# The Development of Turn-Taking and Gaze Behaviour: A Literature Review

Bachelor Thesis Artificial Intelligence – 7,5 ECTS

**Eliza Chung, 5868491**
Faculty of Humanities, Utrecht University
Artificial Intelligence
(e.w.chung@students.uu.nl)

**Supervisor: Gijs Holleman**
(g.a.holleman@uu.nl)
**Second reader: Dr Chris Paffen**
(c.l.e.paffen@uu.nl)

(26 June 2020)

## Abstract

People interact every day with others, whether it be at home, at work, or at school. These spontaneous conversations are highly structured and run smoothly with some overlap and gaps occurring from time to time. How do conversational turn-taking and gaze behaviour develop? In this literature review, a concise overview will be given on the development of these interactional skills in caregiver-child interactions. First, concepts important to understanding turn-taking and gaze behaviour will be discussed: the turn-taking mechanism, types of overlaps during conversations, and the role of gaze in conversations. Then the development of interactional skills in infants, toddlers, and preschoolers. Finally, limitations, avenues for future research, and implications for AI will be discussed.

# CONTENTS

# 1 Introduction

People interact every day with others, whether it be at home, at work, or at school. These spontaneous conversations are highly structured and run smoothly with some overlap and gaps occurring from time to time (Sacks et al., 1978). While face-to-face interaction, characterized by 'back' and 'forth' exchanges between interlocutors, seems like an 'easy' skill that everyone performs effortlessly, researchers in cognitive science and psychology have long marvelled at the complexity of this behaviour. For researchers who aim to explain and understand the underlying mechanisms that make this face-to-face interaction possible, and how these interactional skills develop, this has been an important area of research: e.g. for (psycho)linguists, developmental psychologists, computer scientists, and sociologists (Levinson & Holler, 2019).

The developmental trajectory of interactional skills is necessary for face-to-face interaction, which leads to the main question of this thesis is: *how do these interactional skills develop*? Infants from the moment they are born are thrown into a world full of smells, sounds, and sights. And one of its first interactions is with their caregivers. In the literature on infant social development, it is recognised that certain infant-caregiver interactions are required to facilitate social, emotional, and cognitive development (Masataka, 2003). Concerning conversational turn-taking, it is seen as fundamental to social development in general. One of the reasons is that mother-infant interactions provide the infant with opportunities to learn the structure of conversations (Masataka, 2003). In essence, infant-caregiver interactions are considered to lay the foundation for the development of children's social skills. Aside from turn-taking, gaze plays an essential role in the development of children as well. The gaze helps them to learn when 'to take the floor' (i.e. their turn to talk) and gaze following has also been shown to predict productive vocabulary later in life (Rutter & Durkin, 1987; Tenenbaum et al., 2015).

The thesis is structured as follows: first, an overview of conversation analysis and its concepts relevant to this thesis will be given. Then, the development of these conversational skills will be given in chronological order, from infancy to early childhood. At each stage of development, the timing of vocalization, overlaps, and gaze behaviour will be covered (if possible) and related to milestones in their linguistic development. Although the development of turn-taking and gaze behaviour continues after early childhood, empirical data of turn-taking research from middle childhood (7-11 years) until (pre)adolescence (12-18 years) is scarce, or almost non-existent. Thus, those age groups cannot be included. Finally, limitations, avenues for future research, and implications for AI will be discussed.

## 1.1 Relevance in AI

Robots are booming and there are already many types, all with their specific applications. Examples of such robots include contact assistive robots, which assist people through physical contact; or social interactive robots that entertain through social interaction; and also socially assistive robots (SARs), which goal is to assist with social interaction (Feil-Seifer & Mataric, 2005).

To focus on social robots and socially assistive robots, they are designed to interact with people in a natural and interpersonal manner to achieve positive outcomes in diverse applications such as health, quality of life, education, and therapeutics. (Breazeal et al., 2016). What components are crucial to enable and facilitate these interactions? Besides needing cognitive skills, social abilities are crucial and have to be implemented into the robot as well, including turn-taking and gaze behaviour. Turn-taking skills are vital because, without them, natural communication cannot be achieved (Chao & Thomaz, 2010). And just like in human-human interactions, where gaze behaviour facilitates turn-taking and timing (more details in section 2.3), the same applies to human-robot interactions where the robots have gaze behaviour implemented (Sciutti et al., 2015). Sato and Takeuchi (2014) even found that robots can coordinate multi-party conversations by controlling the robots' eye gaze behaviour.

Another application of robots is, as mentioned, in therapy. These SARs are often applied to assist children with autistic spectrum condition (ASD). An example is Kaspar, which helps such children to learn fundamental social skills such as imitation, turn-taking, and joint-attention – which are the skills that children with autism find challenging (Robins et al., 2018).

To conclude, with social robots and SARs being utilised more and more, insight into turn-taking and eye gaze behaviour is key. That way, it can be applied in the robots to build (humanoid) robots to achieve natural communication and in the case of SARs, it can improve social skills in children with ASD.

