Right on time

Synchronization, overlap, and affiliation in conversation

Published by

LOT
Trans 10
3512 JK Utrecht
The Netherlands

phone: +31 30 253 6111
e-mail: lot@uu.nl

http://www.lotschool.nl

Cover illustration: Majellakerk, Utrecht, the Netherlands. Photograph by Anne van Leeuwen.

ISBN: 978-94-6093-262-5
NUR: 616

Right on time

Synchronization, overlap, and affiliation in conversation

Precies op tijd

Synchronisatie, overlap, en affiliatie in conversatie

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor
aan de Universiteit Utrecht
op gezag van de rector magnificus, prof. dr. G.J. van der Zwaan,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op
vrijdag 15 december 2017
des ochtends te 10.30 uur

door

Anne Ruth van Leeuwen

geboren op 27 mei 1984
te Nijmegen

Promotoren: Prof. dr. J.J.A. van Berkum
Prof. dr. H. Quené

The research project reported here was part of the project 'Moving the language user – Affect and perspective in discourse processing' supported by the Netherlands Organization for Scientific Research under project number 277–89–001.

To Gerke and Doede.

Contents

Acknowledgements

The journey that has led to this thesis would not have been possible without the support I gratefully received from many people around me.

It all started with Jos van Berkum. Jos, you believed in me as a scientific traveler. You were there the whole way, giving me directions and guiding me, even in the darkest moments when I was completely lost. One of the things you learned me was that slowing down can be very valuable, especially when you are in a hurry. Jos, thank you for being my guide, without you I would not have been able to climb this mountain.

Hugo Quené, you were the one who first introduced me to all those things I became to like so much: statistics, phonetics, and off course smiles:-). Your door was always open for me, and I felt that I could tell you literally anything (which I probably did). When I spoke to you, you always made me feel better, even when you criticized my work. Hugo, thank you for being there for me.

I would also like to thank my fellow vici travelers Björn 't Hart, Marijn Struiksma, Hannah de Mulder, Iris Mulders, and Suzanne Dikker. Björn, you showed me how things were done, which made me believe that I could do them too. Marijn, you guided me through the labyrinth of code and programming, and provided me with a shoulder to cry on. Hannah, you are so clever, and so sarcastic at the same time, and this makes you so much fun to be with! Iris, although you are still a mystery to me, your calm energy makes me feel brainless and very happy at the same time. Yes, I will let you know when I found the best fries around. Suzanne, although we did not see each other often, our research interests overlapped, and you proved to me that becoming a mother does not mean that your career is over. Dear vici members, thank you for your company and inspiration.

For the practical side of my work, I am grateful for the various students and student assistants who helped with running my experiments. Thank you Sanne Riemstra, Lea ter Meulen, Myrthe Bergstra, Jorik Geutjes, Ella Bosch, Marjan Bezemer and Paul Verhaar. I would like to thank Chris van Run, Martijn van der Klis, Maarten Duijndam, and Theo Veenker, for helping me with all the nitty–gritty technical stuff. Special thanks to Alexis Dimitriadis for being my LaTeX guru, to Maaike Schoorlemmer for helping me all those years, and to Simone Veld for coaching me along the way. I am very grateful to Arthur Dirksen, who built a custom version of Fluency (speech synthesizer) for me. Arthur, without you I could not even have begun this journey.

I would also like to thank my roommates Mirjam Hachem, Ileana Grama, Hanna de Vries, Xin Li, Bei Peng, Carolien van den Hazelkamp and Stavroula Alexandropoulou. Because you are the ones that had to deal with my daily panic attacks, my loud key strokes, and my endless chit–chatting every day. Thank you for putting up with me.

Carolien van den Hazelkamp, Marloes Herijgers, Jolien Scholten, thank you for your presence and attention during our weekly 'plan' meetings, which greatly helped me to make my deadlines in time. Without you I would still be staring at my computer screen.

Marjolein van Egmond, Liv Persson, Marijn Struiksma, and Carolien van den Hazelkamp, we shared a love for cycling. I tremendously enjoyed our Cyclinguist adventures, especially those in rainy Valkenburg. Thank you for climbing the Cauberg with me.

Anne-France Pinget, we met during our time at the research master Linguistics and soon became friends. I like your attentiveness, your lovely accent, and your sharp mind. I was very honoured to be your paranimph at your defense. Thank you for being my friend and role model.

In no particular order I would like to thank my fellow PhD travelers Hans Rutger Bosker, Marko Hladnik, Sophia Manika, Assaf Toledo, Brigitta Keij, Sandrien van Ommen, Gerdineke Silfhout, Eva Poortman, Louise Nell, Jet Hoek, Suzanne Kleijn, Anne Sara Romøren, Maartje Schulpen, Monica Koster, Renske Bouwer, Franca Wesseling, Anqi Yang, Anja Goldschmidt, Alexia Guerra, Heidi Klockman, Yipu Wei, Anna Volkova, Erin Pretorius, Jingwei Zhang, Dominique Blok, Luying Hou, Desiree Capel, Liquan Liu, Tom Lentz, Hayo Terband, Myrthe Bergstra, Andrea Santana Covarrubias, Rogier Kraf, Shuangshuang Hu, Mengru Han, Jorik van England, and Martijn van der Klis. I am grateful for having lunch with you, for sharing our PhD struggles, and for just having you around. Thank you.

Carolien van den Hazelkamp and Wouter van Leeuwen, my paranimphs, require special attention too. Carolien, we shared an office for almost 3 years and in those years we shared all our struggles, joys, and sorrows. I fills me with joy that we will defend our work almost at the same time. Wouter, my brother, we share a lifetime together. I am very proud to be your little sister. I witnessed your defense as your paranimph, and I cannot wait to see those roles reversed.

I am very grateful for my dear friends Suzan van der Burg, Merel Kahmann, Evelien Davidson, and Maartje Prinsen. Suzan, what would I be without you? You corrected my manuscript, organized our wedding, and cared for our newborn family. Your calm force makes all come into place. Merel, we grew up together and we have collected shared memories for over 20 years. I hope that we keep doing that for a very long time. Evelien, our paths have crossed many times over the last 15 years. The last two years our lives have ran almost parallel, which only strengthened our friendship. I look forward to our shared travels together, big and small. Maartje, while I was busy with my journey, you faced many challenges on your own travels. I admire your courage, strength, and positive vibe in those hard times. Thank all of you, for listening to me, for drying my tears, for just being there for me. Thank all of you for being my friend.

My dear climbing buddies, Joosje Machielse, Thamar van Esch, Julie Tabbers, Dorien Verhulst, Neeltje de Gruyter, Michelle van der Kind, Roanne van Voorst, Anne Raaymakers, Suzanne van der Sluis, Marlies Quack, Martin Fickweiler, Casper ten Sijthoff, Rogier Wouters, Bart Korff, Daniella de Wit, Hanneke Blüm, Celine de Waal Malefijt, Rogier Koppejan, 'Jibbi' and many more. Thank you for holding my rope. You kept me sane.

I would also like to thank the Hoogtelijn editors Femke Welvaart, Peter Daalder, Marieke van Kessel, Mirte van Dijk, Ico Kloppenburg, Ivar Schulte, Florian van Olden, Frank Husslage, and Ernst Arbouw. Thank you for having me on board, and keeping me there, despite the fact that I couldn't contribute as much as I wanted last year. Working at Hoogtelijn magazine reminds me how much fun writing is, and how much I love (thinking about) the mountains. Thank you for keeping that passion alive.

Lastly, I would like to thank my family. Ben and Els, Femke, Edwin, Jasper, Linde, and Hidde, I am so happy to be part of your wonderful family. Thank you for providing a second home for me.

Antoinette and Pieter, my dear parents, you cared for me when I was a baby, raised me when I was a child, and guided me in adulthood. When I fell with climbing and severely injured my back, you took me back home and cared for me again. I am eternally grateful that I am your child, and that you travel with me wherever I go.

Gerke and Doede, only thinking of you two makes me cry of happiness. Gerke, love of my life, you believe in me when I want to give up, you care for me when I am hurt, and you always make me feel loved. The years spent working on this book have been the happiest of my life, and that has everything to do with you. We married in 2015, and in that same, wonderful year, our son Doede was born. Doede, in your eyes I can see the reflection of the love between me and your father. You belong with us and I am looking forward to watch you grow. I love you both so much. Thank you for being in my life.

CHAPTER 1

Introduction

1.1 Being 'in sync'

When a conversation is running smoothly, you know exactly when to nod, hum, or when to start your turn. You feel understood and connected, and you sense that your conversational partner feels the same. However, a conversation may also contain awkward silences, simultaneous starts, and an overall feeling of stuttering and stammering. During such conversations, you are often left with feelings of distance and mutual incomprehension.

Many people share the intuition that the expression of 'being in sync' with someone means that you are somehow in tune, in agreement, or in harmony with the other. This dissertation explores whether this intuition is correct; it investigates whether specific temporal patterns between turn-taking speakers, including synchronization of speech rhythms, shape the affective impression of speakers in conversation. The answer to this question can broaden our understanding of the affective push-and-pull of spoken interaction that we experience every day.

The temporal pattern between two turn-taking speakers can be expressed in a number of ways. For example, it can be expressed in terms of the time distance or interval between turn offset and onset, such as the amount of silence or of overlapping talk between speakers. However, it can also be expressed in terms of speech rhythm or tempo. For example, a speaker in conversation may adjust his speech tempo to match the tempo of the previous turn, or he may place his first 'beat', or first stressed syllable, in such a way that it coincides with the (extrapolated) rhythm of the previous turn.

Here, we specifically explore the affective impact of synchronization and tempo matching between turn-taking speakers, on top of the established effects of silent gap and overlap between them.

In the introduction that follows, we will first discuss how interpersonal temporal patterns in *non-verbal* interaction may convey affective information about the people in that interaction (section 1.2). We will then turn to *spoken* interaction, where we will elaborate on how silent gaps and overlapping talk may give rise to affective impression about those speakers (section 1.3). We will then turn to why certain other temporal patterns between speakers, such as responding 'on the beat', may also be an important factor when investigating the affective impact of temporal relations between speakers in conversation. Finally, we will discuss affect in more detail, and elaborate on how different aspects of a conversation may give rise to emotions (section 1.4).

1.2 Temporal coordination in non-verbal interaction

Playing in an orchestra, dancing the tango, or lifting an object together: humans have a remarkable ability to tightly coordinate their actions in time with no or little apparent conscious effort. These specific examples of coordinated interactions are generally seen as 'joint actions'. A joint action, according to Sebanz, Bekkering, and Knoblich (2006), is "any form of social interaction whereby two or more individuals coordinate their actions in space and time to bring about a change in the environment" (p.70). By definition, accurate temporal coordination plays a crucial role in non-verbal joint action.

Partners in interaction may, for example, *match* their facial expressions or bodily postures. Although these behaviors in itself do not necessarily require precise temporal control within an individual, they often occur shortly after each other in interaction, usually within a second (Dimberg & Thunberg, 1998). There is considerable empirical evidence that humans are very good at matching, intentionally or not, their non-verbal behaviors during joint actions. Humans can imitate each other's bodily postures (Bavelas, Black, Chovil, Lemery, & Mullett, 1986, 1988; Bernieri & Rosenthal, 1991; Shockley, Santana, & Fowler, 2003), they laugh or yawn together (Hatfield, Cacioppo, & Rapson, 1994), and humans tend to automatically imitate each other's facial and vocal expressions during interaction (Hess & Philippot, 2007).

Similarly, rhythmic behaviors too tend to match or *'synchronize'* during interaction. Humans spontaneously synchronize when rocking in a rocking chair together (Demos, Chaffin, Begosh, Daniels, & Marsh, 2012), they synchronize their postural rhythms when talking together (Shockley, Baker, Richardson, & Fowler, 2007), they synchronize their joint tapping when tapping together (Konvalinka, Vuust, Roepstorff, & Frith, 2010; for a detailed review about joint tapping see Repp & Su, 2013), they synchronize their keystrokes in computer-mediated interaction (Campbell, Cothren, & Burg, 2010), and they synchronize their handheld pendulum swinging when walking together (Richardson, Marsh, & Schmidt, 2005).

1.2.1 Non-verbal interpersonal coordination conveys mutual affiliation

Crucially, behavioral matching or synchronization during non-verbal joint actions has been found to reduce social differences and to increase mutual affiliation between individuals in interaction.

Behavioral matching has been long linked to affiliation (Bavelas et al., 1986, 1988). Non-conscious matching of gestures, postures, and mannerisms have been found to enhance the smoothness of interactions and to foster liking (Chartrand & Bargh, 1999).

Moreover, behavioral matching not only *creates*, but also *reflects* affiliation. People intending to affiliate with others are especially prone to match their behaviors (Lakin & Chartrand, 2003), and behavioral matching tends to increase when individuals feel more connected, or when they feel more empathy toward the other person in interaction (Fischer, Rotteveel, Evers, & Manstead, 2004).

Moving in a synchronized fashion leads to similar positive attributions of mutual affiliation. Hove and Risen (2009) had participants tap to a visual metronome together with an experimenter that tapped synchronous to the participant's metronome, asynchronous to the participant's metronome, or did not tap at all. Affiliation ratings were higher if the participant and experimenter tapped in synchrony. Valdesolo and DeSteno (2011) used a similar manipulation of joint tapping and found that successful synchronization experience increased feelings of similarity and compassion, and also increased altruistic behavior. These positive attributions of interpersonal synchronization are already observed very early in life (Cirelli, 2014).

Additionally, not only participating, but merely *observing* movement synchronization has been found to lead to similar positive interpersonal attributions as well. Lakens (2010) and Lakens and Stel (2011) investigated whether movement synchronization signaled affiliation. Both for stick figures and for humans waving, moving in synchrony resulted in greater affiliation ratings, even for different movements performed in the same rhythm. Increasing the difference between movement rhythms lowered the affiliation ratings, suggesting a positive linear relationship between the amount of synchronization and evaluated affiliation of interactants. Similar findings of affiliation were obtained for walking pairs walking in different modes of synchrony (Miles, Nind, & Macrae, 2009).

Taken together, interpersonal coordination, including synchronization and behavioral matching, really appears to somehow serve as a 'social glue' (Chartrand & Bargh, 1999), reflecting and creating affiliation among interactants (see Vicaria & Dickens, 2016, for a detailed review on the interpersonal outcomes of behavioral matching and synchronization during non-verbal interaction).

1.3 Temporal coordination in spoken interaction

Spoken conversations may also be regarded as a specific form of joint action between interlocutors (Clark, 1996; Garrod & Pickering, 2009). Interlocutors converse because they ultimately assume that it is for their individual or mutual benefit to do so (Tomasello, 2008). This is what motivates them to cooperate in getting the message across in the first place ('shared intentionality', Tomasello, 2008, p. 88).

Like non-verbal interaction, spoken conversation requires fine temporal coordination between interlocutors as well. This is clearly visible when looking at the temporal patterns between turn-taking speakers. As already pointed out, the temporal relation between two turn-taking speakers can be expressed in terms of the time distance, or interval, between turn offset and onset between speakers, that is, as the amount of silence (gap) or overlap between speakers (see Figure 1.1).

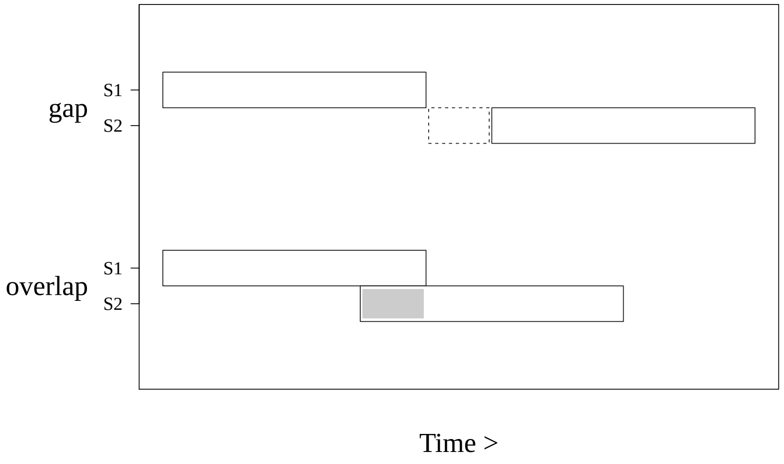

Figure 1.1. Schematic overview of gaps and overlaps between two speakers (S1 and S2). The rectangles represent the individual speaking turn of each speaker. The dashed rectangle represents the silent gap between the two speakers, the grey rectangle the amount of overlap between them.

1.3.1 Gaps and overlaps

The parameters of the amount of silent gap and overlap in conversation appear to be rather universal. In a cross-linguistic corpus investigation on the timing of yes/no questions, Stivers et al. (2009) found broadly similar patterns of turn timing between speakers across the 10 languages under investigation. Although the 10 languages differed in the average duration of the silent gap between speakers, all languages generally appeared to minimize overlapping talk and (long) gaps between speaking turns (Stivers et al., 2009). These findings are in line with Sacks' (1974) influential claim that the timing of turn-taking is so precise that speakers aim at no gap and no overlap between turns. To achieve such smooth turn-taking, interlocutors rely on lexical-syntactic as well as on prosodic information to predict turn ends and to precisely time their subsequent speaking turn (Bögels & Torreira, 2015).

Although short gaps are indeed most common in conversations, longer gaps and overlapping talk does occur as well (Heldner & Edlund, 2010; Ten Bosch, Oostdijk, & de Ruiter, 2004). The most common interval between speaking turns, at least for Dutch, Swedish, and Scottish, is a short gap of about 200 ms, but gaps longer than 500 ms (around 25% of the turn exchanges), and overlaps (around 40% of the turn exchanges) are common too (Heldner & Edlund, 2010). Taken together, short gaps are most common, but longer gaps and overlaps also characterize our spoken interactions.

Gaps

Interestingly, deviations from the conversational 'norms' of 'no-gap-no-overlap' or 'short gap' have communicative significance. Besides reflecting cognitive speaker variables, such as problems with comprehension (Beňuš, Gravano, & Hirschberg, 2011), planning difficulties (Bull & Aylett, 1998), native and non-native fluency (Bosker, 2014), or word frequency (Hartsuiker & Notebaert, 2010), longer gaps occur most often before disconfirmative responses and before non-answers (Stivers et al., 2009). Such responses somehow fail to conform to the terms of the question (e.g., I don't know or I can't remember) or to the questioner's agenda (e.g., A: Is that your car? B: No) (Stivers et al., 2009). These disconfirmative responses or non-anwers were generally delivered between 100 and 500 ms slower than their opposites in all ten languages under investigation (Stivers et al., 2009).

Interestingly, gaps also have communicative significance for outside *observers* of a spoken interaction. Roberts, Margutti, and Takano (2011) presented listeners from three distinct languages with scripted telephone conversations that differed in the silent gap between speakers. Listeners were asked to rate the addressee's willingness to comply with a request or addressee's agreement with an assessment. Regardless of language background, all raters judged speakers to be less willing to comply with requests or agree with assessments, as gaps before agreeing became longer. This suggests that there is a decreased sense of between-speaker agreement, compliance or connectedness with increased gap duration. These interpersonal attributions may be grouped together as reflecting *affiliation* between speakers; longer gaps are taken as less affiliative, or more distancing, than shorter gaps (an elaborated discussion around the concept of affiliation can be found in section 1.4). Crucially, the findings discussed above suggest that longer gaps may somehow convey a lack of affiliation, both for interlocutors as well as for mere observers of a conversation.

Overlaps

In addition to the duration of the silent gap between interlocutors, the presence of overlapping talk between speakers in conversation can also convey disaffiliation between them. Although there may be cultural differences, overlapping talk in English and Dutch has mostly been associated with disaffiliation and strong personality attributions: speakers who overlap are perceived as less agreeable and more assertive (Maat, Truong, & Heylen, 2010; Robinson & Reis, 1989), as more dominant (Beňuš et al., 2011), and less sociable (Robinson & Reis, 1989) than non-overlapping speakers. Goldberg (1990) showed that overlaps are indeed most often linked to displays of power and control, and generally viewed as rude and disrespectful.

