Title
A recipient design in multimodal language on TV: A comparison of child-directed and adult-directed broadcasting

Permalink
https://escholarship.org/uc/item/17k7h7m6

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 45(45)

Authors
zhang, yanran
Gu, Yan

Publication Date
2023

Peer reviewed
A recipient design in multimodal language on TV: A comparison of child-directed and adult-directed broadcasting

Yanran Zhang (yanran.zhang@cuz.edu.cn)
Communication University of Zhejiang

Yan Gu (yan.gu@essex.ac.uk)
Department of Psychology, University of Essex & Department of Experimental Psychology, UCL

Abstract
Child-directed language is a unique multimodal communication behaviour that differs from adult-directed language. We investigated how broadcasters organize their multimodal language production on an adult and child-directed programme to better understand the recipient design in the broadcasting context. Thirty-six future broadcasters produced live programmes for children and adults, respectively, whose linguistic features (utterance=3888), speech prosody, and gestures (N=8486) were analysed as a function of programme. We found that broadcasters used a higher mean pitch but a smaller pitch range, shorter utterances, high(er) frequency words, more questions, pointing and representational gestures but fewer pragmatic gestures in child-directed broadcasting. Gestures were also more salient and slower when addressing children audiences. However, there were no differences in lexical diversity, speaking rate, pausing, or beat gestures between programmes. In conclusion, broadcasters did engage in recipient design multimodally, but the distinction between the speaker and audience orientation is not binary but should be understood across signal channels according to contexts.

Keywords: recipient design; multimodal communication; child-directed language; broadcasting; gesture

Introduction
People are capable of communicating according to various situations. For example, they may adjust their speech, intonation, and gestures according to different addressees, such as when talking to children versus adults. The process of tailoring language for an addressee is called recipient design (Sacks et al., 1974) or audience design (Clark & Murphy, 1982). Child-directed language (CDL) is a typical recipient design where caregivers talk to their children at a higher pitch level and a wider pitch range with shorter and simpler utterances. CDL is found across languages (Cox et al., 2022; Fernald et al., 1989). Even in sign languages, caregivers sign slowly with exaggerated movements or frequently repeat the same sign (Masataka, 1992; Perniss et al., 2017). So far, studies investigating CDL have mainly focused on family dyads, but few have investigated the recipient design in vastly different situations, such as the broadcasting context. Particularly, broadcasters’ speech is often considered more formal, faster, and higher in pitch than ordinary conversations (e.g., Cotter, 1993; Medrado et al., 2005), and there are usually only imagined addressees. It is entirely unknown how broadcasters use different signal channels (speech, intonation, gestures, etc.) as a communicative strategy in child-directed and adult-directed programmes, and whether they adopt recipient design in all information channels. This study compares future broadcasters’ multimodal language production between their child-directed broadcasting (CDB) and adult-directed broadcasting (ADB).

Broadcasting language
Newsreaders have often served as role models within a specific language community (Swerts & Krahmer, 2010). Various aspects of broadcasting languages have already been examined (e.g., Ambrazaitis & House, 2017; Rodero & Cores-Sarría, 2021; Swerts & Marsi, 2012). For example, Rodero and Cores-Sarría (2021) found that the standard speaking style of newsreaders is characterized by specific intonation and rhythm (e.g., a higher pitch level and an upward and downward intonation). Ambrazaitis and House (2017) found that head beats and eyebrow movements were associated with the phonological prosodic structure in Swedish newsreaders. In addition, by analysing the facial expressions and speech of two Dutch newsreaders (children’s versus adults’ programme), Swerts and Krahmer (2010) found that newsreaders strive to make their facial expressions congruent with verbal content that are more expressive when addressing children than adults in general.