# 2 Background information: conversation analysis

So what is CA? It is the study of talk which emerged in the 1960s. The emergence of the field is credited to Harvey Sacks, Emanuel A. Schegloff, and Gail Jefferson. To be more specific, CA is the systematic analysis of the talk produced in everyday situations of human interaction: talk-in-interaction (Hutchby & Wooffitt, 1998). The methodology consists of analysing recordings of naturally occurring interactions. So the recorded interactions are preferably not arranged or set up in laboratories but take place in the mundane lives of people. Those recordings are then (painstakingly) transcribed by the researchers by using a set of conventions. Therefore, CA is the study of recorded, naturally occurring talk-in-interaction. CA aims to ultimately understand how conversational participants understand and respond to one another, with the focus on how sequences of actions are generated: turn-taking. Although the field is referred to as 'conversation analysis', the subject of analysis is not limited to everyday conversations, but also those interactions that occur in a formal or non-formal institutional context. Formal examples include courts of law, interviews, broadcasts, ceremonial occasions, but also job interviews. Non-formal types include more loosely structured, but still task-oriented, lay/professional situations such as counselling sessions, social work and service encounters, radio phone-in conversations, and doctor-patient interactions.

## 2.1 Turn-taking mechanism

So CA aims to understand how people take turns in talk-in-interaction, but what are the underlying mechanisms? Sacks and colleagues (1978) described the turn-taking system for conversations in terms of two components, the turn-constructional component and the turn-allocation component, and a set of rules. The turn-constructional component is a lexical component, which refers to each person's point of view on how to start and continue speaking, thus how a turn is filled. Turn-construction units (TCU) are used to compose a turn and include sentential, clausal, phrasal, and lexical constructions. The end of a TCU is a point where the turn may end and the next speaker may begin. The turn-allocation component is responsible for distributing a turn. This mechanism includes any kind of signal, either verbal or non-verbal, to indicate it is the listener's turn to start speaking. This signal could be a cough, a word, a sound, or a look. The set of rules govern turn construction, provides for the allocation of a next turn, and coordinates transfer to minimise gaps and overlaps between speakers. To conclude, turn-taking involves listening to the speaker, predicting the end of the turn, preparing a response, and articulate this response at the appropriate moment (Corps et al., 2018).

## 2.2 Overlaps

Conversations are not always perfectly timed, which causes overlaps and interruptions to occur. Schegloff (2000) introduced four types of overlapping. First are the *terminal overlaps*, which occur when the listener starts to speak assuming the speaker has or was about to finish, thus creating an overlap. The second type is a *continuer*, which is a listener's feedback to the speaker by backchanneling. A backchannel is a response that can be verbal, non-verbal, or both, such as head nods and interpolations as *uh-huh* and *right*. A continuer indicates that the listener understands the speaker and encourages the speaker to continue. The third type is called a *conditional overlap*¸ which implies that the speaker invites the listener to fill in or complete the turn. For example, in word search instances. The fourth and last type is called *chordal* and it implies that the conversational partners make vocalizations simultaneously, such as laughing together or congratulating someone on finishing their thesis. All in all, overlaps are not always problematic and can encourage the speaker to continue or support the speaker when s/he experiences a tip-of-the-tongue moment.

## 2.3 Cues for turn changing: eye gaze

Now that we know the underlying mechanism of turn-taking, what cues facilitate the whole turn-taking process? Important cues for turn-taking include eye gaze, speech prosody, the grammatical structure of the utterances, body language, and timing (Jokinen et al., 2010). For now, we will focus on eye gaze. Mutual gaze, which is the term for two people making eye contact, facilitates the turn coordination process and is needed for successful turn change (Jokinen et al., 2013). At the point of a turn transition, the current speaker and next speaker might look at each other for a short moment, which improves the accuracy of turn-taking, compared to using only speech features or dialogue acts (Jokinen et al., 2013). Breaking of mutual gaze is associated with the listener not wanting to take the next turn or accepting the next turn (Ho et al., 2015). It has been observed that when the listener takes the next turn, she, he, or it breaks the mutual gaze once the planning of the utterance start. This gaze breaking behaviour is related to increased cognitive load. In conclusion, eye gaze indicates where the speaker's attention is directed and facilitates turn-taking between conversational partners. Generally, speakers end their turn with a direct gaze to the listener and the listener, in turn, begins to speak with averted gaze.

# 3 Development of conversational turn-taking and gaze behaviour

Now we will start with the main question of this thesis: how does turn-taking and gaze behaviour develop throughout the early moments of life?

## 3.1 Infancy (0-12 months)

At this prelinguistic stage, one of its first interactions is with its parents, where infants even before the onset of speech show the ability to take turns in interactions (Levinson, 2016). These caregiver-infant interactions have been called protoconversations and include both turn-taking and overlapping vocalizations. The speech-like vocalizations, which are precursors to speech, have been dubbed protophones (Gratier et al., 2015; Yoo et al., 2018). During early infancy, these protophones could be squeals, vocals, and growls (Oller et al., 2019). The newborn's vocal production is modulated by the presence of maternal vocalization (Rosenthal, 1982). Babies are more likely to vocalize while the mother is speaking.

### 3.1.1 Infant turn-taking

Although there is a consensus that infants first turn-like structures appear at around 2 months, one study by Dominquez et al. (2016) has shown that neonates show the ability to take turns (Casillas, 2017). Dominquez et al. (2016) found that newborns ranging from 2- to 4-days old actively participate in interactions, by creating interactive sequences with their partner, and that these sequences have a turn-taking format. The most common turn-taking sequence involved 2 turns followed by sequences of 3 turns. It became clear that two-thirds of the neonatal vocalizations that follow a maternal vocalization occurred within one second and that the proportion (almost one-third) of latched vocalizations was relatively high. Latched turns are defined as two alternating vocalizations separated by a pause lasting less than 50 ms. These latched turns suggest that neonates anticipate the end of the mother's vocalization with a certain degree of precision. Furthermore, of all newborn vocalization analysed, one-third overlapped with a maternal vocalization. The limitation of this study was that the mothers in the maternity ward were instructed to talk to their infant freely but to avoid using toys or making noise with objects for 10 minutes. So it was not exactly in a natural setting, as the instructions lead to high response and probably unrepresentative rates. However, one could say that based on Dominquez' research a neonate can be prompted to engage in a protoconversation.