Goldberg (1990) also showed that overlaps, even though generally perceived as competitive, can convey cooperation and affiliation between speakers as well. In the previous paragraph, the overlaps can be seen as conveying that the overlapper wants to take over the floor. An overlap can also be non-competitive, as in the case of a back-channel for example. A backchannel is a short verbalization like ‘huh’, ‘mmm’ that occurs while the other speaker is still speaking, which indicate listeners’ attention like ‘I am still here / I hear you and please continue’ (Kurtić, Brown, & Wells, 2013). Under these specific conditions, among others, overlaps can convey rapport or cooperation as well (Goldberg, 1990).

Taken together, the presence of an overlap in conversation can sometimes convey mutual affiliation between speakers. Predominantly, however, overlaps convey the opposite; overlapping speakers are perceived as more rude and less sociable than speakers who do not overlap, with the exception of back-channels. Generally, the presence of an overlap in conversation can thus be taken to convey disaffiliation between speakers.

1.3.2 Synchronization and tempo matching

The preceding sections showed that silent gap or overlap between speakers in conversation can convey information about those speakers. However, as already pointed out, the temporal relation between turn-taking speakers may also be described in terms of speech rhythm or tempo.

Synchronization occurs when a speaker in conversation adjusts his speech rhythm to the (extrapolated) speech rhythm of the previous speaker, so that their speech rhythms coincide or become coupled. Figure 1.2 displays a schematic example of a synchronized exchange between speakers (‘sync’) and of a non-synchronized exchange between speakers (‘no sync’). For the synchronized exchange, S1 has a certain rhythm, which creates a regular beat or pulse (‘extrapolated rhythm S1’). The second speakers (S2), aligns his (or her) rhythm in such a way that it coincides with the (extrapolated) rhythm of S1, as indicated with the dashed vertical line. In other words, S2 responds ‘on the beat’ of the rhythm of S1; the rhythms of S1 and S2 are synchronized (‘sync’). On the other hand, when S2 does *not* align his rhythm to the beat of S1, and instead utters his first beat ‘too early’, then the rhythms of S1 and S2 are not synchronized (‘no sync’).

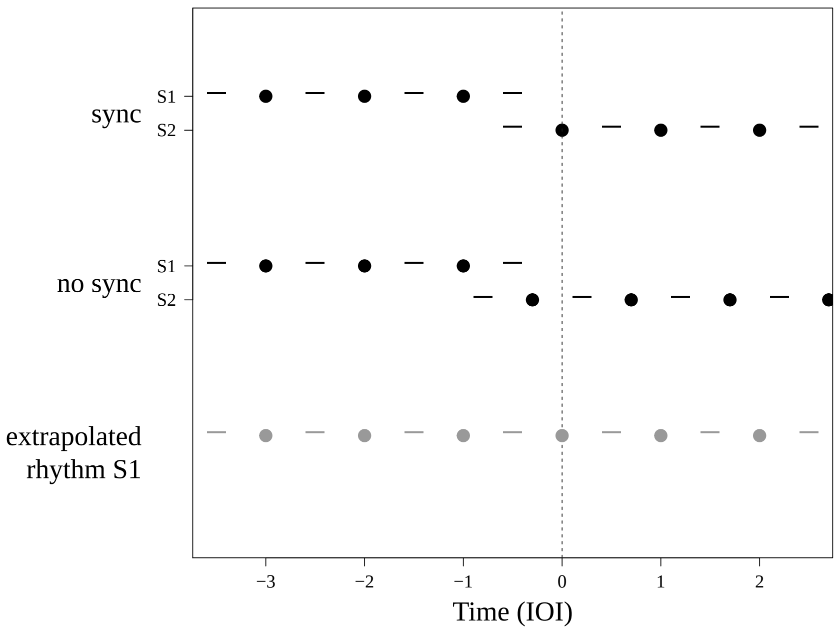

Figure 1.2. Schematic overview of synchronization between speakers. Bullets and dashes represent the strong-weak rhythm of both speakers (S1 and S2). The upper half depicts a synchronized exchange ('sync') between speaker S1 and S2, and the lower half depicts a non-synchronized exchange ('no sync') between speakers S1 and S2. The horizontal gray line at the bottom represents the relevant (extrapolated) rhythm of S1, the grey vertical line represents the rhythmic alignment between speaker S1 and S2, with speaker S2 coming in 'on the beat' (sync) or 'too early' (no sync) given the rhythm of S1. Time is expressed as the inter-onset-interval (IOI) between strong 'beats'.

Tempo matching occurs when a speaker adjusts his speech tempo to match the tempo of the previous turn. Figure 1.3 displays a schematic example of a tempo matched exchange between speakers ('match') and of a tempo mismatched exchange ('mismatch') between speakers. As can be seen in the upper part of Figure 1.3, the first speaker (S1) speaks with a certain tempo. When the second speaker (S2) responds with the same tempo the speech tempos of S1 and S2 are matched. On the other hand, when the second speaker (S2) responds with a *faster* tempo than the first speaker the tempos of S1 and S2B are mismatched.

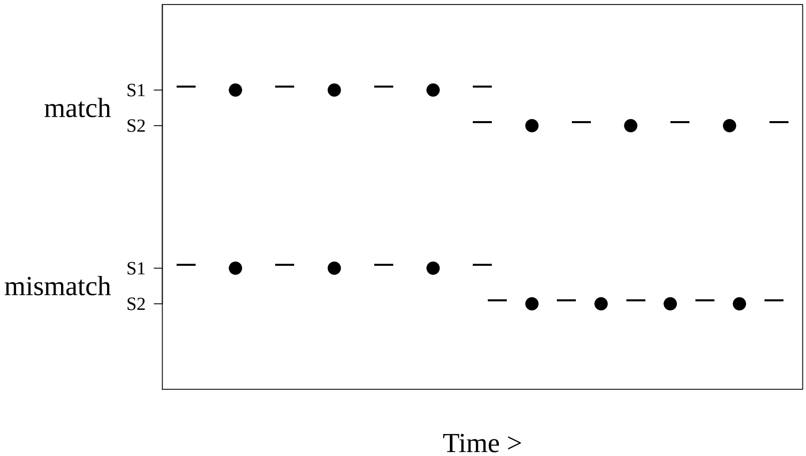

Figure 1.3. Schematic overview of tempo matching and tempo mismatching between speakers. The bullets and dashes represent the speech tempo of the two speakers (S1 and S2). The upper part depicts a tempo-matched exchange ('match') between S1 and S2, and the lower part a tempo-mismatched exchange ('mismatch') between S1 and S2, with speaker S2 speaking too fast given the tempo of S1.

In the remainder of this section we will discuss synchronization and tempo matching in conversation in more detail, and focus on how these specific temporal patterns between turn-taking speakers in spoken interaction may come about.

Synchronization

Rhythm in speech The concept of synchronization is closely connected to the notion of 'rhythm'. We define rhythm as the succession of stronger and weaker elements, periodically recurring in continuous time. Rhythm in this sense creates an expectancy profile of when the next (strong) element or beat is going to be (Barnes & Jones, 2000). In this way, rhythm can be used to generate, satisfy, or frustrate expectations in time (Cummins, 2015).

In speech, there is no regular pulse or beat (Bertinetto, 1989; Dauer, 1983) as is often found in music. Speech is predominantly dictated by the words, and their sequential order in the spoken language. Words differ in number of syllables, stress placement and duration, and syllables differ in structural complexity and number of phonemes. Prosody, pragmatic function, emotion, and articulatory constraints add further complexity (Hawkins, 2014).

Although speech lacks strict periodicity, it is generally perceived as being rhythmical in nature (Lehiste, 1977). When imitating the timing of spoken sentences, listeners tap with less temporal variability than the actual variability in timing of the sentences, whereas imitating noise does not show this pattern (Donovan & Darwin, 1979; Lidji, Palmer, Peretz, & Morningstar, 2011, for speech only).

Historically, the phenomenon of perceived regularity in speech gave rise to the idea that some events, such as the onsets of syllables, might be found to be regularly spaced in time. It became quickly apparent, however, that sequences of words that were aligned perfectly with respect to their acoustic onsets were not perceived as rhythmical (Morton & Marcus, 1976). This phenomenon gave rise to the question at which instant, relative to its physical onset, a certain speech event is perceived. The 'perceptual center' or 'P-center' was introduced to describe this moment in time. A P-center denotes the instant at which a syllable or a vowel of speech is perceived to happen (Chow, Belyk, Tran, & Brown, 2015; De Jong, 1994; Morton & Marcus, 1976; Scott, 1993). The P-center is generally taken as the perceptual estimate of the beat location in speech. Although the P-center does not correspond to any simple acoustic or articulatory feature, the increasing amplitude at the onset of stressed syllables critically influences the perception of the P-center, and therefore, the P-center is generally associated with the onset of stressed vowels.

By actively adapting to external rhythms, such as to P-centers in speech or to beats in music, humans may create a dynamical expectancy profile, or attentional pulse, about when the next event is going to occur (Large & Jones, 1999). Port (2003) proposes that neuro-cognitive oscillators produce this attentional pulse. These neuro-cognitive oscillators apparently influence the motor system by biasing production so that perceptually salient events line up in time close to these neuro-cognitive pulses. Port (2003) suggests that the attentional pulse selects particular instances in real time, and prefers that important events, like P-centers, happen at those points. This kind of role for rhythmic structure in speech has been suggested to facilitate the parsing of the speech stream (Cutler & Mehler, 1993), the acquisition of a first language (Morgan, 1996), spoken-word perception (Quené & Port, 2005), and speech perception in noise (Aubanel, Davis, & Kim, 2016). In everyday interaction, however, speakers seem to be able to control the degree to which they allow these rhythmic attractors to constrain their speech timing (Port, 2003) or, alternatively, these attractors may be overridden by other, more pressing factors, such as intrinsic syllable properties, prosody, and articulatory constraints (Hawkins, 2014). Therefore, it may well be that strict rhythmicity in speech takes second place when it comes to speaking (Hawkins, 2014).

Even though speech is considered to lack strict periodicity, individuals may exaggerate their speech to make it sound more rhythmical. Good examples are, among others, rappers, orators, journalists, poems, auctioneers (Kuiper, 1992) and nursery rhymes (Bergeson & Trehub, 2002). Even though spontaneous speech lacks strict rhythmicity, it can be produced rhythmically under certain circumstances.

The next section focuses on how and under which circumstances this might happen and discusses how speaker and listener may become synchronized to each other on the basis of their speech rhythms.

Speech rhythm synchronization When speaking to a metronome, speakers tend to place their onsets of stressed syllables at temporal harmonics fractions of the metronome cycle (Cummins & Port, 1998; Port, 2003). These observations come from the Speech Cycling Task (Cummins & Port, 1998; Port, 2003). In this task, participants repeat phrases over and over, with their timing stabilized by a simple metronome. Measurements of the location of one or more prominent syllable onsets, or P-centers, can be made and interpreted as phase angles relative to the repetition of the metronome cycle (Cummins & Port, 1998; Port, 2003). In a series of experiments the authors showed that speech which is spoken to a metronome tends to be placed at preferred phase angles of 1:1, 2:1 or 3:1. The authors called this the Harmonic Timing Effect.

Some metrical patterns are easier than others, with the most preferred phase angle of 1:2. This suggests that the strong-weak or weak-strong pattern can be cycled easily, while others (e.g. those containing a Sww foot) are less stable. These observations closely resemble the findings found with the standard tapping paradigm, in which inter-tap-intervals are often simple ratios of the inter-onset-intervals of the induction sequence (Repp & Su, 2013).

Additionally, when speakers are asked to speak in synchrony with one another, they are able to do so, with only very short inter-speaker lags of 40 ms on average (Cummins, 2009b). In the synchronous speech task, two subjects read a prepared text together, attempting to remain in synchrony. The ability of speakers to maintain such tight temporal alignment in the absence of any underlying beat suggests that rhythm might function as an affordance for successful temporal coordination between interlocutors (Cummins, 2009a).

To govern smooth turn-taking in spontaneous conversation, Wilson and Wilson (2005) propose a theoretical 'coupled oscillator' model in which internal oscillators of the brain of speaker and listeners become coupled or synchronized on the basis of the speaker's rate of syllable production. Their idea is that this rhythmic coupling governs when speaker (or listener, as potential next speaker) initiates speech at any given moment.

In line with the coupled oscillator model of Wilson and Wilson (2005), turn-taking speakers do indeed appear to locally synchronize their speech rhythms, even though this synchronization appears to be based on the temporal interval between adjacent prominent syllables (Beňuš, 2009), and not on the rate of syllables, as initially proposed by Wilson and Wilson (2005).

When in conversation together, speakers tend to *locally* display synchronization between turn pairs (Ogden & Hawkins, 2015). The authors analyzed question-answer pairs during spontaneous conversation and found that the initial P-center ('pike'), or the initial sound that the answerer made, such as a click, in-breath, um, etc., was usually synchronized with the last two or three beats of the question. In particular, the last two to three beats of a question tend to have fairly equal inter-onset-intervals, signaling (with other factors) that the end of the turn is imminent. These final beats set up an underlying pulse which seems to provide the next speaker with a structure to coordinate the beginning of the answer to. Moreover, speakers who overlap their turn with that of the preceding speaker tend to locally display synchronization between turns as well (Wlodarczak, Šimko, & Wagner, 2012). Overlapping speakers tend to align their overlap in such a way that the onset of the overlap more frequently coincides with a P-center of the preceding turn than with any other location within the syllable. The overlap onset alignment is, at least to some extent, also dependent on the regularity of the preceding syllables. Together, these results provide evidence for *local* synchronization between speakers in spontaneous conversation. Additionally, these findings suggest that two to three beats across two turns may be sufficient to achieve and confirm successful synchronization in conversation.

Synchronization and interpersonal affect There has not yet been any research on the affective impact of speech rhythm synchronization as defined in this dissertation. However, McGarva and Warner (2003) tested whether on-off vocal activity 'entrainment', or coupling between speakers, can be explained by the factors liking and similarity. On-off vocal activity is basically a measure of speaking activity, that is, it is the speech-pause rhythm of a speaker. The authors instructed participants to engage in a 40-minute introductory conversation. Participants were paired according to attitude and similarity questionnaire outcomes. Before and after the conversation, participants were asked how much they liked their interacting partner and how much they liked working with them. On-off vocal activity cycles for each individual were identified. Mutual entrainment of vocal activity was present when the variability in the vocal activity of the two interlocutors could be explained by the same periodic component and by a phase relationship of anti-phase. Entrainment was found for one third of the participant pairs, another third was slightly entrained and the rest was considered not entrained. Nonetheless, on-off vocal activity synchronization did not appear to be linearly related to judgments of attraction or conversation quality.

Although the approach was promising, the authors did not find the hypothesized effects. As we have seen, speech rhythm synchronization is most pronounced near the edges of turn exchanges (Ogden & Hawkins, 2015). Based on the reviewed literature, the degree of *local* synchronization of vowel onsets of adjacent prominent syllables between speakers appears to be a more promising measure when investigating the social impact of synchronization in conversation.

Tempo matching

The second type of temporal relation between turn-taking speakers explored in this dissertation is tempo matching, which relies on the notion of 'tempo'. Tempo in speech may be defined as the rate at which phonetic events occur over time (Quené, 2008). Generally, the syllable is taken as the central unit; speech tempo is often expressed as syllables per second (Stetson, 1951), or as average syllable duration (Crystal & House, 1990; Goldman-Eisler, 1968). Speech tempo may vary as a function of both between-speaker variables, such as language background (Verhoeven, Pauw, & Kloots, 2004), age (Ramig, 1983; Ryan, 1972) and sex (Verhoeven et al., 2004), and as a function of within-speaker variables, such as phrase length (Quené, 2008), and discourse (van Beinum & van Donzel, 1996).

Importantly, when speaking together, speakers tend to match or accommodate to each others speech rate over time (Giles, Coupland, & Coupland, 1991; Manson, Bryant, Gervais, & Kline, 2013; Schultz et al., 2016; Street, 1984), that is, the tempos of both speakers tend to become more similar during the course of the interaction.

Tempo matching may convey interpersonal affect Like behavioral matching during non-verbal joint action, tempo matching too has been claimed to convey positive speaker evaluations.

Street (1984), and Buller and Aune (1992) assessed the affective impact of tempo matching by presenting listeners with different spoken passages of different speaking rates, after which they rated the speaker on several personality attributions. Before or after this test, the actual speech rate of the participant was measured together with the perceived speech rate of the speaker. These measures were then taken as a predictor for the personality ratings. Listeners who perceived the speaker's rate as faster than their own rated him as more competent and more attractive, whereas the opposite was true for listeners who perceived the speaker's rate as slower than their own (Buller & Aune, 1992; Street, 1984). Given that speaker and listener were never in conversation together and the fact that speech rate similarity was not assessed between interacting speakers, these findings do not serve as convincing evidence that speech rate similarity between speaker and listener is related to mutual affiliation. Nevertheless, these findings do suggest that, at some level, speech rate similarity between interlocutors may indeed reflect affiliation.

A more recent study of Boltz and Miller (2010) showed that actual speech rate similarity between speaker and listener did indeed influence the impression of the interlocutor under investigation, but this time it only concerned the impression of 'truthfulness', which is not easily transportable to our notion of 'affiliation' (which will be discussed in the next section about affect, see section 1.4).

Only Manson et al. (2013) investigated the affective impact of tempo matching during real-life interaction. Contrary to our expectations, this study did not show effects of speech rate similarity on perceived affiliation. Manson et al. (2013) had sets of three participants converse with each other for 10 minutes on any topic(s) they wished after which they played a one-shot prisoner's dilemma and then rated each of their co-participants on "warmth" and "competence" (taken together as affiliation). A sample of the each conversation was then analyzed, besides other things, in terms of speech rate convergence. When speakers' speech rate converged more strongly from the beginning to the end of the conversation, they were more likely to cooperate in the prisoner's dilemma, but no effects on perceived affiliation were found. Matching on other speech dimensions, such as linguistic style matching (see Niederhoffer & Pennebaker, 2002), did increase perceived affiliation. This suggests that even though speech rate similarity may reflect an increased tendency to cooperate, it may not necessarily also lead to mutual impressions of affiliation.

In all, this shows that the affective impact of tempo matching in conversation is not clear. Even though some studies do suggest that speech tempo matching may indeed be of affective relevance, other studies suggest that this is the case for other types of verbal matching, but not for speech tempo matching in particular.

Communication accomodation theory

So far, the idea that synchronization and tempo matching between speakers might convey interpersonal affect has been derived from findings in the non-verbal domain (e.g., Vicaria & Dickens, 2016). However, the idea is also in line with the communication accommodation theory (CAT), developed by Giles and colleagues, (e.g., Giles 2016). During conversation, speakers may adjust their speech, vocal patterns and/or their gestures to each other during the course of the interaction. CAT poses that these adjustments may occur on *all* levels of communication, and that they are often undetectable, to both interlocutors as well as (trained) overhearers. This suggests that these adjustments between speakers are deeply fundamental and that they may be at the base of what makes social interaction possible at all (Levinson & Enfield, 2006).

Crucially, CAT posits that these between-speaker adjustments during conversation serve two major functions; they help facilitate coherent interaction and, importantly, they allow interactants to manage social distance between one another (Chartrand & van Baaren, 2009). The more speakers wish to affiliate, e.g., to decrease social distance, the more they accommodate. Hence, CAT proposes that verbal accommodation during conversation conveys speakers' need for affiliation.

There is abundant empirical evidence in favor of CAT, e.g., the amount of matching of various kinds of vocal behaviours, such as F0 or vowel spectra, have been shown to be modified by social factors such as liking (Babel, 2010), attraction (Babel, 2012) and strength of relationship (Pardo, Gibbons, Suppes, & Krauss, 2012).

One crucial concept in the paragraphs above has to be discussed before we can turn to the general setup of this dissertation. It is the notion of 'affiliation' or 'affect', which is crucial to this work. The next section will discuss these concepts in more detail.