Although a handful of studies have attempted to clarify characteristics in the broadcasting context, they used a between-subject comparison and had a small sample size (e.g., Mok, Fung & Li, 2014; Swerts & Krahmer, 2010). For example, Mok et al. (2014) studied the prosodic features of Cantonese newsreaders and non-newsreader speakers, but the contents of the news announced by each anchor were different, and the control group had one extra day to prepare. Swerts and Krahmer (2010) first attempted to investigate multiple cues in broadcasting programmes, but there were only two newsreaders. There is a paucity of research on social media, such as YouTube and its possible effects on young children’s learning and development (Izci et al., 2019; Neumann & Herodotou, 2020). Also, researchers encouraged broadcasters to consider children audiences’ characteristics when organizing children’s programmes (Abelman & Atkin, 2000), as television programmes are related to children’s physical health, physical activities, interest in the study, etc. (Jenvey, 2007; Gupta et al., 1994). However, it is still unclear how the same broadcasters adjust their multimodal
communication when producing the same content for adults and children in a naturalistic setting.

**Child-directed and Adult-directed language**

CDL differs from ADL in many aspects, and it facilitates children’s language learning and acquisition (e.g., Anderson et al., 2021; Golinkoff et al., 2015; Shi, Gu, & Vigliocco, 2022; Spinelli & Mesman, 2018). For example, CDL usually has exaggerated intonation and more varying pitch, a slower speaking rate, a shorter utterance length, and lower lexicon diversity with more iconic gestures (e.g., Campisi & Özyürek, 2013; Cristia, 2013; Han, De Jong & Kager, 2022), and English caregivers’ use of a higher mean pitch correlates with their children’s immediate unknown word learning outcome (Shi et al., 2022). However, most CDL studies have investigated communication between familiar acquaintances, such as mother/father/grandmother/sibling-child interaction (e.g., Cristia, 2013; Hoff & Krueger, 1991; Lamb & Lewis, 2010; Shute & Wheldall, 1999, 2001), or caregivers in schools (e.g., Kempe, 2009). Furthermore, CDL research typically focuses on characterising a few specific features (e.g., speaking rate, pitch, syntax, etc.) (Raneri et al., 2020; Song, Demuth, & Morgan, 2010; Zangl et al., 2005), and the age of children in such studies is usually below five years old.

So far, rare research has investigated CDL and ADL in the broadcasting context, which is different from typical caregiver-child dyads in terms of addressees and children’s ages. Given the global picture of children being increasingly exposed to social media, which is now an essential part of everyone’s life, broadcasting language is an important area for research into child-directed communication. However, we know little about how TV broadcasters use multimodal language to communicate with children and adults.

**Multimodality in recipient design**

Engaging in recipient design affects speech production (Snow & Ferguson, 1977; Isaacs & Clark, 1987; Galati & Brennan, 2010). For example, Tippenhauer et al. (2020) examined audience design in CDL and ADL, showing that speakers reduced word duration less in CDL than in ADL. Recipient design also shapes other communication means, such as co-speech gestures. Tippenhauer et al. (2020) only looked at recipient design in word duration, but the differences in gesture duration (the temporal length of a gesture) are unknown. Equivalent to the reasoning for word duration in linguistic aspects, the average duration of a gesture may also be shaped by recipient design when talking to different addressees.

Some aspects of co-speech gestures relate to mutually shared knowledge (common ground, Clark, 1996; Holler & Stevens, 2007; Holler & Wilkin, 2009). For example, speakers used more words and gestures when narrating novel contents and produced representational gestures at a higher rate, but significantly reduced the number of words and shortened repeated gestures in duration when common ground existed (e.g., Campisi & Özyürek, 2013; Galati & Brennan, 2010; Holler & Stevens, 2007; Schubotz, Özyürek, & Holler, 2018, 2022).

Also, researchers who study language production debate whether audience design is driven by speakers’ needs or by their audiences’ needs (e.g., Arnold et al., 2012; Aylett & Turk, 2004; Bard et al., 2000), but they always looked at specific features that cannot be generalized to different situations. For example, broadcasting is a unique situation where speakers do not have an addressee but seriously care about their imagined audiences. Hence, to contribute to this debate, we ask two detailed questions: First, how do TV broadcasters use multimodal language (e.g., speech, prosody, and gestures) when addressing different audiences (adults versus children)? Second, is multimodal broadcasting speaker-centred or listener-centred?