In the first weeks after birth infants start to coo and around 4 months old, they start to laugh (Harley, 2015). After around six months, babies start to make speech-like sounds known as babbling, which are characterised by syllables (e.g. "ba" and "ma") and often repeated a few times. How has turn-taking developed now that the infant is a few months old? Gratier et al. (2015) studied the

developmental changes in turn-taking of infants at 2-3 months old and 4-5 months old. At this moment in their development, the most common turn-taking format involves three turns. Non-latched vocalizations occurred within one second as well, with an average of 730 ms. It was also found that the older infants showed switching pauses that were on average 174 ms longer than the younger infants. Latched turns seemed to occur more frequently in 4-5 months old infants than in those that were 2-3 months old, but the difference was not significant. Of all infant produced turns in Gratier's study, 44.5% were latched, in contrast to 26.9% of the latched turns produced by neonates as mentioned before. In terms of overlapping, there was no significant difference between the two age groups where more than one-third of the infant vocalizations overlapped with the maternal vocalization.

Another interesting study with 3- and 6-month old babies was conducted by Striano et al. (2006) who assessed infant and maternal sensitivity to the timing in interaction. The infants interacted with their mothers over television and the amount of time they gazed at each other during real-time live interaction was compared to their behaviours in an interaction where the audio-visual presentation of the mother or the infant was temporally delayed by 1 second. Infants at both ages were able to detect the 1-second delay in maternal responses, which was indicated by a decrease in visual attention and positive affect (smiling and positive vocalizations). The decrease in visual attention is explained by the fact that infants have formed expectations regarding conversations from birth and expect their mothers to respond to their dyadic behaviour (Striano et al., 2006). Thus, the infants might have lost interest in the interaction when the maternal timing differed from normal situations. In summary, the timing of responses (vocalizations), latched turns, and the amount of overlap with maternal vocalizations stays relatively stable from birth to 3-6 months old. And at 3 months old infants are already sensitive to the timing of social interactions.

Around nine months, infants demonstrate the ability to understand that certain sounds occur regularly in certain situations, and perhaps a few words already as well (Harley, 2017). From then on, they start to produce their first recognisable words. Does their linguistic development influence their turn-taking? Hilbrink et al. (2015) examined the gap of infants longitudinally, and in this section the results of 9- and 12-month olds will be discussed (see figure 1 and 2). The turn-timing of 9-month-old infants' response increased significantly to almost 1150 ms (median) compared to the first 6 months, of which an explanation will be given later. At twelve months old, the (median) gap decreases again to around 1000 ms. In terms of overlapping vocalizations, the infants show improvement to just under one-third of the vocalization overlapping with their mothers between 9 and 12 months. Unfortunately, in this age group, the latched turns were not measured.

One would expect that the slowdown would coincide with the emergence of language production at 12 months because at that moment infants must coordinate these developing linguistic skills with their existing turn-taking skills. However, the slowdown occurs earlier at 9 months. As mentioned before, at 9 months old infants start to understand that certain sounds and words occur regularly at certain situations, but other social skills relevant to communicating are emerging as well: joint attention and pointing (Hilbrink et al., 2015). Furthermore, 9-month-old infants start to see others as intentional agents, which is suggested to be a prerequisite for word learning. Another explanation could be that the increase in turn-timing and decrease in overlap might play a role as well, but that has not been investigated yet. In conclusion, infants at this age start to understand the functions of social interactions and combined with developing linguistic skills, it is hypothesized that these factors play a role in the slowdown of timing.

To conclude the turn-taking behaviour of 0-12-months-old infants: neonates show the ability to take turns when prompted to engage in protoconversations, based on the study by Dominquez et al. (2016). The turn-timing and amount of overlap stay relatively stable until they are 6 months old, where it then steadily seems to increase. Around 9 months, the turn-timing of infants increases significantly which could be explained by the emergence of social skills, developing language understanding, and perhaps coincides with decreasing overlap. At 12 months, their turn-timing improves a bit, where the overlap stays the same as at 9 months old. It should be realised that the maternal timing remained stable across the study of Hilbrink (2015), so the changes in infants' turn-taking skills were independent of maternal timing.

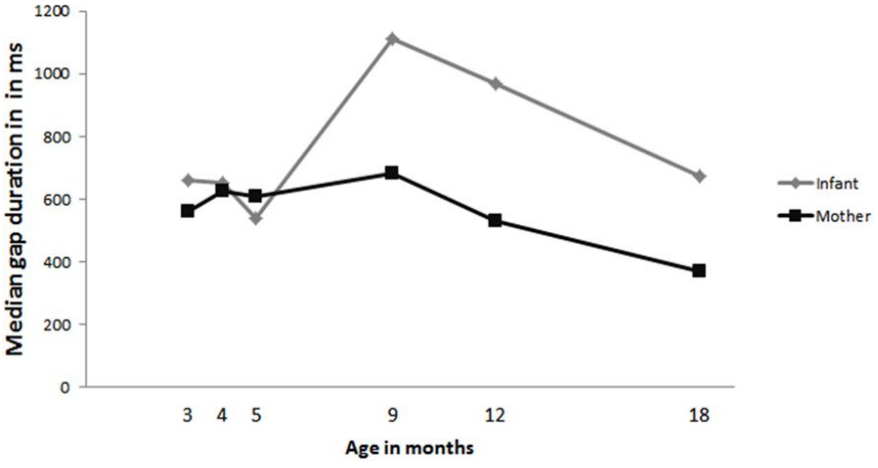


*Figure 1: The median gap of infants age 3 to 18 months old (Hilbrink et al., 2015)*