1.4 Affiliation and affect in conversation

As humans, we are social beings. We like to be together, to collaborate, and to share experiences. However, sometimes we must do the opposite; avoid others, distance ourselves from them, or even compete with them, because we think or feel that they pose a serious treat to us.

The concept of ‘affiliation’ captures these social tendencies. The noun ‘affiliation’ may have different interpretations, but they share one commonality; they all have to do with some sort of social ‘connection’, or ‘closeness’. This can either be directional, like ‘the process of getting closer’, but also non-directional, expressing a state, like ‘being close’. In this dissertation, we define affiliation as enacted or experienced social closeness.

1.4.1 Affiliation is rooted in our affective system

The concept of affiliation is closely linked to the approach – avoidance dimension, which is central to general theories of emotion (Frijda, 2009; Panksepp & Biven, 2012). Without going into too much detail, emotions help us decide whether to affiliate with, or to distance oneself from, someone or something (Fischer & Manstead, 2016; Panksepp & Biven, 2012).

Following Van Berkum (in press) and adapted from Scherer (2005), an emotion is described as a package of short-lived concurrent motivational, behavioral, physiological, and cognitive changes, triggered by the appraisal of a stimulus event relevant to the interests of the organism, and aimed at generating a prioritized functional response to that stimulus event. An emotion does not necessarily have to be consciously experienced to be effective. That is, a weak emotion may unfold and influence our thoughts and behavior without any subjective awareness (see Van Berkum, in press, for discussion). Emotions provide the essential infrastructure for our most basic instinctual behavioral patterns, approach and avoidance, which provide guidance for living (Panksepp & Biven, 2012). Humans, and other animals, approach things that evoke pleasant emotions, such as encounters with food, water, a mate or a friend. This helps them to survive and reproduce. On the other hand, they stay away from things that make them feel unpleasant, like predators, chaotic weather, and rivals and enemies, because such encounters put life and reproductive capacity in jeopardy (Panksepp & Biven, 2012). Such approach – avoidance tendencies, elicited by our emotions, help us to decide whether to affiliate with something or someone or to distance oneself from something or someone. In other words, much of human emotion is inherently socially relevant .

The *affiliative* function of emotions is helping the individual to establish or maintain cooperative and harmonious relations with other individuals (Fischer & Manstead, 2016). Emotions may promote interpersonal closeness and harmony, thereby avoiding social isolation, which enables us to form and maintain intimate relationships. On the other hand, emotions may have a *distancing* function as well (Fischer & Manstead, 2016). They help the individual to be safe, to differentiate from others or to compete with them for social status or power. Examples of emotions that serve a social distancing function are: anger, contempt, disgust, or fear of another person. Thus, affiliation is rooted in our affective system.

Affiliation in conversation

In non-verbal interaction, the cause for (dis)affiliation, that is, the tendency to approach or avoid someone, has to do with what the other person is, or with what the other person does. His (or her) behavior, appearance, or smell, just to name a few, may all elicit an affiliating or distancing response. In *verbal* interaction, this is basically the same; it is the other person, or his behavior, that may elicit a social distancing or affiliating response. Contrary to non-verbal interaction however, verbal behavior also involves *language*. This adds additional complexity when it comes to identifying the trigger of the affective response of speakers in conversation.

For language, there are many different communicative causes that may elicit an affective response (Van Berkum, in press). The meaning of the *individual words* can by itself activate the affective system of the listener. Good examples are taboo words or swearwords. The *referential situation* refers to the concrete mental model of the entire situation that the speaker is referring to; this is often called a 'situation model' (Zwaan, 1999). For example, when the referential situation is "She drove the knife into her husband's belly", this will likely elicit some sort of distancing response. The *affective stance* of the speaker refers to his or her orientation to a particular state of affairs or 'stance object' (Kiesling, 2011; Stivers, 2008). For example, speaker stance can be, deliberately or not, signaled by silence, angry prosody or facial expressions. These signals of speaker stance may be an affective trigger for the listener too. The *social intention* of the speaker refers to the intention of the speaker; what is it that he or she wants to achieve by communicating with the listener? For example, when the social intention of the speaker is to insult the other, this will almost certainly function as a powerfull emotional trigger. The *communicative intention* refers to the intention of the speaker to communicate with the listener. This desire to communicate might by itself might be a potential affective trigger for the listener too; he or she may not be in the mood for interaction with the other person at this very moment. Lastly, Van Berkum (in press) proposes that there is always a *bonus meaning*, which includes those things that the speaker did not mean to convey at all. The bonus meaning is an additional evaluation of the listener about the speaker, or about the world in general, such as “she obviously had a bad day”, or “why do people always need to put other people down?”. Although not part of speaker meaning proper, such bonus meaning will usually strongly contribute to whatever the listener will feel.

Although these various possible causes for (dis)affiliation will not be further explored in this dissertation, it is important to have an idea of what may cause speakers to (dis)affiliate as they respond to a previous speaker’s turn. Becoming aware of this additional complexity is of relevance when discussing the design of our rating task in the next section.

Moreover, knowledge about the various possible affective triggers in spoken language increases our understanding of the affective factors at play during spoken conversation, an area that has largely been ignored in psycholinguistics. As humans, however, we are highly social beings, and we use (spoken) language to influence each other, and to share what we care about. As Van Berkum (in press) puts it: “[we] need to understand why language *works*, i.e., how it does not only inform but also *affect* other people. In that puzzle, emotion is not peripheral, but the key”.

In this dissertation, we aim at taking a first step towards a better understanding of some of the affective forces at play during the dynamics of spoken interaction; it focuses on the *when* of dialogue and seeks to investigate whether and how specific temporal relations between turn-taking speakers can influence the affective impression of speakers in conversation.

1.5 Experimental setup and hypotheses

In this dissertation, we seek to unravel whether certain temporal patterns between speakers during turn-taking may convey affiliation between those speakers, on top of the established effects of overlapping talk or silent gap between them. Specifically, we are interested in the social impact of (de)synchronization of speech rhythms and (mis)matching of speech tempos during conversation.

In three experiments we will explore this question.

1. In Study 1 (Chapter 2) we will explore whether synchronization conveys affiliation between speakers, on top of the effects of the amount of gap or overlap between their turns.

2. In Study 2 (Chapter 3) we will again investigate the affective impact of synchronization between speakers. Moreover, we will also investigate whether tempo matching too conveys positive affiliation between speakers.

3. In Study 3 (Chapter 4) we will again assess our synchronization hypothesis. We will investigate this question as in Study 1, with the exception that Study 3 will contain both an offline as well as an online measure of affiliation.

In the following paragraphs we present an outline of the general methodology of these studies, while Chapter 2, 3, and 4 present the exact details of each individual experiment.

1.5.1 General methodology

To explore whether tempo (mis)matching and/or (de)synchronization between speakers in conversation may convey (a lack of) affiliation between those speakers, we presented participants with short fragments of temporally manipulated dialogues between speakers and asked them to rate the perceived degree of affiliation between those speakers. In Study 3, we also recorded participants' real-time affective response during listening to these fragments.

Manipulations

The fragments presented to our participants consisted of synthesized, mini-dialogues between two speakers in which we manipulated the temporal patterns between, as well as within, each speaker. The use of synthesized speech gave us very precise control over the timing between, as well as within, speaking turns.

Synchronization was manipulated in terms of the location of the onset of the first stressed vowel of the second speaker. The first speaker would always have regular intervals between onsets of stressed vowels, thereby creating a regular rhythm or pulse. Crucially, the first stressed vowel onset of the second speaker came in either 'on the beat' of the rhythm of the first speaker, or it came in 'off the beat', being either too early or too late, relative to the (extrapolated) rhythm of the first speaker.

Tempo matching was manipulated by matching or mismatching the speech tempo of both speakers. In tempo-match exchanges, the tempos of both speakers were exactly the same, while in tempo-mismatch exchanges the tempos of both speakers were different.

Overlap was manipulated by placing the onset of the second turn in such a way that it either overlapped with the preceding turn, or that it did not overlap, containing a silent gap between turns.

Lastly, two different speech *tempos* were included to be able to control for the varying amount of gap and overlap, resulting from our synchronization manipulation. The rationale behind the choice of tempo variation will be discussed in the experimental chapters.

Dependent variables

Affiliation ratings Affiliation was assessed by instructing participants to focus on the reaction of the last speaker, in relation to the preceding speaker, and to look for cues, possibly very subtle, of whether that speaker sounded like he (or she) wanted to decrease or increase the social distance with the preceding speaker.

We chose to query affiliation in this specific way because it closely resembles our definition of affiliation, *viz.* enacted or experienced social closeness. Moreover, it directly connects affiliation to the approach–avoidance dimension, which is, as discussed in the previous section, central to general theories of emotion (Frijda, 2009; Panksepp & Biven, 2012), and emerges in accounts of affiliative versus social distancing functions of emotion (Fischer & Manstead, 2016).

Furthermore, this relatively open-ended notion of perceived affiliation allowed our participants to be sensitive to a very wide range of possible attributions of interpersonal matching and synchronization in conversation, such as the extent to which the second speaker liked, agreed, or approved with the first speaker, his stance, social intention, or the current conversational project (Van Berkum, in press, see section 1.4.1).

Corrugator activity In the Study 3 (Ch. 4), we also measured participants' corrugator, or frowning, activity. Corrugator activity has been shown to be a reliable indicator of the emotional valence of a given stimulus; high activity reflects negative valence, while low activity reflects the opposite (Van Boxtel, 2010). As pointed out, our critical manipulations of synchronization and overlap arise at the turn transition from one speaker to the next. Evaluated affiliation however, is measured at the end of each dialogue, which makes these ratings potentially less sensitive to our critical manipulations. Measuring participants corrugator activity during the experiment allowed us to track the *instantaneous*, and possible subconscious, affective response of our participant to our critical temporal manipulations. In combination with our affiliation ratings, corrugator activity may thus increase our understanding of the, possibly subtle and temporary, affective impact of synchronization and overlap between speakers in conversation.

1.5.2 Hypotheses

Based on the experimental research that we reviewed, as well as on the assumption that speech may be subject to the same phenomena as other forms of joint action, we test the following hypotheses:

Affiliation ratings

H1a Speakers who produce overlapping talk will be rated as *less* affiliative than speakers who do not overlap.

H2 Independent of tempo, speakers who do *not* match their speech tempo to that of the previous speaker will be rated as *less* affiliative than speakers who do match their speech tempo to that of the previous speaker.

H3a Independent of overlap and tempo, speakers who do *not* synchronize their speech rhythm to that of the previous speaker will be rated as *less* affiliative than speakers who do synchronize their speech rhythm to that of the previous speaker.

In addition to the affiliation ratings, we test two of these hypotheses by means of corrugator activity in Study 3.

Corrugator activity

H1b Listening to speakers who produce overlapping talk will *increase* corrugator activity compared to listening to speakers who do not overlap.

H3b Independent of overlap and tempo, listening to speakers who do *not* synchronize their speech rhythm to that of the previous speaker will *increase* corrugator activity compared to listening to speakers who do synchronize their speech rhythm to that of the previous speaker.

The remainder of this dissertation is organized as follows. Chapter 2, 3, and 4 present the three experiments to assess the affective implications of synchronization and tempo matching in conversation. The experimental results are discussed in Chapter 5, with conclusions drawn in the same chapter.

CHAPTER 2

Study 1: Synchronization, overlap and affiliation in conversation

2.1 Introduction

In dialogue, it does not only matter what you say, but also *when* you say it: the timing of a speaking turn has social significance. Silent gaps between speaking turns that are longer than 500-1000 ms may signal sub-optimal understanding (Beňuš et al., 2011) or unwillingness to agree to a request (Roberts et al., 2011). Similarly, overlapping talk between speakers may indicate an asymmetrical dominance relationship (Beňuš et al., 2011) and speakers who overlap are judged to be less agreeable, more rude, and more disrespectful than speakers who do not overlap (Goldberg, 1990; Maat et al., 2010).

Empirical research on the social implications of the timing of a speaking turn in dialogue has largely focused on the time intervals between speaking turns (Beňuš et al., 2011; Maat et al., 2010; Roberts et al., 2011). However, turn timing may also be described in terms of rhythm or tempo. Interlocutors may adjust their speech tempo to one another, or they may time their speech rhythm in such way that their first stressed syllable, or 'beat', is in line with the rhythm of the previous speaker. We will refer to the latter phenomenon as synchronization.

Although speech lacks strict periodicity, it is generally perceived as being rhythmical in nature (Lehiste, 1977). When imitating the timing of spoken sentences listeners tap with less temporal variability than the actual timing of the stressed syllables, whereas imitating noise does not show this pattern (Donovan & Darwin, 1979). Speakers can align their stressed syllables to the beat when speaking with a metronome (Port, 2003), and when speaking together speakers are able to speak in synchrony with one another, with only very short inter-speaker-lags (Cummins, 2009b). Moreover, when speaking together, speakers tend to locally display rhythmic alignment between turns (Ogden & Hawkins, 2015). These results provides evidence for synchronization between speakers in conversation.

Synchronization in spoken interaction is of interest because in other domains of human interaction, moving together in synchrony leads to mutual feelings of liking, compassion, and unity between the interacting partners (Vicaria & Dickens, 2016). Even merely observing synchronized interactions leads to attributions of unity and affiliation of the people being observed (Lakens, 2010; Lakens & Stel, 2011). These observations open up the possibility that in dialogue, synchronization has social significance as well, over and above the established effects of silent gap and overlap.

2.1.1 The present study

In this study we systematically investigated whether speech rhythm (de)synchronization gives rise to impressions of (a lack of) affiliation between speakers, on top of the established effects of silent gap and overlap. To explore this question, participants were presented with short, synthesized fragments of dialogue between speakers, in which we systematically manipulated synchronization, overlap, and tempo. Afterwards participants were asked to rate the degree of perceived affiliation between those speakers.

For our manipulation of synchronization, the second speaker (S2) did or did not synchronize his (or her) first stressed syllable to the rhythm of the first speaker (S1). In this study, the rhythm of the first speaker is defined as the (regular) interval between vowel onsets of stressed syllables, so that the rhythm of S1 functioned as the underlying 'beat' of the dialogue. In synchronized exchanges, the first stressed syllable of S2, as well as subsequent ones, came in on time, that is, 'on the beat' of the rhythm of S1 whereas in non-synchronized dialogues, the first stressed syllable of S2 was either 30% too early or 30% too late given the rhythm of S1. The core hypothesis was that speakers who do *not* synchronize their speech rhythm to the rhythm of S1 will be evaluated as *less* affiliative than speakers who do synchronize their speech rhythm to the rhythm of S1.

In addition, and orthogonal to our synchronization hypothesis, we also investigated the impact of overlap on affiliation. Given previously reviewed research on the social implications of overlapping talk, we expected that speakers producing overlap will be rated as *less* affiliative than speakers who do not produce overlap.

Third, and independent from our synchronization and overlap manipulation, speech tempo was also included as a factor: both speakers either spoke fast or slow. Tempo varied in order to disentangle the effects of synchronization and gap duration, as will be discussed in the section Materials 2.2.2. We had no a priori theoretical reasons to believe that tempo would modify rated affilition.

Perceived affiliation was assessed by asking participants to focus on the reaction of S2 to S1, and to rate on a 7-point scale, to what extent they heard signs of affiliation in that reaction. Affiliation was operationalized in these specific terms because it closely resembles the approach–avoidance dimension (Frijda, 2009; Panksepp & Biven, 2012). Social approach or avoidance tendencies, elicited by our emotions, help humans to decide whether to affiliate, or to distance oneself from someone (Fischer & Manstead, 2016). Query affiliation in this way is intuitive and easy to use, at least for our Dutch participants. Moreover, we believed that this simple rating of affiliation would allow us to (begin to) identify the basic processes underlying the interpersonal affective experience of conversation that we experience everyday.

We measured affiliation by means of post-hoc ratings of overhearers, rather than measuring the affiliative response of the interactants themselves. This may be justified because we know that humans are very sensitive to third-party interactions (Abdai & Miklóosi, 2016). Such sensitivity is often referred to as social eavesdropping. It involves the affective evaluation of third-party interactions, and it appears to develop very early in human infants (Hamlin, 2015). Social eavesdropping is beneficial because it helps us to recognize prosocial and antisocial group members, and it helps us to avoid harmful situations (Abdai & Miklóosi, 2016). Empathy is one of the possible underlying mechanisms of why overhearers are not only very sensitive in general, but also *affectively* responsive to (aspects of) third-party interactions. Empathy is a multidimensional construct comprising dissociable components: emotional resonance, empathic concern, and perspective taking, which allows us to thrive by detecting and responding to significant social events necessary for surviving, reproducing, and maintaining wellbeing (Decety & Cowell, 2014).

As previously discussed in Chapter 1, there are many causes in conversation which may give rise to a (dis)ffiliating response. The words itself, the referential situation, the affective stance, the social intention, the communicative intention, and even a bonus meaning may, amongst others, all be potential causes for (dis)affiliation for interlocutors in conversation (Van Berkum, in press). In our paradigm, we ruled out that our participants would take the individual words and the referential situation, outlined by our speakers, as a potential cause for affiliation. We do so by using ‘muffled’ speech in order to exclude lexical-semantic interference during listening. Nevertheless, even with muffled speech, the sounds of the speech stream, as well as the voices of the speakers, may have an inherent emotional quality (Myers-Schulz, Pujara, Wolf, & Koenigs, 2013; Zuckerman & Driver, 1989; Zuckerman, Hodgins, & Miyake, 1990). To rule out that the specific sequence of sounds, as well as the voice characteristics of the speaker, would influence the ratings of affiliation, we fully counterbalanced the specific sequences of sounds as well as the identity of the speaker, i.e., each sequence of sounds and each speaker occurred both in the first (turn 1) as well as in the second (turn 2) position.

Predictions

H1 Speakers who produce overlapping talk will be rated as *less* affiliative than speakers who do not overlap.

H2 Independent of overlap and tempo, speakers who do *not* synchronize their speech rhythm to that of the previous speaker will be rated as *less* affiliative than speakers who do synchronize their speech rhythm to that of the previous speaker.

2.2 Methodology

2.2.1 Participants

Participants were 60 native speakers of Dutch (age $M = 21.4$ years, $SD = 3.0, 48$ females). None of them reported having any hearing difficulties. Participants signed an informed consent before starting the experiment and they received a small fee for their contribution.

2.2.2 Materials

Stimuli consisted of spoken mini-dialogues between two speaker of opposite sex (3 male and 2 female voices). We used voices of opposite sex in order to facilitate the perceptual separation of the speakers in dialogue (Brokx & Nooteboom, 1982). All speech was synthesized with Fluency text-to-speech software (Derksen & Menert, 2013). Each mini-dialogue consisted of three comparable exchanges between two speakers. Each exchange within a mini-dialogue contained two turns (of two speakers). Each turn within an exchange contained eleven, randomly selected, syllables adhering to Dutch phonotactics. The rhythm of the first speaker's turn realized a regular inter-stress-interval, that is, regular intervals between vowel onsets of strong syllables (Ogden & Hawkins, 2015), thereby creating a regular rhythm or 'beat', consisting of 5 strong beats. Even though vowel onsets were spaced regularly, the spacing of the onsets of the intervening phonemes was not, as a result of their intrinsic duration. Each turn received one of six commonly used Dutch intonational contours ('t Hart, Collier, & Cohen, 1990), and all turns received final lengthening (Crystal & House, 1990) of factor 1.6 at the last stressed and subsequent unstressed syllable.

Synchronization

Synchronization was manipulated by placing the first stressed syllable of S2 'on' or 'off' the beat of the rhythm established by S1 (with 'phase' expressing the shift relative to the last stressed syllable of S1 as a proportion of the inter-stress-interval, ISI, of S1). In synchronized exchanges, the first stressed syllable of S2 was 'on the beat' of the rhythm of S1 ('on time', phase 0.0). In non-synchronized exchanges, the first stressed syllable of S2 was 'off the beat' of the rhythm of S1; the first stressed syllable of S2 was either shifted 30% forward or it was shifted 30% backward relative to the rhythm of S1 ('too early', phase -0.3, or 'too late', phase 0.3). Figure 2.1 shows schematic examples of our synchronization manipulation, containing a silent gap between turns.

