. (2010). Spectral and temporal changes to speech produced in the presence of energetic and informational maskers. *Journal of the Acoustical Society of America*, *128,* 2059-2069.

Crawford, M. D., Brown, G. J., Cooke, M. P., & Green, P. D. (1994). The design, collection and annotation of a multi-agent, multi-sensor speech corpus. In *Proceedings of the Institute of Acoustics, 16,* 183-189.

Gósy, M. (2012). BEA − A multifunctional Hungarian spoken language database. *The Phonetician, 105,* 50-61.

Granlund, S., Hazan, V., & Baker, R. (2012). An acoustic-phonetic comparison of the clear speaking styles of late Finnish-English bilinguals. *Journal of Phonetics, 40,* 509-520.

Granlund, S., Hazan, V. L., & Mahon, H. M. (2018). Children's acoustic and linguistic adaptations of peers with hearing impairment. *Journal of Speech, Language, and Hearing Research, 61,* 1055-1069.

Hasan, S.S., Chipara, O., Wu, Y. H., & Aksan, N. (2014). Evaluating auditory contexts and their impacts on hearing aid outcomes with mobile phones. In *Proceedings of the 8th International Conference on Pervasive Computing Technologies for Healthcare* (pp. 126-133).

Hazan, V. L. (2017). Speech communication across the life span. *Acoustics Today, 13,* 36-43.

Hazan, V. L., & 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,* 2139-2152.

Hazan, V., & Tuomainen, O. (2017). Spontaneous speech adaptations in challenging communicative conditions across the lifespan. In *Book of abstracts of Workshop on Challenges in Analysis and Processing of Spontaneous Speech (CAPSS2017)* (3-4).

Hazan, V., Tuomainen, O., & Pettinato, M. (2016). Suprasegmental Characteristics of Spontaneous speech produced in good and challenging communicative conditions by talkers aged 9 to 14 years old. *Journal of Speech, Language, and Hearing Research, 59,* S1596−S1607.

Jacewicz, E., Fox, R. A., & Wei, L. (2010). Between-speaker and within-speaker variation in speech tempo of American English. *The Journal of the Acoustical Society of America, 128,* 839-850.

Kim, M., Horton, W. S., & Bradlow, A. R. (2011). Phonetic convergence in spontaneous conversations as a function of interlocutor language distance. *Laboratory Phonology, 2,* 125-156.

Lam, J., & Tjaden, K. (2013). Intelligibility of Clear Speech: Effect of Instruction. *Journal of Speech Language and Hearing Research, 56,* 2412-2421.

Lecumberri, M. L. G., Cooke, M., & Wester, M. (2017). A bi-directional task-based corpus of learners' conversational speech. *International Journal of Learner Corpus Research, 3,* 175-195.

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). Dordrecht: Kluwer Academic.

Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes, 27,* 953-978.

McInerney, M., & Walden, P. (2013). Evaluating the use of an assistive listening device for communication efficiency using the Diapix task: A pilot study. *Folia Phoniatrica Logopedia, 65,* 25-31.

Murfitt, T., & McAllister, J. (2001). Comprehension of spoken descriptions by novel listeners in monologue and dialogue. *Language and Speech, 44,* 325-350.

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

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

Scarborough, R., Dmitrieva, O., Hall-Lew, L., Zhao, Y., & Brenier, J. (2007). An acoustic study of real and imagined foreigner-directed speech. *Journal of the Acoustical Society of America, 121,* 3044-3044.

Scarborough, R., & Zellou, G. (2013). Continua of clarity: "clear" speech authenticity and lexical neighborhood density effects in production and perception. *Journal of the Acoustical Society of America, 134,* 3793-3807.

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

Solanki, V., Stuart-Smith, J., Smith, R., & Vinciarelli, A. (2015). Measuring mimicry in task-oriented conversations: the more the task is difficult, the more we mimick our interlocutors. In *Proceedings of InterSpeech 2015* (pp. 1815-1819).

Sørensen, J., Fereczkowski, M., & MacDonald, E. N. (2017). The effect of noise and second language on turn taking in task-oriented dialog. *Journal of the Acoustical Society of America, 141,* 3520.

Stamp, R., Schembri, A., Evans, B. G., & Cormier, K. (2016). Regional Sign Language Varieties in Contact: Investigating Patterns of Accommodation. *Journal of Deaf Studies and Deaf Education, 21,* 70-82.

Tuomainen, O., & Hazan, V. L. (2016a). Articulation rate in adverse listening conditions in younger and older adults. In *Proceedings of Interspeech 2016* (pp. 2105-2109).

Tuomainen, O., & Hazan, V. (2016b). Suprasegmental characteristics of spontaneous speech produced in good and challenging communicative conditions by younger and older adults. *Journal of the Acoustical Society of America, 140,* 3444.

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.

White, L., Wiget, L., Rauch, O., & Mattys, S. L. (2010). Segmentation cues in spontaneous and read speech. In *Proceedings of the 5th Conference on Speech Prosody* 2010 (pp. 1-4). Chicago.

Zurek, P. M., & Desloge, J. G. (2007). Hearing loss and prosthesis simulation in audiology. *Hearing Journal, 60,* 32-33, 36, 38.