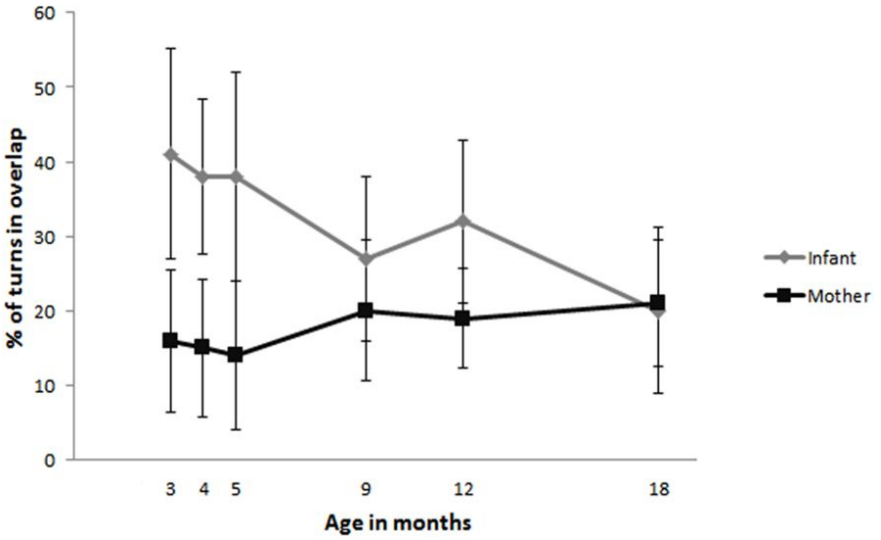


*Figure 2: The percentage of turn produced in overlap (Hilbrink et al., 2015)*

### 3.1.2 Infant gaze behaviour

Regarding gaze behaviour, it seems that during early infancy the gaze is the infants' way of engaging in protoconversations and maintain its participation (Nomikou et al., 2016). During these early interactions, mothers are very responsive to the gaze of the infant and react accordingly. It is suggested that these maternal responses reinforce the behaviour of the infant which set the foundation for protoconversations and thus the development of social interactional skills (Nomikou et al., 2016).

Neonates as young as 1 day old are known to gaze significantly longer at faces with opened eyes than faces with closed eyes (Batki et al., 2000). What this means for neonatal gaze behaviour during dyads with caregivers does not seem to be investigated yet and most studies conducted investigate gaze behaviour starting at around 2/3 months old. Batki (2000) lists some milestones in gaze behaviour of infants: around 2 months old infants, just like neonates, show a preference for looking at the eyes over other regions of the face. By 4 months, infants can discriminate between direct and averted gaze and by 5 months, infants look longer at faces showing direct eye-contact. Infants at 3 months old also turn their eyes to a given target faster when the location of that target has been previously cued by the direction of an adult's gaze.

*Mutual gaze*

Northrup and Iverson (2020) have conducted a longitudinal examination of mother and infant vocal and gaze coordination during toy play interaction across the first year, taking measurements at 3, 6, 9, and 12 months. We will focus on the findings regarding mutual gaze behaviour directed at the partner and not at objects. In short, the amount of time infants and mothers spent in mutual gaze decreases dramatically over across the first year, but the occurrence increases. So although infants

and mothers reduce the time spent looking at each other's faces, they time their moments of mutual gaze better, which suggests that mutual gaze continue to play an essential part during social interactions.

*Gaze as cue*

In terms of gaze behaviour during mother-infant dyads, Rutter and Durkin (1987) conducted a longitudinal study of infants starting from 9 months old. They analysed the onset of the infants' looks at the mothers, which were divided into six categories: at the start, at the middle, and the end of the infant's turn; and correspondingly, the start, the middle, and the end of the mother's turn. "At the start" and "at the end" meant that the look had to be in progress when the vocalization began or ended, and "in the middle" meant that the look occurred at some other point during the turn. The findings in 9- and 12-month-olds are as follows: in terms of infants' gaze during their turns, a slight increase in looking at the mother during the start and end of their turns can be noticed, whereas gazing at the mother in the middle of their turn increases slightly, although the difference between the two ages is minimal (figure 3 and 4). In terms of infants' gaze during the turn of mothers, most of the looks occurred near the end of the mother's turns (figures 3 and 4).

To conclude, although it isn't as accurate yet as in adults, 12-month-olds do seem to demonstrate adult eye gaze behaviour, as mentioned in section 2.3, when they look toward the conversational partner nearing the end of their turns (Rutter & Durkin, 1987). Likewise for gazing towards the end of the conversational partner's turn.