Overlap

Exchanges either contained overlap (phases $-0.3, 0.0$ and 0.3, Figure 2.2) or were non-overlapping with an intermediate silent beat (phases $1.7, 2.0$ and 2.3, Figure 2.1).

In order to disentangle the effects of synchronization and overlap, phase 1.0 was not used because a pilot study showed that this would have resulted in a confound of those factors. That is, using phase 1 would have resulted in unequal distribution of synchronization and overlap: most of the 'too early' exchanges would have contained overlapping talk, while 'too late' exchanges generally would contain no overlap but a silent gap between speakers, and 'on time' exchanges would contain a mixture of overlap and gap.

Tempo

As can be seen, overlap duration (Figure 2.2) or gap duration (Figure 2.1) between S1 and S2 inevitably co-varied with our synchronization manipulation. On average, 'too early' exchanges have shorter gap durations (or larger overlaps), and 'too late' exchanges have longer gap durations (or smaller overlaps) than 'on time' exchanges.

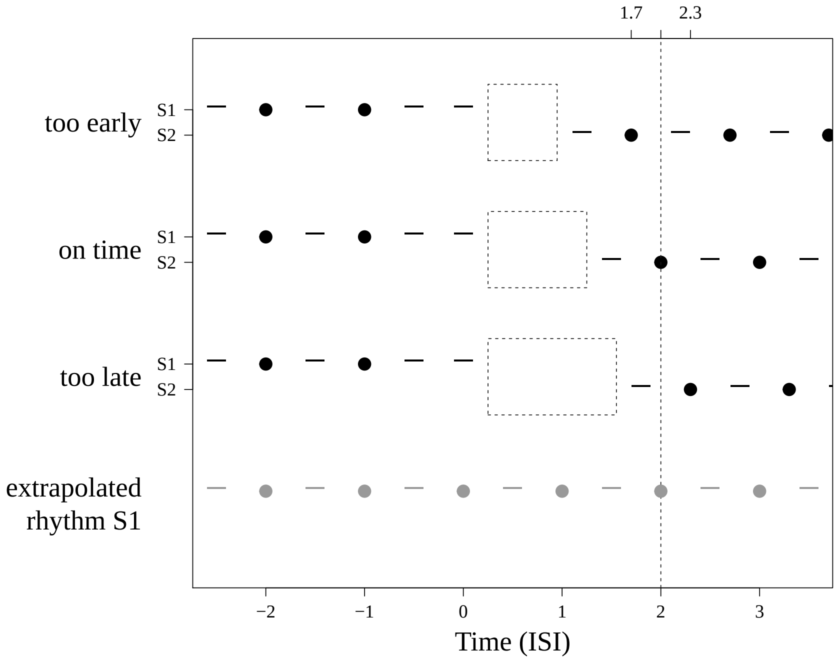

Figure 2.1. Schematic overview of our synchronization (and overlap) manipulation. This figure depicts the synchronization manipulation for exchanges containing a silent gap, and no overlap, between the two speakers (S1 and S2). Bullets represent stressed syllables, dashes unstressed ones, final lengthening is indicated by two dashes at the end of each turn, with the x-axis indicating normalized time in ISI (see text). Rectangular boxes represent the co-varying gap duration. The vertical dashed line represents the relevant, extrapolated, beat of S1, with S2 coming in 'too early', 'on time', or 'too late'.

We used two speech tempos to disentangle the effects of synchronization and of gap duration. In our paradigm, fast speech inherently has smaller gap durations and larger overlap durations than slow speech. If, as predicted, only overlap and synchronization influences the degree of affiliation, then we should find effects of overlap and synchronization only, without any interaction with tempo. We expected that tempo by itself would not influence the ratings of affiliation. Tempo was set at an inter-stress-interval of 525 ms ('slow') and at an inter-stress-interval of 425 ms ('fast').

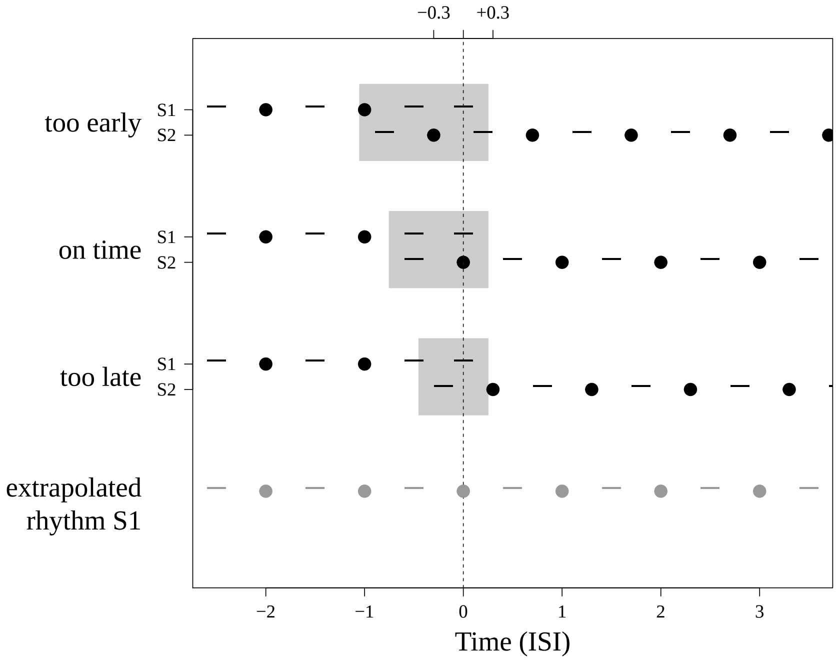

Figure 2.2. Schematic overview of our synchronization (and overlap) manipulation. This figure depicts the synchronization manipulation for exchanges containing an overlap, and no silent gap, between the two speakers (S1 and S2). Bullets represent stressed syllables, dashes unstressed ones, final lengthening is indicated by two dashes at the end of each turn, with the x-axis indicating normalized time in ISI (see text). Rectangular boxes represent the co-varying overlap duration. The vertical dashed line represents the relevant, extrapolated, beat of S1, with S2 coming in 'too early', 'on time', or 'too late'.

The resulting 216 unique exchanges each had 12 conditional variants: 3 synchronization conditions (on time, too early, too late) x 2 overlap conditions (no overlap, overlap) x 2 tempo conditions (fast, slow). Thus, the three factors, synchronization, overlap, and tempo were fully crossed. Each exchange was then combined with two other exchanges of the same conditional type and the same speakers to create a 'mini-dialogue'. This resulted in 72 mini-dialogues, consisting of three exchanges. Each conditional version of each mini-dialogue occurred in only one of 12 different experimental lists, with version, mini-dialogue, speaker voice, turn order, and intonational contour counterbalanced over lists. Additionally, each list was preceded by 4 practice items, consisting of items from a different list.

The resulting mini-dialogues were low-pass filtered (450 Hz cutoff, slope 25 Hz/octave) in order to avoid lexical-semantic interference, while preserving sufficient ecological plausibility and creating the illusion of naturally filtered speech, as if listening to people in the room next door.

2.2.3 Procedure

Participants were seated in a comfortable chair in a sound booth. Participants were instructed to focus on the second speaker in each exchange, and to look for evidence, possibly very subtle, whether the reaction of S2 contained cues that signaled affiliation. A trial consisted of the presentation of one entire mini-dialogue, containing three exchanges, after which the participants were prompted for their response. A visual display suggested that the three exchanges within a mini-dialogue were fragments from the same conversation. Audio presentation of the three exchanges within a trial were self-paced, and participants rated to what extent they heard signs of affiliation in the reaction of the second speaker across the three exchanges. We assumed that having three comparable exchanges would improve stability of the rating. The exact wording, in Dutch, of the question after each mini-dialogue was: *In de reactie van de tweede spreker hoorde ik:* ('In the reaction of the second speaker I heard'), and below that, as anchors for the extremes of the 7-point scale: *toenadering* ('approach') and *verwijdering* ('withdrawal'). A schematic example of a mini-dialogue trial is depicted in Figure 2.3.

Each participant was assigned one of the 12 experimental lists, and rated 72 mini-dialogues, divided in 8 blocks. In any given block, the 9 mini-dialogues shared the same tempo and overlap condition, but differed pseudo-randomly in terms of synchronization. In this way, each participant completed two blocks of fast, overlapping mini-dialogues, two blocks of slow, overlapping mini-dialogues, two blocks of fast non-overlapping mini-dialogues, and two blocks with slow, non-overlapping

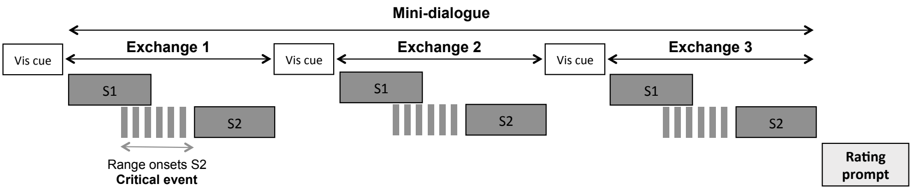

Figure 2.3. Illustration of a mini-dialogue trial. A visual cue helped participants imagine they were hearing three fragments sampled from a wider conversation; and were then presented with three exchanges between speaker 1 (S1) and speaker 2 (S2). The onset of S2 varied as a result of our critical manipulation of synchronization, tempo and overlap.

mini-dialogues. The order of blocks was randomized over lists. In the pauses between blocks a photo of natural scenery was presented, after which the participants had to press a button to proceed. After the experiment, participants filled out an exit questionnaire about the experiment, their language and musical background, and containing three subscales (Empathic Concern, Personal Distress, and Perspective Taking) of the Dutch version of the Interpersonal Reactivity Index (Davis, 1983, translated by Ponnet, Roeyers & Buysse, 2002). As none of these exit questionnaire measures reliably explained variance in the individual difference of the affiliation ratings, these measures will be ignored here. After the exit questionnaire, participants were debriefed and payed.

2.2.4 Analysis

Affiliation ratings were analyzed by means of linear mixed-effects modeling (LMM, Quené & Van den Bergh, 2008) using R (R Core Team, 2015), with package lmer4 (Bates, Maechler, Bolker, & Walker, 2015), and with participant as random intercepts. This model was then extended with different predictors in a particular order, as will be discussed below. After the addition of each predictor, and its interactions, the model fit was evaluated and compared to the previous model without that particular predictor (and interactions). Goodness of fit was explored using likelihood ratio tests (Pinheiro & Bates, 2000) and by inspecting the explained variance (R^2) of the extended model

(Nakagawa & Schielzeth, 2013). On the basis of the evaluation of the model fit we decided which predictor (and interactions) to keep and which to discard.

The main idea of this comparison is to find a model that gives an optimal fit of our data, balancing the chances between Type I and II errors (Matuschek, Kliegl, Vasishth, Baayen, & Bates, 2017). A full model approach, in which the model a priori consists of all experimental predictors and their interactions, may result in overparameterization. This in turn may result in a loss of power and spurious effects, if the complexity of the model is not supported by the data (Baayen, Davidson, & Bates, 2008; Matuschek et al., 2017).

The model fitting sequence was a follows. We first checked for individual differences, and later also for fixed effects, of tempo, then overlap, and finally synchronization. By first assessing whether variation in the data could be explained by our other predictors, i.e., tempo and overlap, ensured us that, would we find effects of synchronization, the effects of synchronization would be attributable to our synchronization manipulation solely, above and beyond the effects of the other variables. This particular sequence of tempo, overlap, and synchronization, was first used to define the random part of the model, before the fixed part was defined. This random-fixed sequence allowed us to assess whether the effects of tempo, overlap, and/or synchronization arise on top of the individual variation of these predictors.

The optimal model resulting from this model fitting procedure was then simplified by removing possible redundant interactions. If the fit did not decrease by removing these interactions, we continued with the simpler model. Finally, the model was then evaluated by bootstrapped confidence intervals (1,000 replications). Together, this conservative approach increased the likelihood that the final model would only contain genuine and robust effects.

2.3 Results

The model comparisons for the affiliation ratings of the current study are displayed in Table 2.2 (at the end of this chapter).

Random part First, the random part of our model was defined by adding item (mini-dialogue) as random intercepts (m0b). This slightly improved [$p = 0.06$] our model, and we decided to keep this predictor in our model. Tempo was then added as a random slope at the subject level (m0c) to assess individual variation in the effect of tempo. However, this addition led to convergence warnings, suggesting a poor fit or a lack of data. Because we had no a priori theoretical belief that tempo would influence the ratings of affiliation, we decided to discard tempo as a random effect. We then added overlap as a random slope (m0d), which improved the model fit. This indicates that subjects display individual differences in evaluated affiliation in response to overlapping and non-overlapping speakers. Synchronization was also added as a random slope (m0e), but this did not increase in the model fit, indicating that participants did not display individual difference in evaluated affiliation with regard to our synchronization manipulation.

Because overlap alone, and not synchronization, influenced the individual differences in rated affiliation, we reran model m0d, but this time without the subject intercepts, and with only the random slopes of overlap and no-overlap (m0f). In this way, we were able to compare the individual differences between the overlap and no-overlap effect directly instead of comparing the subject intercepts (on time, no overlap) with the individual differences of overlap.

Fixed part To define the fixed part of our model we first added tempo (m1). Against expectations, this improved the model fit, indicating that speech tempo does influence evaluated affiliation. Overlap, and its interaction with tempo, was then added (m2), which improved the model fit as well. Finally, we added synchronization, and its interaction with overlap, tempo and the interaction tempo by overlap (m3). In line with our prediction, this improved the model fit as well, indicating that synchronization does indeed influence evaluated affiliation. We also checked for effects of trial (m4), and its interactions, but this did not improve the model fit.

To simplify the model, the optimal model (m3) was then compared to the same model *without* all its interactions (m5), only containing the main effects of synchronization, overlap and tempo. This simpler model m5 was not as optimal as the more complex model m3 $[\chi^2(7) = 16.74, p = 0.02]$. The optimal model (m3) was then compared to another simpler model (m6), containing the interaction synchronization by overlap and the main effects of synchronization, overlap and tempo. Model m6 proved to be an equally good fit compared to the more complex model m3 $[\chi^2(5) = 9.67, p = 0.09]$.

Thus, the model m6, with random slopes of overlap, random intercepts of item, and with overlap, synchronization, and the interaction between overlap and synchronization, as fixed predictors proved to be an optimal fit of our data. The summary of this model m6 is displayed in Table 2.1, the fitted averages are shown in Figure 2.4 and Table 2.3 displays the fitted averages by synchronization and overlap (at the end of this chapter).

Tempo

Against expectations, tempo influenced evaluated affiliation: speakers who spoke slowly were evaluated as slightly more affiliative than speakers who spoke fast $[\beta = 0.11, 95\%\ C.I.(0.03, 0.19), S.E. = 0.04, t = 2.44]$.

Overlap

In line with our predictions, speakers who overlapped with the previous speaker were rated as less affiliative $[\beta = -0.45,\ 95\%\ C.I.\ (-0.65, -0.23), S.E. = 0.11, t = -4.31]$. Moreover, individual differences in rated affiliation for overlapping speakers explained a substantial amount of variance in the ratings, as indicated by the model comparison procedure as well as by the standard deviations of the subject slopes of overlap. This indicates that participants show more variance in rated affiliation in response to overlapping speakers than to non-overlapping speakers. The two subject slopes of no overlap and overlap do not correlate $[\beta = -0.32,\ 95\%\ C.I.\ (-0.76, 0.13)]$, suggesting that individuals with higher ratings of affiliation for non-overlapping speakers do not necessarily rate overlapping speakers as less affiliative, or vice versa. Thus, when listening to overlapping speakers, our participants showed less agreement in evaluated affiliation than when listening to speakers who do not produce overlap.

Table 2.1

Fitted estimates and bootstrapped 95% confidence intervals (C.I.) of evaluated affiliation based on model m6. The intercept refers to the on time, no overlap, fast condition. Evaluated affiliation ranges from 1 (avoidance) to 7 (approach).

Fixed effects:	Estimate (S.E.)	95% C.I.	t-value
Intercept	4.08 (0.07)	(3.94, 4.21)	60.83
Slow	0.11 (0.04)	(0.03, 0.19)	2.44
Overlap	-0.45 (0.11)	(-0.65,-0.23)	-4.31
Too early	-0.16 (0.08)	(-0.32,-0.02)	-2.14
Too late	-0.11 (0.08)	(-0.27, 0.05)	-1.47
Too early:Overlap	0.04 (0.11)	(-0.17, 0.27)	0.34
Too late:Overlap	0.27 (0.11)	(0.05, 0.48)	2.46

Random effects	S.D.	95% C.I.	Corr.	95%C.I.	n
No overlap (subj.)	0.23	(0.01, 0.31)			60
Overlap (subj.)	0.44	(0.33, 0.54)	-0.32	(-0.76, 0.13)	
Items	0.12	(0.00, 0.17)			72
Residual	1.45	(1.41, 1.48)			4320

Synchronization

Unexpectedly, we found an interaction between synchronization and overlap. For non-overlapping speakers we found, in line with our predictions, that being 'too early' was rated as less affiliative [$\beta = -0.16$, $95\%\ C.I.\ (-0.32, -0.02)$, $S.E. = 0.08, t = -2.14$]. Even though the direction of the effect of 'too late' for non-overlapping speakers was in the expected direction, it did not reach significance, as indicated by out bootstrap outcomes [$\beta = -0.11$, $95\%\ C.I.\ (-0.27, 0.05)$, $S.E. = 0.08$, $t = -1.47$].

For overlapping speakers, we found a similar pattern of being 'too early' compared to non-overlapping speakers [interaction $\beta = 0.04$, $95\%\ C.I.\ (-0.17, 0.27)$, $S.E. = 0.11, t = 0.34$]. The effect of being 'too late' however, was in the opposite direction compared to non-overlapping speakers [interaction $\beta = 0.27$, $95\%\ C.I.\ (0.05, 0.48), S.E. = 0.11, t = 2.46$]. That is, being 'too late' for overlapping speakers was rated as *more* affiliative than being 'too late' for non-overlapping speakers .

Thus, for both overlapping as well as for non-overlapping speakers, ‘too early’ was rated as *less* affiliative than being ‘on time’. For ‘too late’ however, we found mixed effects. For non-overlapping speakers, we found no difference in evaluated affiliation between ‘on time’ and ‘too late’, while for overlapping speakers ‘too late’ was evaluated as *more* affiliative.

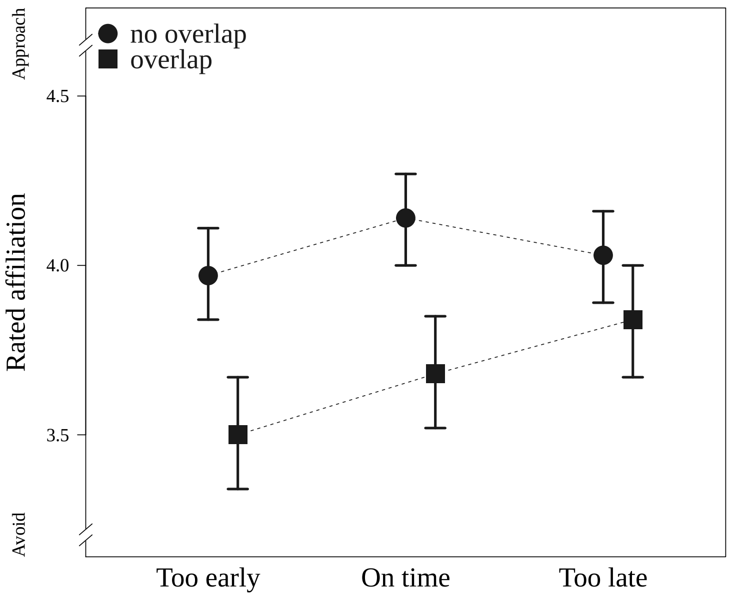

Figure 2.4. Bootstrapped (fitted) means and 95% confidence intervals of evaluated affiliation, based on model m6, broken down by synchronization and overlap. The round symbols represent the bootstrapped fitted means for the non-overlapping conditions, the square symbols the bootstrapped fitted means for the overlapping conditions. The bootstrapped 95% confidence intervals of each condition are represented by the vertical error bars, which may be slightly asymmetrical due to the bootstrapping outcomes. Evaluated affiliation ranges from 1 (avoidance) to 7 (approach).