**Method**

**Participant**

Thirty-six participants majoring in Chinese broadcasting and hosting participated in the study (Mean age=19.7yrs, SD=0.93, all native Mandarin speakers). Several studies have found that mothers and fathers differentiated their language use when talking to their children (Van de Mieroop, Zenner, & Marzo, 2016). Also, mothers vocalized more frequently and for a longer time than fathers (Kiepura, Niedźwiecka, & Kmita, 2021). Moreover, in formal contexts that involved teaching and learning, the coda laterals of mothers were significantly darker than fathers (Sim, 2021). To avoid gender differences, in our study the participants were all female. Participants gave written informed consent to use the data before the recording started.

**Stimuli**

Participants were required to do a live broadcast explaining pictures in front of a camera both on a regular TV (adult-directed) and a children’s (child-directed, to students at primary schools) programme (see Figure 1 for experiment setting). In total, there were four pictures chosen from life-related aspects that adults and children easily understood.

![Figure 1: setting of the experiment.](image)

**Procedure**

Both programmes were recorded on the same day with a lunch break in between (sequences counterbalanced). They
had two minutes to prepare and talked for around one minute for each picture. The sequence of pictures was also counterbalanced. The mean talking duration of each picture in the ADB (M=74.63 sec, SD=21.08) and CDB (M =73.68 sec, SD =26.53) did not differ significantly (β=-1.11, p=.59). The speech was recorded through Audacity (44.1 kHz, 16-bit rate) with a wireless clip-on. The video was recorded by a high-quality mobile phone with a resolution of 2772×1344. Gestures were not mentioned at any point in the experiment.

Coding

Linguistic features of speech

Speech was transcribed via a transcription website (https://www.iflyrec.com/zhuanwenzi.html), and errors were corrected manually. We measured:
(1) **Mean length of utterance** (MLU), the average number of words per utterance (Dickinson & Porche, 2011);
(2) **Lexical diversity**, measured by MATTTR (moving average type-token ratio), which is less biased than TTR (type-token ratio) as it calculates average TTRs for successive sample pieces of a specified token size (Covington & McFall, 2010);
(3) **Sentence types** (statements and questions);
(4) **Word frequency**, calculated through the SUBTLEX-CH corpus (Cai & Brysbaert, 2010), which comprises Chinese word frequency based on Chinese subtitles.

Prosody

Boundaries of sentences were annotated in Praat (Boersma & Weenink, 2019). We measured:
(1) **Speaking rate** (the average number of words per second excluding pauses over 200ms (Han, 2019));
(2) **Pitch** (semitone, ST): mean F0, F0 SD and F0 range;
(3) **Intensity** (dB): mean intensity and intensity range;
(4) **Pauses** (the number of pauses, mean pausing duration and pausing rate of broadcasting each picture).

Gesture

Gestures were coded in ELAN (Wittenburg et al., 2006). We measured:
(1) **Gesture types**:
   - **Representational gestures**: metaphorical or iconic gestures to illustrate the concept by drawing its outline, indicating its shape or representing it (McNeill, 1998; Müller, 1998).
   - **Emblem gestures**: conventionalised gestures in certain areas (McNeill, 1992; Kendon, 2004). For example, a ring formed by the thumb and the index finger means OK in most countries, but it means “zero” in France (Morris et al., 1979). In actual coding, we merged emblems and representational gestures together.
   - **Pointing gestures**: a finger extends in the direction of something or points without any visible target, such as time or location (McNeill, 1992, 1998).
   - **Beat gestures**: simple motoric movements produced along with the rhythm of the speech (e.g., a hand with an open palm flipping outwards) (So, Sim, & Low, 2012). In actual coding,

Data analysis

In total, we had 351.61 minutes of audiovisual recordings. There were 3888 utterances (1891 in ADB and 1997 in CDB), and 8486 gestures (4339 ADB and 4147 CDB).