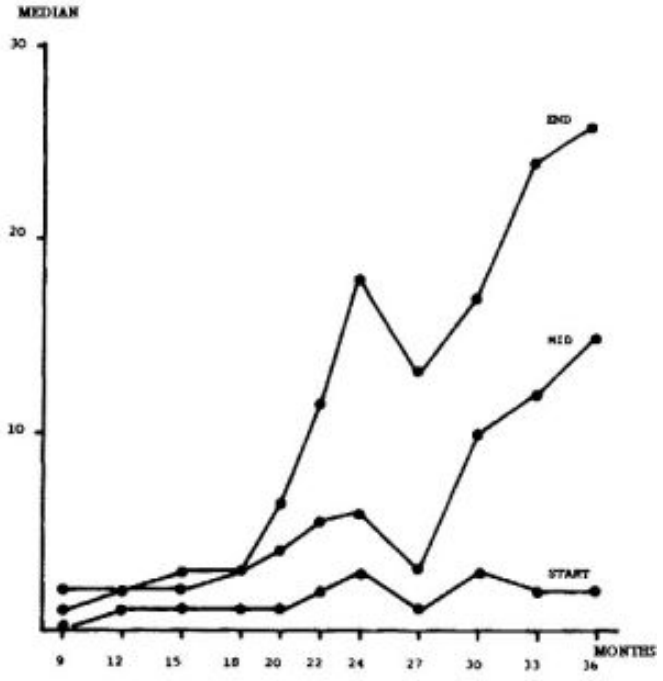


*Figure 3: The distribution of infants' gaze across infants' turns (Rutter & Durkin, 1987)*

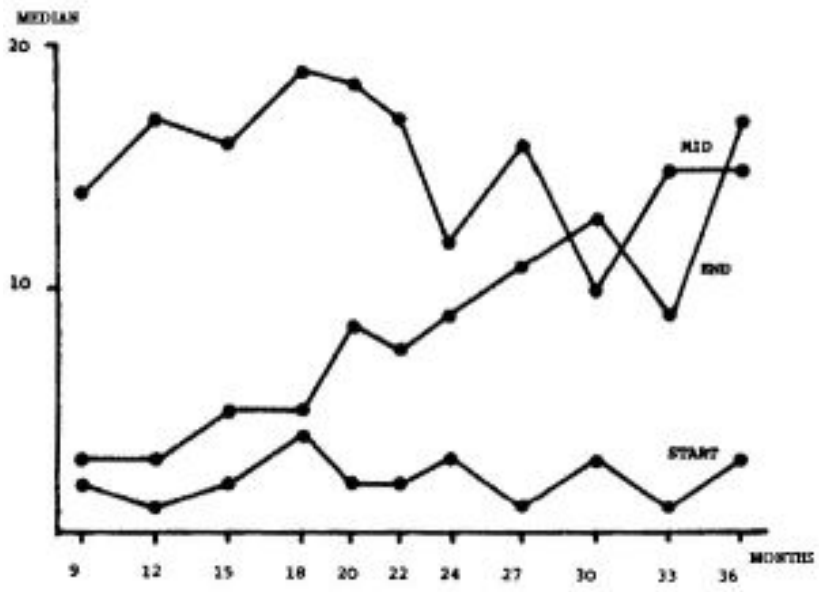


*Figure 4: The distribution of infants' gaze across mothers' turns (Rutter & Durkin, 1987)*

## 3.2 Toddlerhood (1-3 years old)

At around 18 months, a *vocabulary spurt* occurs where toddlers' vocabulary quickly expands. The usage of this newfound vocabulary at the early stage is paired with short utterances, which lack grammatical structures and are abbreviated in nature, such as "bye-bye dog" (Harley, 2017). As they grow, children acquire the finer syntactic details of their language. By two and a half, they start to speak like little adults. Language develops after this point still: syntax improves even further, and vocabulary continues to increase throughout life, although at a slower rate.

### 3.2.1 Toddler turn-taking

Continuing with the study of Hilbrink et al. (2015) toddlers' turn-timing at 18 months continue to improve compared to 9-12 months old where the gap decreases from 1000 ms at 12 months to around 700 ms, reaching almost the speed of a 5-month-old (see figures 1 and 2). In terms of overlap, around one-fifth of the toddler vocalizations overlap with maternal vocalizations, which is around adult level and less than at 12 months.

A factor that influences the turn-timing is the complexity of questions toddlers have to answer, of which Tice et al. (2011) examined the development. The toddlers were between age 1;8 to 3;4 (year; month) and there were simple and complex questions. Simple questions were yes/no questions, whereas the more complex questions were the *wh*-questions (what/why/who/where). What they found was that the older the toddlers became, the faster they were able to answer the questions. Overall, toddlers were faster in answering the simple questions and when the answer was mentioned in preceding utterances. Within the *wh*-questions, they were slower to answer *who*

questions than *what/where* questions, which is consistent with children's order of acquisition for wh-words (Tice et al., 2011). So although the toddlers' gap shrank with age, the complexity of the questions influenced the duration of the gap.

Rutter and Durkin (1987) conducted a longitudinal study examining vocalizations and gaze until 36 months. They examined whether toddlers would interrupt (overlap) less as they grow older but unfortunately did not analyse the timing of vocalizations. As shown in figure 5, the number of overlaps steadily increase until around 24 months, but then decrease sharply for both the infants and the mothers, and then increase again around 30 months. In line with the findings of Hilbrink et al. (2015), toddlers produce relatively less overlaps the older they become. What was interesting is that Rutter and Durkin did not seem to find an increase in overlap between 9- and 12-month old infants, but that could be explained by the fact that Rutter and Durkin did not differentiate between the amount of turns mothers and infants made – they took the median of produced turns of both mother and infant together (see table 1).

Beyond 24 months old, one can see a significant change. Whereas the number of turns continued as before, the frequency with which children overlapped with their mothers fell dramatically, whereas the rate of maternal overlap remained the same (see table 1). The implication, therefore, is that the third year is the time when the baby first begins to play a significant part in controlling the sequencing of vocalizations.

To conclude toddler turn-taking, from 12 to 18 months, their timing improves. When answering questions, the complexity of the questions will influence the duration of toddlers' replies. In terms of overlap, toddlers produce more in overlap with their mothers, but the frequency in which they produce overlap decreases because more turns are taken, thus with age toddlers overlap less and less (see table 1).