2.4 Discussion

2.4.1 Overlap

As expected (Beňuš et al., 2011; Goldberg, 1990; Maat et al., 2010), speakers who overlapped their turns with that of the preceding speaker were rated as less affiliative than speakers who did not produce overlap. Moreover, individual differences in evaluated affiliation for overlapping were larger than for non-overlapping speakers, indicating that overlap may weight differently on evaluated affiliation for each listener.

2.4.2 Synchronization

Importantly, and in line with earlier research on synchronization-induced attributions of affiliation (Lakens & Stel, 2011; Vicaria & Dickens, 2016), synchronization contributed to evaluated affiliation, at least under some conditions. Irrespectively of tempo and overlap, speakers were evaluated as more affiliative when their turns were synchronized (on time) than when their turns were not synchronized, albeit only when they were too early relative to the rhythm of the previous speaker. Specifically, being 'too early' was evaluated as less affiliative as being 'on time', both for non-overlapping as well as for overlapping speakers.

Against the expectation that the affiliation ratings would be equally low for both non-synchronization conditions, being 'too late' did not result in lower ratings of affiliation. Instead we found an unexpected interaction between overlap and synchronization. For non-overlapping speakers, being 'too late' was equally affiliative as being 'on time'. For overlapping speakers on the other hand, being 'too late' was rated as relatively *more* affiliative than being 'on time'.

One possible explanation for the asymmetry between 'too early' and 'too late', at least for our non-overlapping speakers, might be that being 'too late' is not as socially disruptive as being 'too early' because being too late may also be interpreted, by the listener, as a cue for other speaker variables, such as fluency (Bosker, 2014), problems with comprehension (Beňuš et al., 2011), planning difficulties (Bull & Aylett, 1998), or word frequency (Hartsuiker & Notebaert, 2010). It may be that participants take these possible speaker challenges into account when evaluating affiliation between speakers, thereby attenuating the synchronization effect of 'too late'.

Alternatively, the null effect of being 'too late' may be the result of low statistical power. Low power reduces the likelihood of detecting a true effect (Ioannidis, 2005). We already suspected that our study may suffer from low statistical power because all our effects were small and because of the computational convergence warnings when assessing the individual differences of our predictors. Because the null effect of 'too late' may be a false-negative, our next study (Study 2, Chapter 3) aims reassessing the effects of synchronization for non-overlapping speakers.

Amount of overlap

For overlapping speakers, being 'too late' was rated as relatively *more* affiliative than being 'on time'.

This results of being 'too late' led to the suspicion that the effect of synchronization may be different, or even absent, in the contexts of the presence of an overlap, causing synchronization to yield different effects under different conditions. For the effect of synchronization between speaker who do overlap, evaluated affiliation increased step-wise between 'too early', 'on time', and 'too late'. Comparing the amount of overlap in Figure 2.2 with our levels of synchronization suggests that this step-wise increase in evaluated affiliation may be driven by the *amount* of overlap, and not by the hypothesized effect of synchronization. 'Too early' has generally larger overlaps, while 'too late' has generally smaller overlaps between speakers compared to the intermediate overlaps of 'on time'. This may well explain why 'too late' for overlapping speakers was rated as more affiliative, since those conversations contained only small overlaps between speakers. This suggests that overlap, and the amount of it, may constitute a more salient cue than synchronization, thereby washing away the possibly subtle effects of synchronization in the overlapping dialogues.

Nevertheless, as can be seen from the model comparisons (Table 2.2), there were no interactions with tempo, as one would expect if the amount of overlap influences evaluated affiliation. This indicates that gap duration, or overlap duration, alone cannot explain the ratings of affiliation. However, as previously pointed out, the absence of such tempo interaction may be explained by low statistical power as well.

Thus, even in the absence of interactions with tempo, we cannot completely rule out that, beyond the simple presence of overlap itself, the duration of overlap has social connotations too.

2.4.3 Speech tempo

Not anticipated was the effect of speech tempo. Irrespectively of overlap, speakers who spoke slowly were evaluated as more affiliative than speakers who spoke fast. As previously discussed, the rationale to include both tempos was to be able to disentangle the effects of the amount of gap or overlap from our synchronization manipulations. Our data however, shows main effects of all three predictors, i.e., synchronization, overlap and tempo, without any interactions with tempo. As previously pointed out, the absence of an interaction with tempo suggests that the duration of the silent gap or overlap between speakers, cannot explain the ratings of affiliation. Somehow slow speech has unknown social connotations by itself, above the effects of synchronization and overlap.

The higher ratings of affiliation for slow speech are in line with research about the pleasantness of speech and music that showed that fast speech is evaluated as less pleasant, and having greater tension than slow speech (Ilie & Thompson, 2006). Moreover, when talking to strangers, people tend to use longer turns and slower speech rates than when talking to acquaintances or friends (Yuan, Liberman, & Cieri, 2006). In our paradigm, the dialogues sounded muffled, and both turns were relatively long. Participants may well have perceived these dialogues as conversations between strangers. Especially under these circumstances, the slower speech might be perceived as more pleasant and more friendly than faster speech, and as a result, may have led to higher affiliation ratings.

Together, this indicates that listeners take temporal aspects *between* speakers, such as overlap and synchronization, as well as temporal aspects *within* speakers, as in the case of speech tempo, into account when evaluating affiliation between speakers.

2.4.4 Other issues

We have obtained our effect of turn synchronization with synthesized and low-pass filtered speech, which may raise questions about the ecological validity of our materials. However, our attempt to use filter settings that would lead to a 'muffled speech' experience seems to have been successful: many listeners expressed judgments such as 'it sounded Italian' or 'sometimes they sounded really agitated', indicating that they did experience the materials as natural conversational speech. It therefore seems reasonable to assume that in real everyday conversation, synchronization will have similar, albeit modest, social connotations.

In our paradigm, speaker affiliation was rated by overhearers, rather than by the interactants themselves. It may well be that overhearers weigh the temporal cues between and within speakers differently. It seems highly unlikely, however, that the perceptions of overhearers and interactants would involve radically different processing machinery. The more parsimonious position is that, like overlap, synchronization is a cue for all listeners, regardless of their role. Nevertheless, it might well be that by regarding the overhearer as a proxy of the interlocutor, the observed effect of synchronization is attenuated, which might also explain the small effect of synchronization in this study. For our overhearing participants, synchronization might not be of the same salience and relevance as for speakers in real-life conversations.

In this first study, we choose to use affiliation in the approach–avoidance dimension. We deliberately did not specify to our participants how they should interpret affiliation exactly, allowing them to be sensitive to a maximally wide range of possibilities, potentially covering the extent to which the second speaker liked, approved, aligned with or felt comfortable with the first speaker, his stance, or the current conversational project (principled distinctions derived from the Affective Language Comprehension model Van Berkum, in press). Which of these possibilities can drive turn synchronization needs to be further investigated.

2.5 Summary

Prior research on the timing of conversational turns has provided ample evidence for the social significance of overlapping talk and of silent gaps. Here we explored the possibility that other temporal patterns between speakers, in our case being in or out of ‘sync’ with the rhythm of the preceding speaker, has social significance as well. Although we do not yet fully understand why the phenomenon is limited in the way it is here, this study does reveal that, in addition to the presence of overlap, synchronization between speakers can indeed be taken as conveying affiliation between them.

Table 2.2

Models explaining the affiliation ratings of Study 1

Name	Model	R^2_{marg}	R^2_{cond}	Comparison	-2LogLik	Chi-squared
m0a	$1 + (1 \mid subj)$	0	0.019		-7902	
m0b	$1 + (1 \mid subj) + (1 \mid item)$	0	0.025	m0a vs m0b	-7900	$\chi^2(1) = 3.45, p = .06$
m0c	$1 + (1 + temp \mid subj) + (1 \mid item)$	0				Failed to converge
m0d	$1 + (1 + ovl \mid subj) + (1 \mid item)$	0	0.078	m0b vs m0d	-7833	$\chi^2(2) = 134.56, p < 0.001$
m0e	$1 + (1 + sync + ovl \mid subj) + (1 \mid item)$	0	0.078	m0d vs m0e	-7828	$\chi^2(7) = 8.88, p = 0.26$
m0f	$1 + (0 + ovl \mid subj) + (1 \mid item)$	0	0.078	m0b vs m0f	-7828	$\chi^2(2) = 134.56, p < 0.001$
m1	$1 + temp + (0 + ovl \mid subj) + (1 \mid item)$	0.001	0.077	m0e vs m1	-7830	$\chi^2(1) = 6.00, p = 0.01$
m2	$1 + ovl * temp + (0 + ovl \mid subj) + (1 \mid item)$	0.016	0.077	m1 vs m2	-7819	$\chi^2(2) = 20.59, p < 0.001$
m3	$1 + sync * ovl * temp + (0 + ovl \mid subj) + (1 \mid item)$	0.021	0.082	m2 vs m3	-7808	$\chi^2(8) = 22.53, p < 0.01$
m4	$1 + trial * sync * ovl * temp + (0 + ovl \mid subj) + (1 \mid item)$	0.025	0.087	m3 vs m4	-7800	$\chi^2(12) = 15.46, p = 0.22$
m5	$1 + sync + ovl + temp + (0 + ovl \mid subj) + (1 \mid item)$	0.017	0.078	m3 vs m5	-7808	$\chi^2(7) = 16.74, p = 0.02$
m6	$1 + sync + ovl + temp + sync : ovl + (0 + ovl \mid subj) + (1 \mid item)$	0.019	0.079	m3 vs m6	-7813	$\chi^2(5) = 9.67, p = 0.09$

Note: Following Bates et al. (2015), models are expressed as $1 + V1 + (1 + V1|V2)$ where the fixed part is represented by the components outside the brackets, and the random part by the components inside the brackets. In the fixed part, a 1 denotes to the global intercept, and V1 the fixed predictors. In the random part, a 1 denotes the intercepts of V2, and V1 the slope of the effect of V2, for each level of V2. A 0 is used when there are no intercepts. Asterisks (*) express interactions, including all component parts, i.e., all sub-interactions and main effects. Colons (:) express specific interactions. R^2_{marg} and R^2_{cond} are calculated based on Nakagawa and Schielzeth (2013).

Table 2.3

Bootstrapped (fitted) means and 95% confidence intervals (C.I.) of the affiliation ratings of model m6. Evaluated affiliation ranges from 1 (avoidance) to 7 (approach).

	No overlap		Overlap	
Synchronization	Mean	95% C.I.	Mean	95% C.I.
Too early	3.97	(3.84, 4.11)	3.50	(3.34, 3.67)
On time	4.14	(4.00, 4.27)	3.68	(3.52, 3.85)
Too late	4.03	(3.89, 4.16)	3.84	(3.67, 4.00)

CHAPTER 3

Study 2: Synchronization, tempo matching, and affilition in conversation

3.1 Introduction

In the introductory chapter, we reviewed evidence that suggests that certain temporal patterns between turn-taking speakers, namely synchronization and tempo matching, may convey affiliation between those speakers, on top of the established effects of silent gap and overlap between them.

Study 1 showed that this was indeed the case for synchronization. Speakers in non-synchronized dialogues were rated as less affiliative than speakers in synchronized dialogues, irrespectively of the presence of overlap. However, synchronization only influenced perceived affiliation when speakers responded ‘too early’, and not when they were ‘too late’, given the rhythm of the previous speaker.

For responding ‘too late’ we found conflicting evidence. While we predicted equally low affiliation ratings for speakers responding ‘too early’ and ‘too late’, our data revealed an interaction between overlap and synchronization.

For non-overlapping speakers responding 'too late' did not lower evaluated affiliation. Given that the (null) effect of being 'too late' in a non-overlapping context was in the right direction, we hypothesized that this might have been the result of low statistical power.

For overlapping speakers on the other hand, responding 'too late' was rated as *more* affiliative, while being 'too early' was rated as *less* affiliative than being 'on time'. Overlap duration co-varied with our synchronization manipulation, that is, 'too early' has largest, 'too late' smallest, and 'on time' intermediate overlaps. This co-variance led us to believe that the effect of synchronization for overlapping speakers was in fact the result of the amount of overlap, instead of a true synchronization effect. Our post-hoc reasoning to explain why we only found an effect of synchronization for non-overlapping speakers was that the presence and amount of overlap may constitute a more salient cue than synchronization, thereby washing away the possibly subtle effects of synchronization in the overlapping dialogues.

Moreover, although the effect of synchronization was corroborated by bootstrap testing procedures, the effect was only very small. Affiliation ratings for speakers in non-synchronized and non-overlapping dialogues were lowered with 0.16 points (too early), on a 7-point Likert scale, relative to speakers in synchronized dialogues (on time) of the same type. This is a decrease in evaluated affiliation of only 2% (see Table 2.1, Chapter 2).

Therefore, the first aim of the present study was to reassess our synchronization hypothesis in a narrower, more targeted context, namely specifically for speakers who do not produce overlap. We hoped that the less variable context of exchanges without any overlap would reinforce the possible subtle effects of synchronization.

3.1.1 Tempo matching

Secondly, the current study furthermore aimed at extending the findings of synchronization with tempo matching. Tempo matching is another form of temporal adjustment between speakers in which speakers adjust their speech tempo to match the speech tempo of the previous speaker.

As also discussed in Chapter 1, when speaking together, the tempos of both speakers tend to become more similar during the course of the interaction (Giles et al., 1991; Manson et al., 2013; Schultz et al., 2016; Street, 1984). Like behavioral matching during non-verbal joint action (Vicaria & Dickens, 2016), tempo matching too has been claimed to convey positive speaker evaluations.

Listeners who perceive a speaker's rate as faster than their own rate him as more competent and more attractive, whereas the opposite was true for listeners who perceived a speaker's rate as slower than their own (Buller & Aune, 1992; Street, 1984). Although speaker and listener were never in conversation together in these studies, these findings do suggest that, at least at some level, tempo matching may indeed reflect affiliation.

However, when investigating the impact of speech rate similarity on perceived mutual affiliation in real life conversations, Manson et al. (2013) found that when speakers' speech rate converged more strongly from the beginning to the end of the conversation, they were more likely to cooperate, but no effects on perceived affiliation were found. On the other hand, matching on other speech dimensions, such as linguistic style matching (see Niederhoffer & Pennebaker, 2002), did increase perceived affiliation. This suggests that even though speech rate similarity may reflect an increased tendency to cooperate, it may not necessarily also give rise perceptions of mutual affiliation.

Although it it unclear whether speech tempo matching in particular conveys affiliation between speakers, the amount of matching of all kinds of other vocal behaviors has been shown to be modified by interpersonal factors (Babel, 2010, 2012; Pardo et al., 2012). These findings are in line with the previously discussed communication accommodation theory (CAT, see Chapter 1, section 1.3.2, e.g., Giles, 2016). This theory poses that, in general, the more verbal matching ('accommodation') between interlocutors during communication the smaller the social distance between them and the more positive the evaluation of the interaction and the speakers within.

Matching in the *non-verbal* domain, i.e., behavioral matching ('mimicry'), has been long linked to affiliation too (Bavelas et al., 1986, 1988). People intending to affiliate with others are especially prone to match their behaviours (Lakin & Chartrand, 2003) and even non-conscious matching of gestures, postures, and mannerisms has been found to enhance the smoothness of the interaction and to foster liking (Chartrand & Bargh, 1999).

Taken together, the above suggest that matching of all kinds of verbal and non-verbal behavior during interaction leads to and reflects positive affiliation. Besides synchronization, tempo matching may therefore also be an important factor when investigating the social impact of temporal patterns between turn-taking speakers.

3.1.2 The present study

First, the current study sought to replicate the effect of synchronization of Study 1. Study 1 revealed that speakers who do not synchronize their speech rhythm to the previous speaker were evaluated as less affiliative than speakers who do synchronize, albeit in the context of responding ‘too early’. In the current study we again expected that speakers who do *not* synchronize their speech to the rhythm of the previous speaker will be evaluated as *less* affiliative than speakers who do synchronize their speech rhythm in this way. Given the findings of Study 1, we may particularly expect to find effects of synchronization on evaluated affiliation for speakers being ‘too early’.

Secondly, and in addition to our synchronization hypothesis, the present study also assessed whether tempo matching conveys similar positive evaluations of affiliation. We expected that speakers do *not* match their speech tempo to that of the previous speaker will be evaluated as *less* affiliative than speakers who do match their speech tempo to that of the previous speaker.

As in Study 1, participants were presented with synthesized, and temporally manipulated, mini-dialogues between two speakers after which participants were asked to rate perceived affiliation. In each exchange, the speech rhythm of the second speaker was either synchronized (on time) or not synchronized, i.e., ‘too early’ or ‘too late’ given the rhythm of the previous speaker. In addition to the previous study, and orthogonal to our synchronization manipulation, speakers would either have the same tempo (‘match’) or a different tempo (‘mismatch’).

Speech tempo was also included as a factor, and independent from our synchronization and tempo-matching manipulation. Speakers either spoke fast or slow. Tempo varied in order to disentangle the effects of synchronization and gap duration, as previously discussed in Chapter 2. Initially, we had no a priori theoretical reasons to believe that the speech tempo itself (compared to tempo matching or mismatching) would influence perceived affiliation. However, given the results of Study 1 we were prepared to find higher ratings of affiliation for slow compared to fast speech. Moreover, we expected that this effect would be more pronounced for the tempo of the second speaker, since participants were asked to focus on the reaction of this speaker in particular.

As in Study 1, affiliation was assessed by asking participants to focus on the reaction of S2 to S1, and to rate on a 7-point scale, to what extent they heard signs of affiliation in that reaction.

Predictions

H1 Speakers who do *not* synchronize their speech rhythm to that of the previous speaker will be rated as *less* affiliative than speakers who do synchronize their speech rhythm to that of the previous speaker. Given the findings of Study 1, we were prepared to find effects on evaluated affiliation of being 'too early' only.

H2 Independent of tempo, speakers who do *not* match their speech tempo to that of the previous speaker will be rated as *less* affiliative than speakers who do match their speech tempo to that of the previous speaker.

3.2 Methodology

3.2.1 Participants

Participants were 60 native speakers of Dutch (age $M = 22.0$ years, $SD = 4.1, 52$ females). None of them reported having any hearing difficulties. Participants gave written informed consent before starting the experiment and they received a small fee for their contribution.

3.2.2 Materials

The materials of this study consisted of a subset of the materials used in Study 1 (repeated here), supplemented with a restructured version of the same subset, as discussed in the next sections.

Stimuli consisted of spoken mini-dialogues between two speaker of opposite sex (3 male and 2 female voices) with all speech synthesized with Fluency text-to-speech software (Derksen & Menert, 2013). Each mini-dialogue consisted of three comparable exchanges between two speakers. Each exchange within a mini-dialogue contained two turns (of two speakers). Each turn within an exchange contained eleven, randomly selected, syllables adhering to Dutch phonotactics. The rhythm of the first speaker's turn realized a regular inter-stress-interval, that is, regular intervals between vowel onsets of strong syllables (Ogden & Hawkins, 2015), thereby creating a regular rhythm or 'beat', consisting of 5 strong beats. Even though vowel onsets were spaced regularly, the spacing of the onsets of the intervening phonemes was not, as a result of their intrinsic duration. Each turn received one of six commonly used Dutch intonational contours ('t Hart et al., 1990), and all turns received final lengthening (Crystal & House, 1990) of factor 1.6 at the last stressed and subsequent unstressed syllable.