Linear mixed-effects models in the R environment (R Core Team, 2022) were used to assess differences in linguistic features, prosody, and gestures (dependent variables) as a function of broadcasting programmes (adult or child). Logistic regressions were used when dependent variables were binary (e.g., sentence types of question vs. statement). Participants were included as a grouping variable with a random intercept. The four pictures were control variables.

Result

Table 1 presents the results of linguistic cues, where broadcasters produced shorter utterances, words of a higher frequency, and a larger proportion of questions in CDB than in ADB (all ps<.001). The results suggest that broadcasters structured their speech to be more easily understood. However, in terms of lexical diversity, there were no significant differences (p=.19) between the two programmes.

<table>
<thead>
<tr>
<th>Dep variables</th>
<th>ADB</th>
<th>CDB</th>
<th>β and p</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLU</td>
<td>26.05(4.97)</td>
<td>23.81(4.49)</td>
<td>-2.24***</td>
</tr>
<tr>
<td>Lexical diversity</td>
<td>0.88(0.03)</td>
<td>0.89(0.03)</td>
<td>4.03</td>
</tr>
<tr>
<td>Word frequency</td>
<td>3.34(0.08)</td>
<td>3.37(0.08)</td>
<td>3.37***</td>
</tr>
<tr>
<td>Question</td>
<td>0.69(1.04)</td>
<td>2.02(1.80)</td>
<td>1.13***</td>
</tr>
</tbody>
</table>

Note: ***p < .001, **p < .01, *p < .05.

For prosody (Table 2), broadcasters produced a higher mean pitch, a greater pitch variation, a larger pitch maximum and pitch minimum, and a smaller pitch range in CDB than ADB (all ps<.001), indicating that broadcasters used a more exaggerated vowel to attract children’s attention. However, there were no significant differences in the speaking rate
(p=.65), raising rate (p=.36), mean pausing duration (p=.54), mean intensity (p=.12) and intensity range (p=.82).

Table 2: Means (SD) and results (beta and p) of prosodic cues for each condition.

<table>
<thead>
<tr>
<th>Dep variables</th>
<th>ADB</th>
<th>CDB</th>
<th>β and p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean F0(ST)</td>
<td>27.15(1.22)</td>
<td>27.99(1.29)</td>
<td>0.83***</td>
</tr>
<tr>
<td>SDF0 (ST)</td>
<td>58.47(6.39)</td>
<td>60.73(6.98)</td>
<td>2.31***</td>
</tr>
<tr>
<td>Max F0 (ST)</td>
<td>36.07(1.52)</td>
<td>36.53(1.37)</td>
<td>0.47***</td>
</tr>
<tr>
<td>Min F0 (ST)</td>
<td>16.14(2.07)</td>
<td>17.04(1.91)</td>
<td>0.89***</td>
</tr>
<tr>
<td>F0 range (ST)</td>
<td>15.08(2.59)</td>
<td>14.28(2.30)</td>
<td>-0.76***</td>
</tr>
<tr>
<td>Speaking rate</td>
<td>0.71(0.03)</td>
<td>0.71(0.04)</td>
<td>-0.002</td>
</tr>
<tr>
<td>Avg. Intensity (dB)</td>
<td>58.13(3.33)</td>
<td>58.30(3.57)</td>
<td>0.18</td>
</tr>
<tr>
<td>Intensity range (dB)</td>
<td>34.91(2.47)</td>
<td>34.88(2.55)</td>
<td>-0.034</td>
</tr>
<tr>
<td>Pausing rate</td>
<td>0.33(0.07)</td>
<td>0.33(0.07)</td>
<td>0.36</td>
</tr>
<tr>
<td>Avg. pause dur (sec)</td>
<td>0.40(0.09)</td>
<td>0.40(0.10)</td>
<td>0.54</td>
</tr>
<tr>
<td>Numbers of pauses</td>
<td>24.72(8.60)</td>
<td>23.83(8.64)</td>
<td>-0.94</td>
</tr>
</tbody>
</table>

Note: ***p < .001, **p < .01, *p < .05.