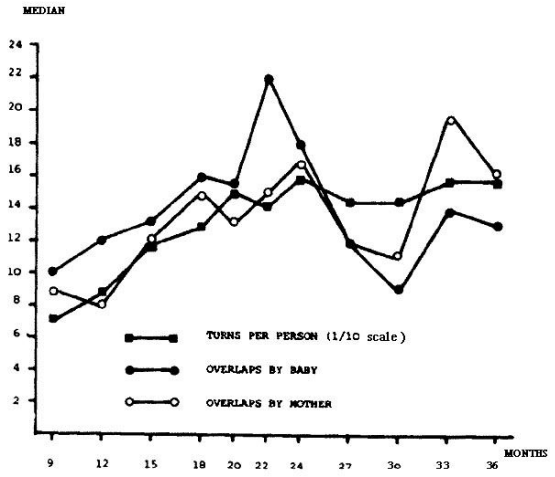


*Figure 5: The median values of turns and overlaps the sessions (Rutter &Durkin, 1987).*

| Variable | Age in months | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 9 | 12 | 15 | 18 | 20 | 22 | 24 | 27 | 30 | 33 | 36 |
| **Turns per person** | 70 | 87 | 117 | 128 | 150 | 141 | 159 | 145 | 144 | 158 | 158 |
| **Mothers' overlaps** | 9 | 8 | 12 | 15 | 13 | 15 | 17 | 12 | 11 | 20 | 16 |
| **Babies' overlaps** | 10 | 12 | 13 | 16 | 15.5 | 22 | 18 | 12 | 9 | 14 | 13 |
| **% babies' overlap/turns per person** | 14.3 | 13.8 | 11.1 | 12.5 | 10.3 | 15.6 | 11.3 | 8.3 | 6.3 | 8.9 | 8.2 |

*Table 1: median values of turns, overlaps, and percentage overlaps produced by the infants across age (Rutter and Durkin, 1987).*

### 3.2.2 Toddler gaze behaviour

As mentioned before, Rutter and Durkin (1987) analysed gaze behaviour beyond infancy into toddlerhood. What becomes apparent is that toddlers from 12 months old on increasingly start to look at their mothers' faces nearing the end of their turns: nine times as many at 2 years as at 1 year and 13 times as many at 3 years as at 1 year (see figure 3). Looking at the mother's face at the end of her turn also significantly increased starting from 18 months. A correlation was found between gazing nearing the end of a mother's turn and infant overlap. Just as mentioned in section 2.3 on eye gaze acting as a cue, so too, do toddlers receive cues as to when they can take over the turn and overlaps decrease. So by the end age of 2, toddlers start to demonstrate gaze behaviour on adult levels.

## 3.3 Early Childhood (3-6 years old) and beyond

Where during infancy and toddlerhood most analyses were taken from free play interactions between mother and infant/toddler, with the older age group that is less the focus of examining turn-taking and gaze-behaviour development. Children in this age group, also called preschoolers,

are still learning to use complex language and can hold conversations, albeit still not at adult-level speed (Lindsay et al., 2019). The focus is more on how the conversational context influences the turn-taking speed and gaze behaviour, such as answering simple or complex questions, and the ability to coordinate language comprehension and production. So this section (3.3) will be different from the sections on infancy and toddlerhood.

### 3.3.1 Child turn-taking

As mentioned in section 2.1, turn-taking involves listening to the speaker, predicting the end of the turn and at the same time preparing a response, and articulate this response at the appropriate moment. Lindsay et al. (2019) tested how this ability to coordinate comprehension (listening to the speaker) and production (preparing the answer and respond at the right moment) develops in preschool children (ages 3-5 years) and compared it to the response times of adults. The instructions for the children were as follow: the goal of the game they were going to play was to help Peter Pan navigate a set of mazes while searching for either one or two animals and that they should answer his questions as quickly as possible. The yes/no questions were varied in predictability (e.g. early: “Is Po (a fictional character in the game) hiding the animal?” vs. late: “Is the animal behind Po?”) but were controlled for complexity. The response time was then measured. What Lindsay et al. found was that when children were able to predict a question’s ending, they then left shorter gaps before responding, but compared to adults, they were still slower. As read in the previous section on toddler turn-taking, toddlers were faster when answering yes/no questions, so why were the children in Lindsay’s study still slower than adults while their answer was simply a yes or no? These findings do suggest that factors other than linguistic planning difficulties may contribute to the slower responses of children’s turn-taking. Lindsay et al. propose thus that more general factors might be at play, such as children’s ability to switch from comprehension to formulation.

In terms of overlap, Bedrosian et al. (1988) examined overlap and other conversational violations in mother-child interactions during free play situations. Children aged 2;10 to 6;3 (year; month) were included in the study. Unfortunately, the data of the number of overlaps per age group was not given in the study, but a correlation was found between the children’s age and frequency of overlap: overlaps decreased significantly with age.

To conclude, although children are still learning to use complex language, they can flexibly coordinate language comprehension and production, which result in the optimization of their turn-timing, and overlaps continue to decrease.

### 3.3.2 Child gaze behaviour

As previously mentioned in section 2.3, gaze eversion is not only related to the switching of turns but also increased cognitive load. And to continue with the theme of children's interactional behaviour when answering questions, Doherty-Sneddon et al. (2002) found that children show gaze aversion behaviour when the cognitive load increased. Children aged 5 and 8 years old were included in this study and the questions they had to answer were verbal or arithmetic and were either easy or hard.