Synchronization

To test our synchronization hypothesis, we used the non-overlapping materials of Study 1, because the targeted synchronization effect seemed only present in this context. As before, synchronization was manipulated by shifting the first beat of the second speaker (S2) 30% backwards or 30% forwards relative to the extrapolated rhythm of the first speaker (S1). This resulted in exchanges in which S2 was 'on time' (phase 2.0), 'too early' (phase 1.7), or 'too late' (phase 2.3) given the (extrapolated) beat of S1. 'Phase' expresses the shift relative to the last stressed syllable of S1 as a proportion of the inter-stress-interval (ISI) of S1. Figure 3.1 depicts a schematic example of our synchronization manipulations.

Tempo matching

In Study 1, speakers would always have the same tempo; both of them either spoke fast or slowly. Here, we used the two speech tempos to create tempo-matched ('match') and tempo-mismached ('mismatch') exchanges. The original materials, as described in Chapter 2, were used for the tempo-match exchanges. In these exchanges the tempo of both speakers was always exactly the same (fast-fast or slow-slow).

The tempo-mismatch exchanges were created on the basis of the tempo-match exchanges. Turn pairs (exchanges) of the original materials were split and then matched to their alternative tempo half. For example, the individual turns of a fast-fast and of a slow-slow tempo exchange were separated after which they were again combined crosswise, thus yielding a fast-slow and a slow-fast exchange. Importantly, the new turn partner would have the same sequence of syllables and the same speaker voice as the original turn partner. Thus, the recombined turn pairs yielded the tempo-mismatch exchanges (fast-slow and slow-fast) and the original turn pairs yielded the tempo-matched exchanges (fast-fast and slow-slow).

Tempo S2

Because affiliation was measured by asking participants to focus on the reaction of the second speaker, tempo will be expressed as the tempo of the second speaker ('tempo S2'). In this way, the tempo of S1 can be deduced from the combination of the factors tempo matching and tempo S2. 'Mismatch-fast' refers to slow-fast exchanges, 'mismatch-slow' to fast-slow exchanges, 'match-fast' to fast-fast exchanges, and 'match-slow' to slow-slow exchanges.

Schematic examples of our synchronization, tempo-matching, and tempo-S2 manipulation can be found in Figures 3.1 (match) and 3.2 (mismatch). As displayed in Figure 3.1, tempo matching is, in principle, independent from synchronization, that is, tempo matching does not necessarily entail that the rhythm of the two speakers are synchronized at the turn transition from one speaker to another (Figure 3.1, too early, and too late). On the other hand, synchronization is, in principle, also independent from speech tempo, as long as the first beat of S2 is synchronized with the beat of S1 (Figure 3.2, on time).

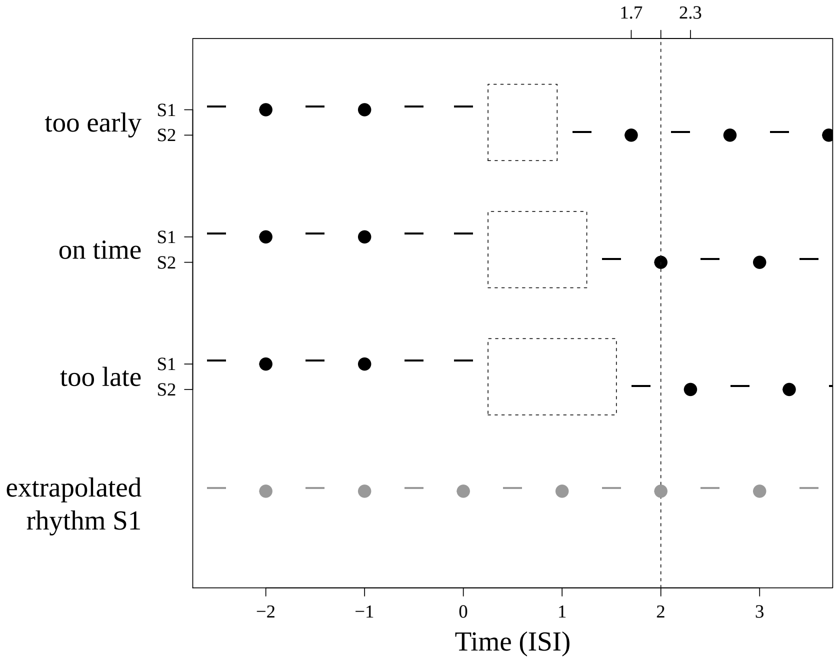

Figure 3.1. Schematic overview of our synchronization (and tempo-matching) manipulation. This figure depicts the synchronization manipulation for exchanges with matched speech tempos between the two speakers (S1 and S2). Bullets represent stressed syllables, dashes unstressed ones, final lengthening is indicated by two dashes at the end of each turn, with the x-axis indicating normalized time in ISI (see text). Rectangular boxes represent the co-varying gap duration. The vertical dashed line represents the relevant, extrapolated, beat of S1, with S2 coming in 'too early', 'on time', or 'too late'.

The resulting 216 unique exchanges each had 12 conditional variants; three synchronization conditions (on time, too early, too late), two tempo-S2 conditions (fast, slow), and two tempo-match conditions (match, mismatch). Thus, the three factors, synchronization, tempo S2, and tempo matching were fully crossed. Each exchange was then combined with two other exchanges of the same conditional type and the same speakers to create a 'mini-dialogue'. This resulted in 72 mini-dialogues, consisting of three exchanges.

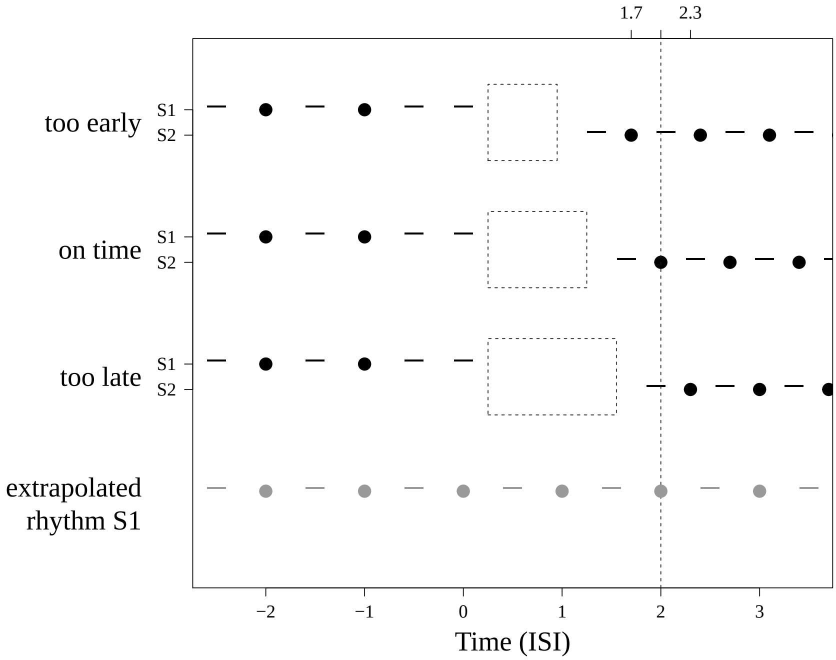

Figure 3.2. Schematic overview of our synchronization and tempo-matching manipulation. This figure depicts the synchronization manipulation for exchanges with mismatched speech tempos between the two speakers (S1 and S2). Bullets represent stressed syllables, dashes unstressed ones, final lengthening is indicated by two dashes at the end of each turn, with the x-axis indicating normalized time in ISI (see text). Rectangular boxes represent the co-varying gap duration. The vertical dashed line represents the relevant, extrapolated, beat of S1, with S2 coming in 'too early', 'on time', or 'too late'.

As in Study 1, the resulting mini-dialogues were low-pass filtered (450 Hz cutoff, slope 25 Hz/octave) in order to avoid lexical-semantic interference, while preserving sufficient ecological plausibility and creating the illusion of naturally filtered speech, as if listening to people in the room next door.

Each conditional variant of each mini-dialogue occurred in one of the 12 different experimental lists. Each participants completed one of the 12 experimental lists of 6 blocks of 12 mini-dialogues. The order of blocks was randomized over lists. The presentation of synchronization, of tempo S2, and of tempo matching was pseudo-randomized within list. As before, turn order (first turn or second turn) and speaker voice were fully counterbalanced over lists. Each list was preceded by 4 practice items, consisting of items from a different list. In the pauses between blocks a photo of natural scenery was presented, after which the participants had to press a button to proceed.

3.2.3 Procedure

The procedure and assessment of affiliation was the same as in Study 1, and is repeated here.

Participants were seated in a comfortable chair in a sound booth. Participants were instructed to focus on the second speaker in each exchange, and to look for evidence, possibly very subtle, whether the reaction of S2 contained cues that signaled affiliation. A trial consisted of the presentation of one entire mini-dialogue, containing three exchanges, after which the participants were prompted for their response. A visual display suggested that the three exchanges within a mini-dialogue were fragments from the same conversation. Audio presentation of the three exchanges within a trial were self-paced, and participants rated to what extent they heard signs of affiliation in the reaction of the second speaker across the three exchanges. We assumed that having three comparable exchanges would improve stability of the rating. The exact wording, in Dutch, of the question after each mini-dialogue was: *In de reactie van de tweede spreker hoorde ik:* ('In the reaction of the second speaker I heard'), and below that, as anchors for the extremes of the 7-point scale: *toenadering* ('approach') and *verwijdering* ('withdrawal'). A schematic example of a mini-dialogue trial is depicted in Figure 3.3.

After the experiment, participants filled out an exit questionnaire about the experiment, their language and musical background, and containing three subscales (Empathic Concern, Personal Distress, and Perspective Taking) of the Dutch version of the Interpersonal Reactivity Index (Davis, 1983, translated by Ponnet, Roeyers & Buysse, 2002). As none of these exit questionnaire measures reliably explained variance in the individual difference of the affiliation ratings, these measures will be ignored here. After the exit questionnaire, participants were debriefed and payed.

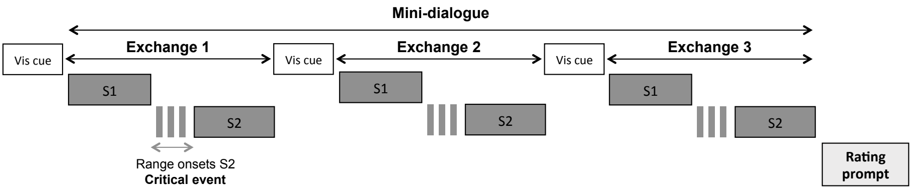

Figure 3.3. Illustration of a mini-dialogue trial. A visual cue helped participants imagine they were hearing three fragments sampled from a wider conversation; and were then presented with three exchanges between speaker 1 (S1) and speaker 2 (S2). Figure illustrates that the onset of S2 varied as a result of our critical manipulation of synchronization, tempo S2, and tempo matching manipulation.

3.2.4 Analysis

Similar to Study 1, the affiliation ratings were analyzed by linear mixed-effects modeling (LMM, Quené & Van den Bergh, 2008) using R (R Core Team, 2015), with package lmer4 (Bates et al., 2015), and with participant as random intercepts. This model was then extended with different predictors in a particular order, as will be discussed below. After the addition of each predictor, and its interactions, the model fit was evaluated and compared to the previous model without that particular predictor (and interactions). Goodness of fit was explored using likelihood ratio tests (Pinheiro & Bates, 2000) and by inspecting the explained variance (R^2) of the extended model (Nakagawa & Schielzeth, 2013). On the basis of the evaluation of the model fit we decided which predictor (and interactions) to keep and which to discard.

The model fitting sequence was a follows. We first checked for individual differences, and later also for fixed effects, of tempo S2, then of tempo matching, and finally synchronization. This particular sequence of tempo S2, tempo matching, and synchronization, was first used to define the random part of the model, before the fixed part was defined.

The optimal model resulting from this model fitting procedure was then simplified by removing possible redundant interactions. If the fit did not decrease by removing these interactions, we continued with the simpler model. Finally, this final model was then evaluated by bootstrapped confidence intervals (1,000 replications).

Together, this conservative approach increased the likelihood that the final model would only contain genuine and robust effects. A more detailed discussion about the rationale behind this model fitting procedure can be found in Chapter 2.

3.3 Results

The model comparisons for the affiliation ratings of the current study are displayed in Table 3.2 (at the end of this chapter).

Random part To define the random part of the model we first added item (mini-dialogue) as random intercepts (m0b). This did improve the model fit. However, we then checked for individual differences of the tempo of S2 (m0c). This addition of tempo S2 led to convergence warnings, but when we excluded item as random intercepts, and reran the model with tempo S2 as random slopes (m0d), the model fit increased. This indicates that model (m0c) was overparametrized given our data (Baayen et al., 2008). For this reasons we decided to continue with model m0d, and discard item as random intercepts. We then added tempo matching as a random slope (m0e) at the subject level. This did not improve the model fit, indicating that our participants did not display individual variation in evaluated affiliation in response to tempo matching or mismatching between speakers We also checked forindividual effects of synchronization (m0f), but this did not improve the model fit either, again indicating that our participants did not display individual variation in evaluated affiliation in response to our synchronization manipulation.

Because the tempo of S2, and not synchronization or tempo matching, influenced the individual ratings of affiliation, we reran model m0d without the subject intercept, and with only the random slopes of fast and slow speech (m0g). In this way, we were able to compare the individual differences on rated affiliation of the slow and fast tempo effects directly, instead of comparing the subject intercepts (on time, match) with the individual differences in evaluated affiliation for the slow tempo.

Fixed part For the definition of the fixed part, we first added tempo S2 (m1), which improved the model fit. We then checked for effects of tempo matching (m2). Against expectations, tempo matching did not lead to a better fit of the data. Similarly, when synchronization was added (m3), the model fit did not increase either, indicating that neither synchronization nor tempo matching did influence evaluated affiliation. Adding both tempo cmatching and synchronization, and their interactions (m4) did not improve the model fit, nor did adding trial (m5).

The model fitting procedure leads to model m4 as the optimal model for the purpose of the current analysis. Note that this is not the optimal fit given our data (which is m1), but model m4 includes both synchronization and tempo matching, the primary targets of the current study. A simpler model (m6) with only main effects of synchronization, tempo matching and tempo S2, and random subject slopes of tempo S2, without the subject intercepts, was an equally good fit of the data compared to model m4 [$\chi^2(7) = 7.44, p = 0.38$]. Here, we will present the results of this simpler model m6. The summary of model m6 can be found in Table 3.1, the fitted averages of model m6 are displayed in Table 3.3 (at the end of this chapter), and shown in Figure 3.4.

Table 3.1

Fitted estimates and bootstrapped 95% confidence intervals (C.I.) of evaluated affiliation, based on model m6. The intercept refers to the on time, S2 fast, match condition. Evaluated affiliation ranges from 1 (avoidance) to 7 (approach).

Fixed effects:	Estimate (S.E.)	95% CI	t-value
Intercept	3.85 (0.06)	(3.75, 3.96)	68.07
S2 Slow	0.24 (0.06)	(0.13, 0.36)	4.21
Mismatch	-0.06 (0.05)	(-0.16, 0.02)	-1.40
Too early	-0.01 (0.06)	(-0.11, 0.11)	-0.12
Too late	0.05 (0.06)	(-0.05, 0.16)	0.98

Random effects	S.D.	95% C.I.	Corr.	95%C.I.	n
S2 fast (subj.)	0.19	(0.06, 0.27)			60
S2 Slow (subj.)	0.25	(0.14, 0.32)	0.28	(-0.44, 1.00)	
Residual	1.51	(1.47, 1.54)			4320

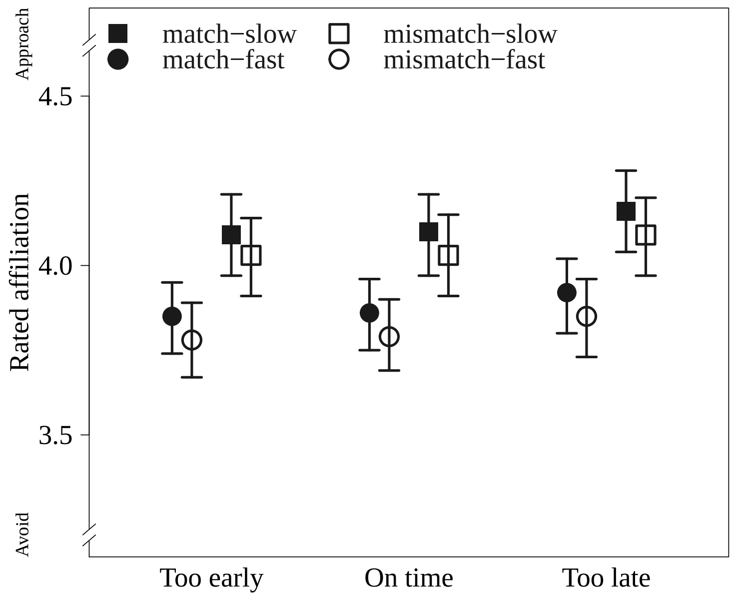

Figure 3.4. Bootstrapped (fitted) means and bootstrapped 95% confidence intervals of evaluated affiliation by synchronization, tempo S2, and tempo matching, based on model m6. The round symbols represent the bootstrapped fitted means for the fast S2 conditions, square symbols the bootstrapped fitted means for the slow S2 conditions. The filled symbols depict tempo-match conditions, the empty symbols tempo-mismatch conditions. The bootstrapped 95% confidence intervals of each condition are represented by the vertical error bars, which may be slightly asymmetrical due to the bootstrapping outcomes. Evaluated affiliation ranges from 1 (avoidance) to 7 (approach).

Tempo S2

First, irrespectively of tempo matching and synchronization, the tempo of the second speaker influenced the evaluations of affiliation; slow speech was rated more affiliative [$\beta = 0.24$, 95% $C.I.$ $(0.13, 0.36)$, $S.E. = 0.06, t = 4.21$].

Moreover, individual differences in rated affiliation for slow S2 speech were larger than for the individual differences in rated affiliation for fast S2 speech. No correlation was found between the two subject slopes of tempo S2 [$r = 0.28, 95\% \, C.I.(-0.44, 1.00)$].

Tempo matching

Second, and unexpectedly, tempo mismatching did not lead to lower ratings of affiliation [$\beta = -0.06$, $95\% \, C.I. \, (-0.16, 0.02)$, $S.E. = 0.05$, $t = -1.40$].

Synchronization

Third, and crucially, evaluated affiliation was not affected by synchronization either; being 'too early' did not influence the ratings of affiliation [$\beta = -0.01$, $95\% \, C.I. \, (-0.11, 0.11)$, $S.E. = 0.06, t = -0.12$], nor did being 'too late' [$\beta = 0.05$, $95\% \, C.I. \, (-0.05, 0.16)$, $S.E. = 0.06, t = 0.98$].

In sum, the predicted effects of tempo matching and synchronization were not confirmed. Neither tempo matching, nor synchronization influenced evaluated affiliation. Instead, the tempo of the second speaker influenced evaluated affiliation; when second speakers spoke slowly they were evaluated as more affiliative than when they spoke fast.

3.4 Discussion

The first aim of the current study was to replicate the synchronization finding of Study 1 for speakers who do not overlap. The second aim of this study was to extend these findings with tempo matching, and explore whether tempo matching leads to similar positive evaluations of affiliation.

Against expectations, the present study did not reveal evidence for synchronization nor for tempo matching on the impression of affiliation. Perceived affiliation was, only influenced by the speech tempo of the second speaker; second speakers who spoke slowly were evaluated as more affiliative than second speakers who spoke fast.

3.4.1 Experimental design

Although the methodology of Study 2 was as similar as possible to the one used in Study 1, it contained one potentially important dissimilarity. In Study 1 the different synchronization conditions were presented in a mixed fashion, while the additional conditions of tempo and overlap were presented in a blocked fashion. In this way, participants would be maximally sensitive to our synchronization manipulation. In Study 2 however, synchronization as well as tempo were presented in a mixed fashion. This was the result of our choice to present the conditions of tempo matching, like those of synchronization, in a mixed fashion as well. In other words, in Study 1 only synchronization varied per trial within a given block, whilst in Study 2 this was the case for both synchronization, tempo, and tempo matching.