As for the results of gestures (Table 3), first, broadcasters produced a higher proportion of representational (23.55% vs. 17.72%, β=0.41, p=.001) and pointing gestures (19.58% vs. 17.67%, β=0.28, p=.001) in CDB than in ADB. Second, the average duration of each representational gesture in CDB was 180 ms longer than in ADB (p=.002). Additionally, gestures in CDB were more salient (p=.006) (see Figure 2). These results demonstrated that broadcasters relied more on representational gestures and produced more salient gestures to convey their information to children. By contrast, broadcasters produced a higher proportion of pragmatic gestures in ADB (37.11%) than in CDB (29.69%) (β=0.44, p<.001). The differences were not significant for the use of beat gestures (CDB: 25.1%, ADB: 27.5%, β=-0.10, p=.053). However, broadcasters reduced the number of repetitions within a beat gesture in CDB (p=.012).

Table 3: Means (SD) and results (beta and p) of gestures for each condition.

<table>
<thead>
<tr>
<th>Dep variables</th>
<th>ADB</th>
<th>CDB</th>
<th>β and p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Repre_rate (per sec)</td>
<td>0.07(0.06)</td>
<td>0.10(0.08)</td>
<td>2.51***</td>
</tr>
<tr>
<td>Point_rate (per sec)</td>
<td>0.06(0.05)</td>
<td>0.08(0.07)</td>
<td>0.02***</td>
</tr>
<tr>
<td>Prag_rate (per sec)</td>
<td>0.19(0.11)</td>
<td>0.12(0.10)</td>
<td>-0.04***</td>
</tr>
<tr>
<td>Beat_rate (per sec)</td>
<td>0.11(0.06)</td>
<td>0.10(0.07)</td>
<td>-0.01</td>
</tr>
<tr>
<td>Average beat times</td>
<td>2.26(1.52)</td>
<td>1.94(0.89)</td>
<td>-0.32**</td>
</tr>
<tr>
<td>Mean duration of rep. (sec)</td>
<td>1.17(0.59)</td>
<td>1.35(0.54)</td>
<td>0.19**</td>
</tr>
<tr>
<td>Gesture saliency</td>
<td>6.32(0.97)</td>
<td>6.55(1.43)</td>
<td>0.23**</td>
</tr>
</tbody>
</table>

Note: ***p < .001, **p < .01, *p < .05.

Figure 2: A broadcaster explaining “two rabbits” in ADB (left) and CDB (right).

Discussion

This study provides a first insight into how broadcasters adapt their multimodal language to adults and in live broadcasting contexts. We found that broadcasters provided relatively less multimodal information when there was some assumed shared knowledge (i.e., addressing adult audiences). However, in the broadcast context, broadcasters need to prioritize the completion of broadcasting content, thus their language production is not always audience-centred (e.g., speaking rate; pausing; lexical diversity) (Table 4).

Table 4: The choice of recipient design in multimodal cue

<table>
<thead>
<tr>
<th>Multimodal Language</th>
<th>Features</th>
<th>Audience-centred</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MLU</td>
<td>✓</td>
</tr>
<tr>
<td>Linguistic features</td>
<td>Word frequency</td>
<td>✓</td>
</tr>
<tr>
<td>Prosodic cues</td>
<td>Speaking rate</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>Pitch</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>Intensity</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>pauses</td>
<td>ns</td>
</tr>
<tr>
<td>Gestures</td>
<td>Gesture type</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>Gesture rate</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>Beat gestures</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>Repeat times in a beat</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>Mean duration of rep.</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>Gesture saliency</td>
<td>✓</td>
</tr>
</tbody>
</table>

Note: ns = no significant evidence, 1 without including beat gestures; 2 without including beat gesture rate; 3 including the type of beat gestures and beat rate