The first finding was that the older children averted their gaze more than the younger children for difficult questions. The second finding was that 5-year-olds were less consistent with their gaze aversion to cognitive difficult questions, where only difficult verbal questions showed a significant effect. These 5-year-old children increased aversion while planning their response to difficult verbal questions, but not while responding, unlike 8-year-olds who averted their gaze both during planning and responding. So gaze aversion while responding to difficult questions is acquired with age. The third finding was that 8-year olds averted their gaze more when thinking and speaking compared to listening. 5-year-olds showed similar aversion behaviour when thinking, but not when speaking. What Doherty-Sneddon also found was that there was no correlation between accuracy of response and gaze aversion – so aversion did not influence performance. The fact gazes were averted in response to both difficult and arithmetic question confirmed that gaze aversion functions to reduce cognitive load (i.e. less visuospatial demands to compete over brain resources).

To conclude, gaze aversion develops over time and is used in a similar way as adults: to reduce cognitive load when answering (difficult) questions.

# 4 Discussion

This thesis aimed to answer the question of how interactional skills develop in the first years of life. The focus was in particular on the development of turn-taking and gaze behaviour in caregiver-child interactions. To accomplish this, a literature review was performed to create an overview of the development of interactional abilities in different from infancy, to toddlerhood, and finally early childhood.

As you might have noticed, the further we explored the development of turn-taking and gaze behaviour of older age groups, the more different the focus and methodology was. That is related to the development of the child: during early and mid-infancy, they cannot vocalise their intentions and thoughts but do show the ability to engage in protoconversations. Thus, investigating the temporal organization of their vocalizations reveals the development of the ability to predict the end of turns. As they grow older, they start to talk and have to coordinate language comprehension and production, but other social skills emerge as well. That is reflected in their turn-timing (gap), which does not improve linearly. The turn-timing stays relatively stable from birth until around 6 months, then the turn-timing becomes considerably slower with the peak at 9 months, after which it slowly improves again over the next few years. Having to coordinate language comprehension and production considerably influences their timing, as difficult questions require a more complex answer, thus slows down a child's reply. In terms of overlap, children improve with age. Whether they reply slow or interrupt, in dyads with caregivers, children get the chance to learn and improve their turn-taking skills.

Interactions aren't only about turn-taking, but also about gaze – which facilitates turn-taking. In infancy, babies are known to stare more at eyes than any other part of the face. The amount of time infants and mothers spent in mutual gaze decreases dramatically over across the first year, but the occurrence increased, which suggest infants start to time their gaze better. Just like adults, the older children are, the more they look towards their conversational partner during the end of the turns, as opposed to during the beginning or in the middle of the turn. Gaze aversion plays a role as well, which develops during early childhood. Children avert their gaze when answering difficult questions, and especially when thinking about and producing the answer.

This thesis focused on conversational turn-taking and gaze behaviour, however, many other factors guide the course of conversations, such as speech prosody, the grammatical structure of the utterances, body language and more. They couldn't all be included, because discussing all factors would not fit in a thesis of this size. Humans converse almost effortlessly every day, yet when trying

to research it, it becomes apparent how complicated it is (which applies for other phenomena as well, such as trying to replicate consciousness in humanoid robots).

As mentioned in the introduction, the age groups after early childhood were supposed to be included as well, but because there is a gap in the literature between middle childhood and adolescence, that was not possible. One explanation could be that after early childhood, the changes in turn-timing and gaze behaviour during dyads change subtly so that conducting research would not show significant results. However, considering that as children enter their teenage years and require more autonomy and freedom, the way of interacting with their parents change. It is often characterised by conflict between parents and teenagers (Mastrotheodoros et al., 2020; myself). Although there is more conflict with parents, at the same time peer and friend interactions become more relevant in the lives of teens and adolescents. Perhaps these factors (conflict and friend interactions) influence turn-taking and gaze behaviour, so these could be interesting contexts for future research in conversation analysis.

Finally, what are the implications for AI? Section 1.1 touched upon social robots and the reason why turn-taking and gaze behaviour have to be implemented in robots: it allows for human-like interactions between humans and robots. Especially with robots being used in educational and interventional settings (ASD intervention), the ability to hold human-like conversations and being able to adapt to the conversational partner (whether it be a young child and the usage of child-directed speech or senior) will be more and more important.

This thesis aimed to provide a concise overview of the development of conversational turn-taking and gaze behaviour across infancy, toddlerhood, and early childhood. Although it is far from complete due to the reasons mentioned above, it still covered the development of two components that are relevant to human interactions: turn-taking and gaze behaviour.

# References


Batki, A., Baron-Cohen, S., Wheelwright, S., Connellan, J., & Ahluwalia, J. (2000). Is there an innate gaze module? Evidence from human neonates. *Infant Behavior and Development*, *23*(2), 223-229.

Bedrosian, J. L., Wanska, S. K., Sykes, K. M., Smith, A. J., & Dalton, B. M. (1988). Conversational turn-taking violations in mother-child interaction. *Journal of Speech, Language, and Hearing Research*, *31*(1), 81-86.

Breazeal, C., Dautenhahn, K., & Kanda, T. (2016). Social robotics. In *Springer handbook of robotics* (pp. 1935-1972). Springer, Cham.

Casillas, M., & Frank, M. C. (2017). The development of children's ability to track and predict turn structure in conversation. *Journal of Memory and Language, 92*, 234-253.

Chao, C., & Thomaz, A. L. (2010). Turn taking for human-robot interaction. In *2010 AAAI Fall Symposium Series*.

Corps, R. E., Gambi, C., & Pickering, M. J. (2018). Coordinating utterances during turn-taking: The role of prediction, response preparation, and articulation. *Discourse Processes*, *55*(2), 230-240.