Although this kind of randomizing reduces the risk of accidental biases, the less restricted context of varying tempo might have been too variable for our participants to be able to attune to our subtle manipulations of synchronization and tempo matching. This idea is reinforced by the current effect of the tempo of the second speaker (tempo S2), as if participants only used the last temporal cue available.

The tempo of S2 is present throughout the whole turn of the second speaker, while synchronization and tempo matching are much more subtle and short-lived; they have to be inferred from the interplay between the two speakers. For our participants, the tempo of the second speaker appeared to be a more proximate or salient cue than synchronization and tempo matching. Thus, in Study 2, the already subtle effect of synchronization found in Study 1 might have been washed away under the more variable context of mixed tempo.

3.4.2 Tempo

In our previous study, slow tempo led to higher affiliation ratings. This slow tempo effect was confirmed by the current study; when second speakers spoke slowly, affiliation ratings were higher than when second speakers spoke fast. As already discussed in Study 1, the higher ratings of affiliation for slow speech are in line with research about the pleasantness of speech and music (Ilie & Thompson, 2006). The tempo effect of the current study was independent from synchronization and tempo matching, indicating that the tempo of the *responding* speaker was the main cause of this effect, at least in our current paradigm with tempo presented in a mixed fashion, and with listeners instructed to pay special attention to the reaction of the second speaker.

3.4.3 Sensitivity

The small estimation of the effect of synchronization in Study 1 together with the null result of synchronization and tempo matching in the current study led to questions about the sensitivity of our measure of affiliation. We measured the amount of affiliation offline, after the end of each dialogue. It might well be that at that point the reaction to our critical manipulation has already been extinguished by other, more current, or more salient factors, such as the tempo of the second speaker (current study), or overlap (Study 1). The potentially subtle effects of our synchronization and tempo matching manipulation might thus be too weak to cause an effect large enough to be noticeable in our post-hoc affiliation ratings.

An online measure might be a good addition to our coarse offline 7-point affiliation scale. Measuring participants' corrugator or 'frowning' activity during the experiment might be a good candidate for such online measure. Corrugator, i.e., 'frowning', activity is a reliable indicator of the emotional significance of a given stimulus (Larsen, Norris, & Cacioppo, 2003; Van Boxtel, 2010). Moreover, facial electromyography, by which we reliably record corrugator activity, can accurately track rapid changes in corrugator activity, even if these changes are visually undetectable or occur subconsiously (Tassinary, Cacioppo, & Vanman, 2007; Van Boxtel, 2010). Measuring participants corrugator activity during the experiment allowed us to track the *instantaneous*, and possible subconscious, affective response of our participant to our critical temporal manipulations. In combination with our affiliation ratings, corrugator activity may thus increase our understanding of the, possibly subtle and temporary, affective impact of synchronization between speakers in conversation.

Therefore, our next study (Chapter 4) will again investigate the affective impact of synchronization between speakers in conversation, yet this time affiliation will, in addition to our post-hoc affiliation ratings, also be assessed by tracking participants' corrugator or 'frowning' activity.

Table 3.2

Models explaining the affiliation ratings of Study 2

Name	Model	R^2_{marg}	R^2_{cond}	Comparison	-2LogLik	Chi-squared
m0a	$1 + (1 \mid subj)$	0	0.013		-7958	
m0b	$1 + (1 \mid subj) + (1 \mid item)$	0	0.022	m0a vs m0b	-7954	$\chi^2(1) = 7.73, p < 0.01$
m0c	$1 + (1 + tempS2 \mid subj) + (1 \mid item)$	0				Failed to converge
m0d	$1 + (1 + tempS2 \mid subj)$	0	0.027	m0a vs m0d	-7949	$\chi^2(2) = 19.50, p < 0.001$
m0e	$1 + (1 + match + tempS2 \mid subj)$	0	0.029	m0d vs m0e	-7846	$\chi^2(3) = 4.40, p = 0.22$
m0f	$1 + (1 + sync + tempS2 \mid subj)$	0	0.029	m0d vs m0f	-7846	$\chi^2(7) = 4.20, p = 0.76$
m0g	$1 + (0 + tempS2 \mid subj)$	0	0.027	m0a vs m0d	-7949	$\chi^2(2) = 19.50, p < 0.001$
m1	$1 + tempS2 + (0 + tempS2 \mid subj)$	0.006	0.027	m0d vs m1	-7941	$\chi^2(1) = 15.50, p < 0.001$
m2	$1 + match * tempS2 + (0 + tempS2 \mid subj)$	0.007	0.028	m1 vs m2	-7939	$\chi^2(2) = 4.47, p = 0.11$
m3	$1 + sync * tempS2 + (0 + tempS2 \mid subj)$	0.009	0.029	m1 vs m3	-7840	$\chi^2(4) = 1.73, p = 0.79$
m4	$1 + match * sync * tempS2 + (0 + tempS2 \mid subj)$	0.009	0.029	m1 vs m4	-7835	$\chi^2(10) = 10.86, p = 0.37$
m5	$1 + trial * match * sync * tempS2 + (0 + tempS2 \mid subj)$	0.013	0.034	m1 vs m5	-7926	$\chi^2(22) = 30.59, p = 0.10$
m6	$1 + sync + match + tempS2 + (0 + tempS2 \mid subj)$	0.007	0.028	m4 vs m6	-7835	$\chi^2(7) = 7.44, p = 0.38$

Note: see Table 2.2 (Chapter 2) and Bates et al. (2015) regarding the format of the models. R^2_{marg} and R^2_{cond} are calculated based on Nakagawa and Schielzeth (2013).

Table 3.3

Bootstrapped (fitted) means and bootstrapped 95% confidence (C.I.) intervals of the affiliation ratings, based on model m6 broken down by tempo S2, synchronization, and tempo matching. Evaluated affiliation ranges from 1 (avoidance) to 7 (approach).

Tempo S2	Synchronization	Tempo match		Tempo mismatch	
		Mean	95% C.I.	Mean	95% C.I.
	Too early	3.85	(3.74, 3.95)	3.78	(3.67, 3.89)
S2 fast	On time	3.86	(3.75, 3.96)	3.79	(3.69, 3.90)
	Too late	3.92	(3.80, 4.02)	3.85	(3.73, 3.96)
	Too early	4.09	(3.97, 4.21)	4.03	(3.91, 4.14)
S2 slow	On time	4.10	(3.97, 4.21)	4.03	(3.91, 4.15)
	Too late	4.16	(4.04, 4.28)	4.09	(3.97, 4.20)

CHAPTER 4

Study 3: Synchronization, overlap, affiliation, and corrugator activity in conversation

4.1 Introduction

In Study 1, we tested our synchronization hypothesis for the first time. In line with our predictions, we found that speakers who synchronized their speech rhythm to that of the previous speaker were evaluated as more affiliative than speakers who did not adjust their speech rhythm in this way. The effect of synchronization, however, was small and only present in part of the design. Desynchronization only led to a decrease in evaluated affiliation if speakers responded too early, but not when they responded too late given the rhythm of the first speaker. In Study 2, we therefore reassessed our synchronization hypothesis in a different context. Contrary to our predictions, the previous effect of synchronization was not replicated in Study 2.

Sensitivity

The conflicting results of these two studies lead to questions about the sensitivity of our measure of affiliation. In both studies, participants first listened to (manipulated) speakers in dialogue, and were then asked to *rate* the extent of affiliation between those speakers. In hindsight, this experimental setup may have suffered from two drawbacks. The first drawback of asking participants to rate the stimuli is that such rating primarily relies on an episodic memory of the *conscious* affective experience or evaluation of the participants. (De)synchronization between speakers might impact a deeper, more subconscious level of affective processing or evaluation as well, which self-reported ratings might not be able to track. Another possible drawback of our self-reported rating is that we measured it at the *end* of each dialogue. At that point, the reaction of our participants to our critical synchronization manipulation, which occurs at the turn transition from the first speaker (S1) to the second (S2), may already have been extinguished; either by itself, or it may have been washed out by other, more recent or more salient factors, such as the tempo of the second speaker (Study 2), or the presence of overlap (Study 1). The effect of synchronization might be too subtle, too weak or too short-lived, to impact the post-hoc affiliation ratings. The lack of reproducibility, the small effect, and the fact that we only found evidence for synchronization in part of our design is consistent with this view.

Corrugator activity

Measuring corrugator activity during the experiment might resolve these two issues. Because we reliably frown more when we evaluate something as being negative and less when we evaluate something as being positive (Larsen et al., 2003) the activity of the corrugator supercilii or 'frowning muscle' is a useful indicator of the emotional valence of a given stimulus (Van Boxtel, 2010). The activity of the corrugator has a strong negative linear relationship with emotional valence, ranging from positive to negative (Larsen et al., 2003; Van Boxtel, 2010); corrugator activity increases when we evaluate a given stimulus as negative, and decreases when we evaluate a given stimulus as positive (Van Boxtel, 2010). Surface facial electromyography (facial EMG) can accurately record corrugator activity (Van Boxtel, 2010). By using facial EMG, affective responses may be detected, even if they are visually undetectable or occur subconsciously (Van Boxtel, 2010). Moreover, the temporal resolution of fEMG is such that it can capture fast and subtle facial musculature changes, so that rapid changes in corrugator activity can be reliably tracked (Tassinary et al., 2007; Van Boxtel, 2010).

Corrugator activity might serve as a measure of the affective valence of our critical temporal manipulations, in particular of our synchronization manipulation. To know the affective valence of our critical temporal manipulations is important because we assumed that the impression of affiliation arises under the influence of the affective system of the listener. Especially in relation to the ratings of affiliation, information about the affective valence of our temporal manipulations may broaden our understanding of the, possible subtle, affective impact of synchronization between speakers in conversation that we experience every day.

In addition to post-hoc self-report ratings of affiliation, the current study therefore also assessed our synchronization hypothesis by means of participants' activity of their corrugator muscle. In this way, we were able to track participants' post-hoc evaluation of affiliation (ratings) of our manipulated dialogues, as well as their instantaneous, affective (corrugator) response at the critical transition from S1 to S2. The transition from S1 to S2 is the moment in time at which our synchronization manipulation kicks in: S2 either responds 'on time' (synchronized), 'too early' or 'too late' given the rhythm established by S1.

Affective sensitivity to third-party conversations

As before, we measured affiliation and affective valence by means of *overhearers*, rather than measuring the affective response of the interactants themselves. This is justified because we know that humans are very sensitive to third-party interactions (Abdai & Miklóosi, 2016). Such sensitivity is often referred to as social eavesdropping. It involves the affective evaluation of third-party interactions, and it appears to develop very early in human infants (Hamlin, 2015). Social eavesdropping is beneficial because it helps us to recognize prosocial and antisocial group members, and it helps us to avoid harmful situations (Abdai & Miklóosi, 2016).

Empathy is one of the possible underlying mechanisms of why overhearers are not only very sensitive in general, but also *affectively* responsive to (aspects of) third-party interactions (Abdai & Miklóosi, 2016).

The nature and definition of empathy is a topic of ongoing debate, but most scholars agree that empathy should be seen as a complex multi-dimensional construct involving cognitive, affective, and motor processes that all concern the ability to understand and, more importantly, to share anothers emotional state (Decety & Cowell, 2014). Emotional sharing, or resonance, reflects the capacity to share, resonate or become affectively aroused by the emotions of others (in at least in valence, tone, and intensity). This may happen automatically, without a distinction between ones own emotions and the emotions of the other ('emotional resonance'), or it may happen with at least some differentiation between self and other, involving mental reasoning and social competence ('perspective taking, see Decety & Cowell, 2014). If something about S2s affective state is expressed by our synchronization manipulation, then participants might be able to pick up onto this feeling, and, as a result, also experience this emotion themselves, at least to some extent.

Our overhearing participants may also be affectively responsive to our synchronization manipulation because, for them, (de)synchronization conveys interpersonal discord or friction, signalling a potentially harmful situation that should be avoided. The detection of such a possible harmful situation might by itself elicit an affective response, without taking any of the speakers' affective states into consideration. Taken together, empathy, as well as more general conflict sensitivity may underlie our sensitivity to the affective dynamics of third-party conversation.

4.1.1 The present study

In the current study, we again investigated our synchronization hypothesis, but in addition to post-hoc self-report ratings of affiliation, the affective responses of our participants were also measured by recording participants' corrugator activity. We investigated whether consciously rated speaker affiliation (ratings) and subconscious affective responses (corrugator) were modified as a function of synchronization between overheard speakers in conversation. We expected to find an *increase* in corrugator activity, i.e., more frowns, and *lower* affiliation ratings for speakers who do *not* synchronize their speech rhythm to that of the previous speaker, compared to speakers who do synchronize.

As in Study 1, and orthogonal to our synchronization hypothesis, we also investigate the impact of the presence of overlapping talk on affiliation. Overlap is another temporal cue between speakers that might be relevant for the affective dynamics of interpersonal communication. Speakers who produce turn overlap are generally viewed as rude and disrespectful (Goldberg, 1990). The results of Study 1 confirmed this view: speakers who overlapped with the talk of the preceding speaker were evaluated as less affiliative than speakers who did not produce turn overlap. Here, we expected to find an *increase* in corrugator activity (more frowns) and *lower* affiliation ratings for speakers who overlap, compared to speakers who do not produce overlap.

Independent from our synchronization and overlap manipulation, speech tempo was also included as a factor; both speakers either spoke fast or slowly. Tempo varied in order to disentangle the effects of synchronization and gap duration, as will be discussed in the section Materials. We had no a priori theoretical reasons to believe that tempo would influence corrugator activity or rated affiliation. However, given the results of Study 1 and 2 we were prepared to find higher ratings of affiliation for slow compared to fast speech.

As in our previous studies, we presented listeners with synthesized dialogues between two speakers after which they were asked to rate perceived affiliation between those speakers. Contrary to Study 1 and 2, only two instead of three exchanges were presented per trial. This was the result of the inclusion of corrugator activity, which required a slightly adapted trial structure, as will be discussed in section 4.2. As in Study 1 and 2 we also investigated affiliation by self-report again, to be able to examine the replicability of our previous synchronization findings.

Predictions

- H1a Speakers who produce overlapping talk will be rated as *less* affiliative than speakers who do not overlap.

- H1b Listening to speakers who produce overlapping talk will *increase* corrugator activity compared to listening speakers who do not overlap.

H2a Independent of overlap and tempo, speakers who do *not* synchronize their speech rhythm to that of the previous speaker will be rated as *less* affiliative than speakers who do synchronize their speech rhythm to that of the previous speaker.

H2b Independent of overlap and tempo, listening to speakers who do *not* synchronize their speech rhythm to that of the previous speaker will *increase* corrugator activity compared to listening to speakers who do synchronize their speech rhythm to that of the previous speaker.

4.2 Methodology

4.2.1 Participants

Participants were 60 native speakers of Dutch (age $M = 21.5$ years, $SD = 3.5, 48$ females). None of them reported having any hearing difficulties, and none of them participated in any of the previous experiments. Participants signed an informed consent before starting the experiment and they received a small fee for their contribution.

4.2.2 Materials

Stimuli consisted of a subset of the materials of Study 1 (repeated below). In contrast to Study 1, each trial was preceded by a 2 seconds interval necessitated by the fEMG recordings. The added interval consisted of the presentation of an unrelated picture presented for 2 seconds before the start of each trial. In order to compensate for the extra time of the added interval; only the first two, instead of three, exchanges in the mini-dialogues of Study 1 were presented.

As before, stimuli consisted of spoken mini-dialogues between two speaker of opposite sex (3 male and 2 female voices) with all speech synthesized with Fluency text-to-speech software (Derksen & Menert, 2013). Each mini-dialogue consisted of two comparable exchanges between two speakers. Each exchange within a mini-dialogue contained two turns (of two speakers). Each turn within an exchange contained eleven, randomly selected, syllables adhering to Dutch phonotactics.

The rhythm of the first speaker's turn realized a regular inter-stress-interval, that is, regular intervals between vowel onsets of strong syllables (Ogden & Hawkins, 2015), thereby creating a regular rhythm or 'beat', consisting of 5 strong beats. Even though vowel onsets were spaced regularly, the spacing of the onsets of the intervening phonemes was not, as a result of their intrinsic duration. Each turn received one of six commonly used Dutch intonational contours ('t Hart et al., 1990), and all turns received final lengthening (Crystal & House, 1990) of factor 1.6 at the last stressed and subsequent unstressed syllable.

Synchronization

Synchronization was manipulated by placing the first stressed syllable of S2 'on' or 'off' the beat of the rhythm established by S1 (with 'phase' expressing the shift relative to the last stressed syllable of S1 as a proportion of the inter-stress-interval, ISI, of S1). In the synchronized exchanges, the first stressed syllable of S2 was 'on the beat' of the rhythm of S1 ('on time', phase 0.0). In non-synchronized exchanges, the first stressed syllable of S2 was 'off the beat' of the rhythm of S1; the first stressed syllable of S2 was either shifted 30% forward or it was shifted 30% backward relative to the rhythm of S1 ('too early', phase -0.3, or 'too late', phase 0.3). Figure 4.1 shows schematic examples of our synchronization manipulation, containing a silent gap between turns.

Overlap

Exchanges either contained overlap (phases $-0.3, 0.0$ and 0.3, Figure 4.2) or were non-overlapping with an intermediate silent beat (phases $1.7, 2.0$ and 2.3, Figure 4.1).

Tempo

As can be seen, overlap duration Figure 4.2) or gap duration (Figure 4.1) between S1 and S2 inevitably co-varied with our synchronization manipulation. On average, 'too early' exchanges have shorter gap durations (or larger overlaps), and 'too late' exchanges have longer gap durations (or smaller overlaps) than 'on time' exchanges. We used two

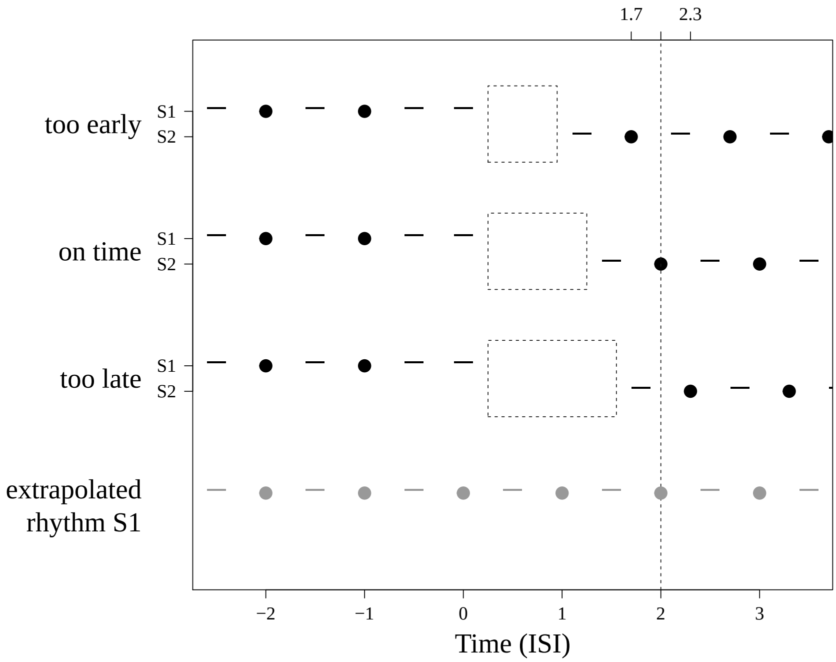

Figure 4.1. Schematic overview of our synchronization (and overlap) manipulation. This figure depicts the synchronization manipulation for exchanges containing a silent gap, and no overlap, between the two speakers (S1 and S2). Bullets represent stressed syllables, dashes unstressed ones, final lengthening is indicated by two dashes at the end of each turn, with the x-axis indicating normalized time in ISI (see text). Rectangular boxes represent the co-varying gap duration. The vertical dashed line represents the relevant, extrapolated, beat of S1, with S2 coming in 'too early', 'on time', or 'too late'.

speech tempos to disentangle the effects of synchronization and of gap duration. In our paradigm, fast speech inherently has smaller gap durations and larger overlap durations than slow speech. If, as predicted, only overlap and synchronization influences the degree of affiliation, then we should find effects of overlap and synchronization only, without any interaction with tempo. We expected that tempo by itself would not influence the ratings of affiliation. Tempo was set at an inter-stress-interval of 525 ms ('slow') and at an inter-stress-interval of 425 ms ('fast').