Effects of common ground on recipient design

In line with previous research comparing CDL and ADL in non-broadcasting contexts, we discovered that broadcasters used shorter utterances, high(er)-frequency vocabulary (cf. Jaffe et al., 2001; Phillips, 1973; Snow, 1972) and more questions (cf., e.g., Newport, Gleitman, & Gleitman, 1977; Soderstrom et al., 2008), carried a higher pitch and more significant pitch variation (cf. Cotter, 1993; Grawunder et al., 2008; Medrano, Ferreira, & Behlau, 2005; Strangert, 2005), and produced more representational and pointing gestures (cf. Gogate, Bahrick, & Watson, 2000; Matatyahou-Bullaro et al., 2014; So et al., 2012) in CDB than in ADB. It has been claimed that CDS utterances tend to be shorter and grammatically simpler than those in ADS (Fletcher & MacWhinney, 1996; Saxton, 2009; Snow & Ferguson, 1977). When hosting a programme aimed at children, broadcasters used more common and frequent words as well as shorter utterances in CDB. Usually, children have a smaller vocabulary size than adults, which means that they may need help processing longer utterances and low-frequency words. Also, broadcasters must complete their speech in a limited time, so they may adopt shorter utterances and higher frequency words to make their speech easily understood by children while conveying as much information as possible.

There was a larger proportion of questions in CDB than in ADB. Questions used in CDS are a helpful type of input as they challenge children to reason and provide verbal
explanations (e.g., Duong et al., 2021; Purpura & Ganley, 2014; Tompkins et al., 2017; Rowe, Leech, & Cabrera, 2016). For example, caregivers’ use of questions is shown to relate to children’s unknown word learning (Dong et al., 2021). Although broadcasters cannot get audiences’ feedback when hosting a programme, they still care about what children might think of and respond to their speech, as children may have a poorer comprehension of news than adults. Moreover, even if the primary goal of broadcasters’ use of questions is not to get feedback or promote language acquisition, it can maintain or attract children’s attention. Children have smaller mnemonic spans (Cowan et al., 1991) and constraints in working memory (Hulme & Tordoff, 1989; Luciana & Nelson, 1998), which allow them to hold less information compared to adults—asking a question benefits learning because it is an open communication form that allows children to generate hypotheses (Yu et al., 2018). To keep their attention, despite raising their pitch, broadcasters can also ask questions to promote thinking and keep children engaged in watching. Questions can be used as a language strategy when broadcasters organize their speech.

Intonation used in broadcasting discourse is usually carried by a higher pitch and a larger pitch variation than in ordinary conversations (Price, 2008). In the context of CDB, the pitch level is even higher, and the pitch variation is larger when broadcasters deliver children’s programmes. However, we found an inconsistent finding with previous research. Researchers have argued that CDS has a wider pitch range than ADS (e.g., Fernald et al., 1989; Han et al., 2022; Narayan & McDermott, 2016), but in our study, CDB had a smaller pitch range than ADB. This is likely due to a ceiling effect that limited the broadcasters to enlarging the pitch maximum. The pitch maximum for CDB in our study was 0.47 semitones higher than ADB (p<.001), but the pitch minimum was 0.89 semitones higher than ADB (p<.001). The increase in pitch maximum in the CDB was not as significant as the pitch minimum, thus making the pitch range smaller when addressing children audiences.

Furthermore, broadcasters make use of gestures to help explain their speech. Campis and Özyırek (2013) found that the rate of iconic gestures increased for children when explaining how to use a coffee machine compared to when explaining to adults. Our study shows that broadcasters still produced more iconic and pointing gestures for the children’s programme, even though no addressee was presented. In addition, broadcasters often exaggerate the crucial features when hosting children’s programs. Besides raising pitch to attract their young audiences, they would also make gestures with larger sizes and extend the duration of representational gestures to make them clearer and more vivid to children. Also, pointing could map what they say with what is straightforwardly shown in the pictures to help children’s understanding. The common ground between speakers and addressees affects various aspects of prosody and linguistic contents (see Isaacs & Clark, 1987) and gestures.

**Effects of context requirement on recipient design**

In contrast to the above results, we also found broadcasters did not adapt their programmes in some features between the two conditions. This suggests that the language strategies broadcasters adopted were not always designed especially for audiences, they still needed to take the unique parts of the broadcasting context into consideration.