Doherty-Sneddon, G., Bruce, V., Bonner, L., Longbotham, S., & Doyle, C. (2002). Development of gaze aversion as disengagement from visual information. *Developmental psychology*, *38*(3), 438.

Dominguez, S., Devouche, E., Apter, G., & Gratier, M. (2016). The Roots of Turn-Taking in the Neonatal Period. *Infant and Child Development, 25*(3), 240-255.

Feil-Seifer, D., & Mataric, M. J. (2005). Defining socially assistive robotics. In *9th International Conference on Rehabilitation Robotics, 2005. ICORR 2005.* (pp. 465-468). IEEE.

Gratier, M., Devouche, E., Guellai, B., Infanti, R., Yilmaz, E., & Parlato-Oliveira, E. (2015). Early development of turn-taking in vocal interaction between mothers and infants. *Frontiers in psychology, 6*, 1167. https://doi.org/10.3389/fpsyg.2015.01167

Harley, T. A. (2017). Children. In *Talking the talk: Language, psychology and science*. Psychology Press.

Hilbrink, E. E., Gattis, M., & Levinson, S. C. (2015). Early developmental changes in the timing of turn-taking: a longitudinal study of mother-infant interaction. *Frontiers in psychology, 6*, 1492.

Ho, S., Foulsham, T., & Kingstone, A. (2015). Speaking and listening with the eyes: gaze signaling during dyadic interactions. *PloS one*, *10*(8), e0136905.

Hutchby, I., & Wooffitt, R. (1998). *Conversation analysis*. Polity.

Jokinen, K., Harada, K., Nishida, M., & Yamamoto, S. (2010). Turn-alignment using eye-gaze and speech in conversational interaction. In *Eleventh Annual Conference of the International Speech Communication Association*.

Jokinen, K., Furukawa, H., Nishida, M., & Yamamoto, S. (2013). Gaze and turn-taking behavior in casual conversational interactions. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, *3*(2), 1-30.

Lindsay, L., Gambi, C., & Rabagliati, H. (2019). Preschoolers optimize the timing of their conversational turns through flexible coordination of language comprehension and production. *Psychological science*, *30*(4), 504-515.

Levinson, S. C. (2016). Turn-taking in human communication–origins and implications for language processing. *Trends in cognitive sciences, 20*(1), 6-14.

Mastrotheodoros, S., Van der Graaff, J., Deković, M., Meeus, W. H., & Branje, S. (2020). Parent–adolescent conflict across adolescence: trajectories of informant discrepancies and associations with personality types. *Journal of youth and adolescence*, *49*(1), 119-135.

Masataka, N. (2003). The development of the ability to take turns. In *The Onset of Language* (Cambridge Studies in Cognitive and Perceptual Development, pp. 27-63). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511489754.002

Nomikou, I., Leonardi, G., Rohlfing, K. J., & Rączaszek-Leonardi, J. (2016). Constructing interaction: the development of gaze dynamics. *Infant and Child Development*, *25*(3), 277-295.

Northrup, J. B., & Iverson, J. M. (2020). The development of mother-infant coordination across the first year of life. *Developmental psychology*, *56*(2), 221–236. https://doi.org/10.1037/dev0000867

Oller, D. K., Caskey, M., Yoo, H., Bene, E. R., Jhang, Y., Lee, C. C., ... & Vohr, B. (2019). Preterm and full term infant vocalization and the origin of language. *Scientific reports, 9*(1), 1-10.

Robins, B., Dautenhahn, K., & Nadel, J. (2018). Kaspar, the social robot and ways it may help children with autism–an overview. *Enfance*, (1), 91-102.

Rosenthal, M. (1982). Vocal dialogues in the neonatal period. *Developmental Psychology, 18*(1), 17.

Rutter, D. R., & Durkin, K. (1987). Turn-taking in mother–infant interaction: An examination of vocalizations and gaze. *Developmental psychology*, *23*(1), 54.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1978). A simplest systematics for the organization of turn taking for conversation. In *Studies in the organization of conversational interaction* (pp. 7-55). Academic Press.

Sato, R., & Takeuchi, Y. (2014). Coordinating turn-taking and talking in multi-party conversations by controlling robot's eye-gaze. In *The 23rd IEEE International Symposium on Robot and Human Interactive Communication* (pp. 280-285). IEEE.

Schegloff, E. A. (2000). Overlapping talk and the organization of turn-taking for conversation. *Language in society*, *29*(1), 1-63.

Sciutti, A., Schillingmann, L., Palinko, O., Nagai, Y., & Sandini, G. (2015, March). A gaze-contingent dictating robot to study turn-taking. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts* (pp. 137-138).

Striano, T., Henning, A., & Stahl, D. (2006). Sensitivity to interpersonal timing at 3 and 6 months of age. *Interaction studies*, *7*(2), 251-271

Tenenbaum, E. J., Sobel, D. M., Sheinkopf, S. J., Malle, B. F., & Morgan, J. L. (2015). Attention to the mouth and gaze following in infancy predict language development. *Journal of Child Language*, *42*(6), 1173-1190.

Tice, M., Bobb, S., & Clark, E. (2011, September). Timing in turn-taking: Children's responses to their parents' questions. In *Proceedings of the 15th Workshop on the Semantics and Pragmatics of Dialogue* (pp. 202-203).

Yoo, H., Bowman, D. A., & Oller, D. K. (2018). The origin of protoconversation: An examination of caregiver responses to cry and speech-like vocalizations. *Frontiers in psychology, 9*, 1510.