In linguistic features, broadcasters did not show a significant difference in lexical diversity between CDB and ADB, whereas earlier research showed that CDS was significantly less diverse (Henning, Striano, & Lieven, 2005; Hills, 2012; Kaye, 1980; Phillips, 1973). The inconsistency may be explained by language modality and context. For example, in written CDS, picture books for children contain more unique words than oral speech (Cameron-Faulkner & Noble, 2013; Hayes & Ahrens, 1988; Montag, 2019; Montag, Jones, & Smith, 2015). In book-reading contexts, picture-book provides more lexically sophisticated speech than other caregiver-child activities (Crain-Thoreson, Dahlín, & Powell, 2001; Salo et al., 2015; Sosa, 2016; Weizman & Snow, 2001). It could be the case that, like picture books, broadcasting programmes contain prepared drafts before live broadcasting. Even in our experiment, participants had two minutes to prepare. Also, the degree of formality between broadcasting contexts and ordinary conversations directed to children is different. Speech used in broadcasting programs is usually well-organized and prepared.

In prosody, broadcasters did not slow down their speaking rate or reduce the number/ rate of pauses, and pause duration in CDB. CDS usually has a slower speaking rate and longer pauses in many languages compared to ADS (e.g., Bernstein-Ratner, 1985; Fernald & Simon, 1984; Marklund et al., 2014; Sjons et al., 2017; Tang & Maidment, 1996). One explanation for the differences between our and past results may be that, unlike stress-timed languages (such as English), Mandarin is closer to the typological extreme of syllable-timed languages (e.g., Mok & Dellwo, 2008). In stress-timed languages, stressed and unstressed syllables are distinguished in terms of duration and syllable weight, while syllable-timed languages have nearly equal weight and time in all syllables. Also, the durational variability is more remarkable in stress-timed languages compared to syllable-timed languages (Grabe & Low, 2002). For example, Han et al. (2022) showed that the speaking rate of CDS is not consistently slower than ADS across languages, ages, or across the board. They compared Dutch and Chinese and found that Dutch CDS was generally slower than ADS, while Chinese CDS did not. Another explanation may be that speech on television news often displays at a fast pace and is lack of pauses compared to non-announcing speech (e.g., Bolinger, 1982, 1989; Castro et al., 2010; Van Leeuwen, 1984). The nature of broadcasting thus does not allow much room for the adjustment of speaking rate or pausing, especially since the time setting was 1 minute for both programmes. The design of multimodal utterances has been examined in both speech and co-speech gestures. For example, people increase both speech and gesture efforts in noisy environments (i.e., an increase in speech intensity and
a pitch shift, see Trujillo et al., 2021; Rasenberg et al., 2022),
and people use fewer words and gestures as common ground
increases (Holler & Bavelas, 2017). However, unlike pitch
modulation, speaking rate exhibits a trade-off between
efficiency and information transmission. A slower speaking
rate will sacrifice efficiency if a speaker is restricted to time.
The speaker will tend to speak faster and reduce the pause
duration to complete the speech under a given short-time
instruction, while pitch adaptation will not take notably more
time but only more effort.

Even though the proportions of beat gestures did not differ
significantly in the two programmes, broadcasters beat more
repeatedly and rapidly within a beat gesture in ADB. Beat
gestures are usually used to bring listeners’ attention to
certain highlighted parts and emphasize information in speech (Ekman, 1999; McNeill, 1992; Wang & Chu, 2013).
So et al. (2012) have found that beat gestures support memory,
but beats seem to provide benefits to preschool children only
when presented in pragmatically relevant contexts (Igualada,
Esteve-Gibert, & Prieto, 2017). For native speakers and
language learners, Rohrer, Delais-Roussarie, and Prieto
(2020) found that beat gestures had no robust effect on
information recall for native listeners and even led to
significantly lower comprehension and memory recall for
non-native language learners. When beat gestures are
repeatedly used in discourse, they inherently lose their
saliency as markers of important information (Rohrer et al.,
2020). Due to the limited cognitive ability, children, to some
degree, can be viewed as language learners analogous to adult
second language learners who also experience cognitively
more demands in an L2. Thus, too frequent beat gestures may
increase the perceivers’ cognitive load, which may hinder
comprehension. As broadcasters, encoding words with
appropriate beat gestures could help reinforce their content
and, consequently, better emphasize the important content for
their audience. However, broadcasters seem to avoid
repeating beat gestures too much in CDB as it may decrease
its power of marking information structure for children.

We found a novel result that broadcasters used more
pragmatic gestures in ADB but tried to avoid such gestures in
CDB. This may be because children do not use pragmatic
gestures as proficiently as adults (e.g., Dowling, 2022).
Unlike referential gestures such as iconic and deictics that
serve cognitive benefits (e.g., memory, So et al., 2012) for
audiences (Goldin-Meadow, 2003; Hadar & Butterworth,
1997; Kita, 2009), pragmatic gestures neither indicate or
iconically represent a referent in conversation, nor do they
have a conventionally agreed upon symbolic meaning (Ferré,
2012; Kendon, 2004, 2017; McNeill, 1992; Prieto et al., 2018;
Shattuck-Hufnagel et al., 2016). In particular, children’s
ability to understand pragmatic gestures is poorer than adults’
when a gesture’s meaning is abstracted from its context. It
has also been found that speakers produce pragmatic gestures,
relating not to lexical content but to aspects of complicated
interaction (Graziano & Gullberg, 2013). When talking to
adults, broadcasters are talking to more competent audiences
who share more common ground with them than children.

They may have more difficulties speaking to competent
audiences while speaking to children is more relaxed and
effortless.

Why do broadcasters adapt some features for their
audiences (audience-centred) but others not? As seen from
the discrepancies in the above results, broadcasters were not
constantly adapting to addressees. The distinguishing
features of broadcasting may help answer the question.
The primary goal of broadcasters is to produce their programmes
of a good “quality”. Therefore, they may give an equal weight
to adults and children in some features of their productions.
For example, the degree of diversity in their vocabulary and
how they highlight prominent content in their programmes
can also be viewed as factors contributing to good
programmes. The trade-off effect on efficiency and effort can
also play a role in adopting language strategies. Within the
time limitation, sometimes broadcasters must sacrifice
speaking rate or pause duration to complete their speech.
Finishing a speech on time also contributes to maintaining the
“quality” of live-broadcasting programmes. Besides, the
notion of global adaptation, which is highly linked to the
notion of expectation (Brennan, Galati, & Kuhlen, 2010),
may also contribute to the complexity of recipient design.
Global adaptation happens at the beginning of planning
communicative expression and is grounded on speakers’
expectations for different audience groups and individuals
(Campis & Özyürek, 2013). In our experiment, even though
broadcasters produced the same pictures for adults and
children, they may expect that their child audiences have less
background information about the pictures with them so they
would produce CDB with more iconic gestures, rely more on
higher-frequency words and shorter utterances. That is to say,
recipient design is affected by the addressee and there is a
direct relationship between different addressees and
broadcasters’ language choices. Furthermore, there are some
features that the adaptation to the audience design was not
significant. However, this does not mean that broadcasters do
not engage in audience design or can automatically consider
it as being speaker-centred. From a view of multimodal
production in broadcasting language, it is driven at least in
part by addressees, some of the signal channels are not
audience-oriented whereas at the same time, other signal
channels compensate for this through an audience design
(e.g., gesture duration vs. speech rate). Thus, the distinction
between speaker or audience orientation is not binary, but
their mechanism should be understood as the incorporation
of both processes simultaneously via different information
channels according to contexts.

In sum, our study made the first attempt to document
multimodal child-directed and adult-directed communication
in a broadcasting context. It provides a better understanding
of recipient design, which should be studied multimodally
and viewed holistically across signal channels. Moreover,
differences between CDS and CDB patterns could be
attributed to the role of child-directed broadcasting in
children’s learning and development, and future studies can
also expand to gender differences in multimodal broadcasting.
References


Han, M. (2019). The role of prosodic input in word learning: A cross-linguistic investigation of Dutch and mandarin Chinese infant-directed speech. LOT Publisher: Amsterdam.