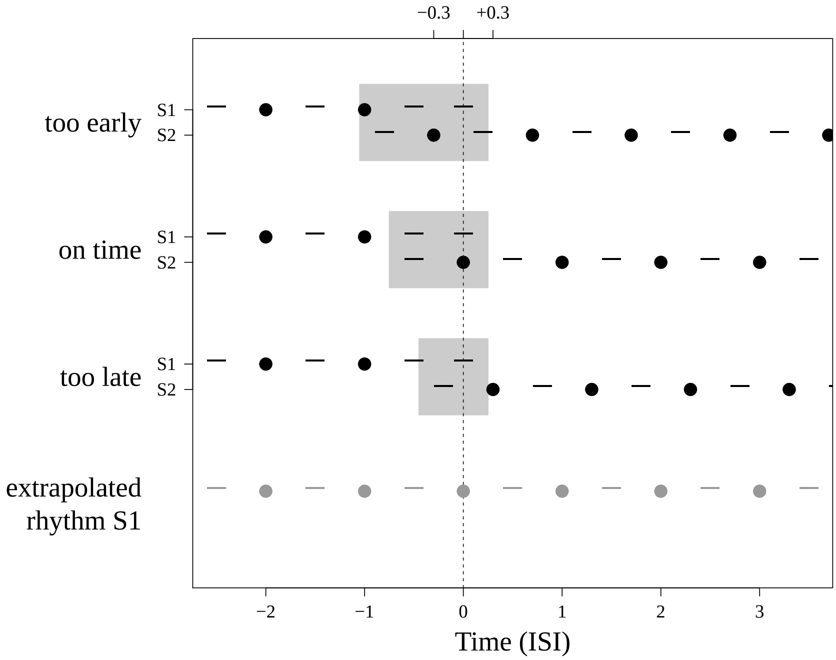

Figure 4.2. Schematic overview of our synchronization (and overlap) manipulation. This figure depicts the synchronization manipulation for exchanges containing an overlap, and no silent gap, between the two speakers (S1 and S2). Bullets represent stressed syllables, dashes unstressed ones, final lengthening is indicated by two dashes at the end of each turn, with the x-axis indicating normalized time in ISI (see text). Rectangular boxes represent the co-varying overlap duration. The vertical dashed line represents the relevant, extrapolated, beat of S1, with S2 coming in 'too early', 'on time', or 'too late'.

The resulting 216 unique exchanges each had 12 conditional variants: three synchronization conditions (on time, too early, too late), two overlap conditions (no overlap, overlap), and two tempo conditions (fast, slow). Thus, the three factors, synchronization, overlap, and tempo were fully crossed. Each exchange was then combined with one, instead with two as in Study 1, other exchange of the same conditional type and the same speakers to create a 'mini-dialogue'. This resulted in 72 mini-dialogues, consisting of two exchanges. Each conditional version of each mini-dialogue occurred in only one of 12 different experimental lists, with version, mini-dialogue, speaker voice, turn order, and intonational contour counterbalanced over lists. Additionally, each list was preceded by 4 practice items, consisting of items from a different list. The resulting mini-dialogues were low-pass filtered (450 Hz cutoff, slope 25 Hz/octave) in order to avoid lexical-semantic interference, while preserving sufficient ecological plausibility and creating the illusion of naturally filtered speech, as if listening to people in the room next door.

4.2.3 Procedure

The procedure was the same as in Study 1, with the exception that before the experiment, participants were prepared for the fEMG recording. Facial EMG data were collected using Ag/AgCl electrodes with a surface area of 2 mm over the right side of the face using the NeXus-10 MKII (Mind Media) bio-amplifier unit (2048 Hz sampling rate). Both the activity of the corrugator supercilii and zygomaticus major were measured. We included the zygomaticus ('smiling') muscle to help with disambiguating artefact recognition. However, the zygomaticus is primarily a marker of positive affect only and does not reflect emotional valence in the same way the corrugator does (Larsen et al., 2003). Positive affect decreases and negative affect increases corrugator activity, while zygomaticus activity only increases with positive affect, and does not show effects of negative affect (Larsen et al., 2003; Van Boxtel, 2010). Moreover, this suggests that a stimulus, which is ambivalent in affective valence, may result in increased zygomaticus activity, while corrugator activity remains unchanged under these conditions (Larsen et al., 2003). Given the results of Study 1 and 2, it is likely that participants will experience such ambivalence in affective valence while presented with our materials. We therefore focused only on corrugator activity as a measure of affective valence.

Participants were seated in a comfortable chair in a sound booth. As in Study 1 and 2, participants were instructed to focus on the second speaker in each exchange, and to look for evidence, possibly very subtle, whether the reaction of S2 contained cues that signaled affiliation. The mini-dialogues were presented over headphones at approximately 70 dB SPL. A trial consisted of the presentation of one entire mini-dialogue, containing two exchanges, with 1500 ms silence between exchanges, after which the participants were prompted for their response. A visual display suggested that the two exchanges within a mini-dialogue were fragments from the same conversation. Participants rated to what extent they heard signs of affiliation in the reaction of the second speaker across the two exchanges. As before, the exact wording (in Dutch) of the question after each mini-dialogue was: *In de reactie van de tweede spreker hoorde ik:* ('In the reaction of the second speaker I heard'), and below that, as anchors for the extremes of the 7-point scale: *toenadering* ('approach') and *verwijdering* ('withdrawal'). We chose to query affiliation in these terms because the approach-avoidance dimension is central to general theories of emotion (Frijda, 2009; Panksepp & Biven, 2012), and emerges in accounts of affiliation versus social distancing functions of social emotions (Fischer & Manstead, 2016). A schematic example of a mini-dialogue trial is depicted below in Figure 4.3.

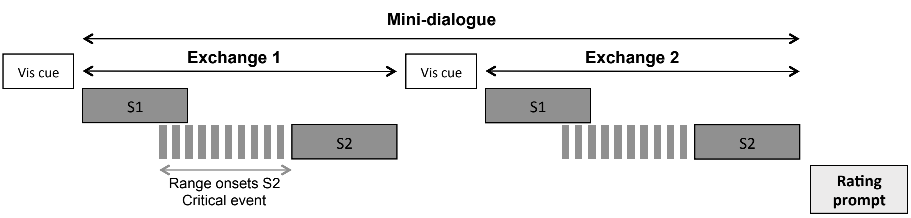

Figure 4.3. Illustration of a mini-dialogue trial. A visual cue helped participants imagine they were hearing two fragments sampled from a wider conversation; and were then presented with two exchanges between speaker 1 (S1) and speaker 2 (S2). Figure illustrates that the onset of S2 varied as a result of our critical manipulation of synchronization and of overlap.

Each participant was assigned one of the 12 experimental lists, and rated 72 mini-dialogues, divided in 8 blocks. As in Study 1, synchronization was presented in a mixed fashion, while block and overlap were presented in a blocked fashion. In other words, in any given block, the 9 mini-dialogues shared the same tempo and overlap condition, but pseudo-randomly differed in terms of synchronization in any given block. In this way, each participant completed two blocks of fast, overlapping mini-dialogues, two blocks of slow, overlapping mini-dialogues, two blocks of fast non-overlapping mini-dialogues, and two blocks with slow, non-overlapping mini-dialogues. Order of blocks was randomized over lists. In the pauses between blocks a photo of natural scenery was presented, after which the participants had to press a button to proceed. After the experiment, participants filled out an exit questionnaire about the experiment, their language and musical background, and containing three subscales (Empathic Concern, Personal Distress, and Perspective Taking) of the Dutch version of the Interpersonal Reactivity Index (Davis, 1983, translated by Ponnet, Roeyers & Buysse, 2002). As none of these exit questionnaire measures reliably explained variance in participant performance of facial EMG, these measures will be ignored here. After the exit questionnaire, participants were debriefed and payed.

4.2.4 Preprocessing

The raw facial EMG data were band-pass filtered between 20-500 Hz (48 dB/octave roll-off) and were additionally filtered with a notch filter at 50 Hz (see Van Boxtel, 2010 for justification of filter parameters), followed by signal rectification, using BrainVision Analyzer 2 (Brain Products GmbH). As previously pointed out, we also included a 2 s interval before each trial. For each mini-dialogue trial, this interval was inspected for remaining artefacts. We selected maximally long continuous pre-trial intervals epochs, with the requirement of a minimum of 500 ms of artefact-free signal for both corrugator and zygomaticus simultaneously. Trials were excluded if such a 500 ms pre-trial interval epoch could not be found. Exclusion led to a data loss of 4.2% with no asymmetry between conditions.

Following artefact rejection, data were exported to MatLab, and then segmented based on two critical segmentation points (CSPs): (1) time-locked at the first *beat* of S2 to assess effects of synchronization only, or (2) time-locked at the *onset* of S2 to assess effects of overlap, and, possibly, of tempo. See Table 4.1 for a summary of the two CSPs. For both CSPs, each post-CSP window was 2100 ms long. This should be long enough to detect rapid affective responses, but short enough not to be influenced by the following exchange or rating. The activity in the post-CSP window was expressed as a percentage of baseline-window activity[1]. Expressing responses as a percentage of a baseline helps to reduce random variance both within and between individuals; see Van Boxtel (2010). For each trial, data were averaged across the two exchanges per mini-dialogue.

Table 4.1

Summary of the two critical segmentation points (CSPs, in ms before the baseline-window or after the post-window)

Factor	CSP	Baseline window	Post window
1. Synchronization	1st beat S2	-900 to -400	0 to 2100
2. Overlap (& Tempo)	Onset S2	-500 to 0	0 to 2100

The two locations of the baseline-windows were somewhat different from each other. As can be seen from Table 4.1, the baseline window for synchronization (1) is placed slightly further away from its CSP compared to the baseline window for overlap (2). The second speaker always first utters a weak syllable, which is between 245 ms and 394 ms long, before uttering his or her first beat. Having the baseline-CSP window directly before the first beat would result in a baseline window containing at least some amount of corrugator activity in response to the onset of S2, which may result in a post-CSP activity that is contaminated with S2-onset effects. To avoid this possibility, we shifted the baseline window of synchronization 400 ms forwards given its CSP, so that it would solely include corrugator activity to S1, and not to S2.

4.2.5 Analysis

Post-window corrugator activity for both CSPs (section 4.3.1) as well as for the affiliation ratings (section 4.3.2) were analyzed by linear mixed-effects modeling (LMM) (Quené & Van den Bergh, 2008) using the R (R Core Team, 2015) package lmer4 (Bates et al., 2015), with participant, and, if the model fit increased, also with item (mini-dialogue) as

[1]The previously discussed 2 s interval interval was initially designed as a baseline and for artefact rejection of the facial EMG recordings. However, corrugator activity in response to S1 varied more than we expected. To cope with this unexpected within-trial variability, we selected a short interval directly before the turn of S2 as a new baseline: the two baseline windows.

random intercepts. Similar to Study 1, this model was then extended with different predictors in a particular order, as will be discussed below. After the addition of each predictor, and its interactions, the model fit was evaluated and compared to the previous model without that particular predictor (and interactions). Goodness of fit was explored using likelihood ratio tests (Pinheiro & Bates, 2000) and by inspecting the explained variance (R^2) of the extended model (Nakagawa & Schielzeth, 2013). On the basis of the evaluation of the model fit we decided which predictor (and interactions) to keep and which to discard.

The model fitting sequence was a follows. We first checked for individual differences, and later also for fixed effects, of tempo, then of overlap, and finally of synchronization. This particular sequence of tempo, overlap, and synchronization, was first used to define the random part of the model, before the fixed part was defined.

The optimal model resulting from this model fitting procedure was then simplified by removing possible redundant interactions. If the fit did not decrease by removing these interactions, we continued with the simpler model. Finally, the model was then evaluated by bootstrapped confidence intervals (1,000 replications).

Together, this conservative approach increased the likelihood that the final model would only contain genuine and robust effects. A more detailed discussion about the rationale behind this model fitting procedure can be found in Chapter 2.

4.3 Results

4.3.1 Corrugator activity

CSP 1: Synchronization-oriented analysis

For the model comparisons of CSP1, see Table 4.5 and Table 4.6 (at the end of this chapter).

Random Part We first defined the random part of our model by adding tempo as a random slope at the subject level (m0c). However, the model did not converge, suggesting a poor fit or a lack of power. Additionally, overlap was added as a random slope at the subject level (m0d). This improved the model fit, suggesting that participants may have different patterns of corrugator activity in response to our overlap manipulation. We then checked for individual differences of synchronization (m0e). This did improve the model fit as well, suggesting that participants also display differences in corrugator activity in response to our critical synchronization manipulation. However, visually inspecting the random slopes of model m0e shows that the individual differences are predominantly driven by a single participant. This specific participant displayed very large corrugator activity differences, while the other participants did not display clear individual variation in respect to our synchronization manipulation. Therefore, we decided to exclude this participant and to run the model comparisons again (see Table 4.6).

As before, tempo as a random slope (m0c) did not improve the model fit, while overlap (m0d) as a random slope did. When inspecting the random part of model m0d we found a perfect $[r = 1.00]$ correlation between the individual differences of the intercept and the effect of overlap. Very high correlations between intercept and slopes of random effects may indicate that the model has been overparameterized (Baayen et al., 2008). Moreover, adding synchronization (m0e) as a random slope led to convergence warnings, again suggesting that the model may be overparameterized. For these reasons, we decided to continue with a model without any random slopes, and with only participant as random intercepts.

Fixed Part For the definition of the fixed part, we added the same sequence of predictors as for the definition of the random part. Tempo (m1) did not improve the model fit. When overlap (m2) was added as a fixed predictor, the model fit improved. Crucially, we then checked for an effect of synchronization, and its interaction with overlap (m3). Against expectations, this did not improve the model fit. We found no effects of synchronization on corrugator activity. Finally, we also checked for effects of trial (m4), and of interactions with tempo (m5), but this did not improve the model fit either.

The model fitting procedure leads to model m3 as the optimal model for the purpose of the current analysis. Note that this is not the optimal model fitting our data (which is m2), but model m3 includes synchronization, which is the primary target of the current analysis. Tempo was not included in this model because we had no a priori theoretical reasons to expect effects of tempo on corrugator activity. A simpler model (m6) with only main effects of synchronization and overlap, and random intercepts of subject was an equally good fit of the data compared to model m3 [$\chi^2(2) = 0.43, p = 0.80$]. Therefore, we will present the results of this simpler model m6. The summary of this model m6 can be found in Table 4.2, and the fitted averages of model m6 are shown in Figure 4.4 and displayed in Table 4.9 (at the end of this chapter).

Table 4.2

Fitted estimates and bootstrappen 95% confidence intervals (C.I.) of corrugator activity in response to synchronization and overlap, based on model m6. Corrugator activity is expressed as a percentage of the activity in the baseline-window of that particular condition. The intercept refers to the on time, no overlap condition.

Fixed effects:	Estimate (S.E.)	95% CI	t value
Intercept	101.88 (0.56)	(100.74, 102.93)	180.42
Overlap	1.96 (0.43)	(1.14, 2.85)	4.52
Too Early	0.66 (0.53)	(-0.44, 1.76)	1.23
Too Late	-0.02 (0.53)	(-1.03, 1.03)	-0.03

Random effects	S.D.	95% CI	n
Subjects (intercept)	2.77	(3.79, 7.72)	59
Residual	19.58	(19.29, 19.91)	8126

Overlap In line with our predictions, we found a main effect of overlap [$\beta = 1.96,\ 95\%\ C.I.(1.14, 2.85), S.E. = 0.43, t = 4.52$]. However, and as previously pointed out, the current synchronization-oriented analysis is not optimal for assessing effects of overlap, hence we will not discuss the results of overlap in detail here (see section 4.3.1, for the overlap-oriented analysis).

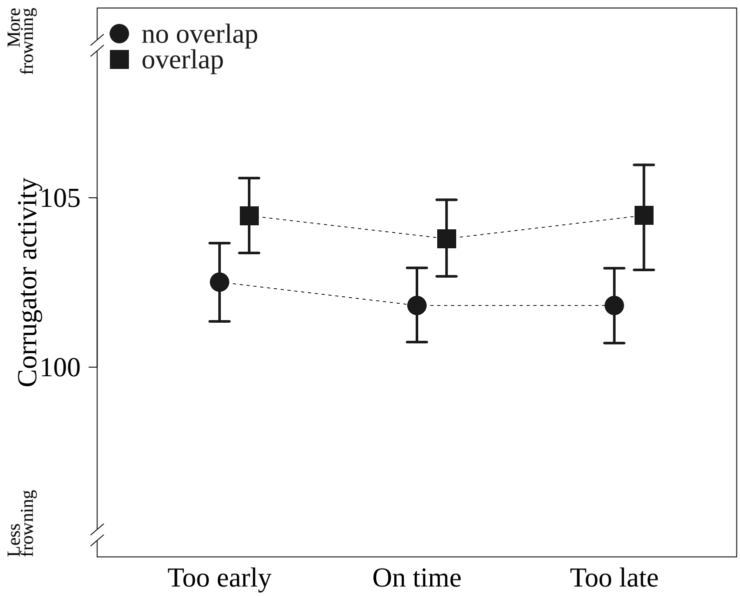

Figure 4.4. Bootstrapped (fitted) means and 95% confidence intervals (C.I.) of corrugator activity based on model m6, broken down by synchronization and overlap. The round symbols represent the bootstrapped fitted means for the non-overlapping conditions, square symbols the bootstrapped fitted means for the overlapping conditions. The bootstrapped 95% confidence intervals of each condition are represented by the vertical error bars, which may be slightly asymmetrical due to our bootstrapping outcomes. Corrugator activity is expressed as a percentage of the activity in the baseline-window of that particular condition.

Synchronization Confirming our goodness of fit exploration, but against our predictions, we found no effects of synchronization. Corrugator activity was not affected in response to speakers who were 'too early' [$\beta = 0.66, 95\%\ C.I.(-0.44,\ 1.76), S.E. = 0.53, t = 1.23$], nor in response to speakers who were 'too late' [$\beta = -0.02, 95\%\ C.I.(-1.03,\ 1.03), S.E. = 0.53, t = -0.03$].

In sum Thus, we found no effects of synchronization on corrugator activity, not as a main effect, nor as an effect at the individual level. Instead we found that only overlap affected participants' corrugator activity. The next analysis, which is more appropriately time-locked for assessing the effects of overlap, will investigate these effects in optimal fashion.

CSP 2: Overlap-oriented analysis

To investigate effects of overlap on corrugator activity, we again compared different model fits of our data, which was this time time-locked at the onset of S2 (CSP2). The different models are displayed in Table 4.7 (at the end of this chapter).

Random Part Tempo as a random slope at the subject level (m0c) did not improve the model fit. We then added overlap as a random slope (m01d). This model, however, failed to converge, suggesting a poor fit given our data.

Fixed Part We then turned to the fixed part of our model, where we added tempo (m1). As expected, adding tempo did not improve the model fit. Overlap was then added as a fixed predictor (m2), and, as predicted, this did improve the model fit. As before, we also checked for effects of trial (m3), and for effects of the interactions between our predictors tempo, overlap, and trial (m4). These more complex models were not a better fit of the data than our simpler model with only overlap and the random subject intercepts (m2). The summary of this best model (m2) is displayed in Table 4.3.

Overlap As expected, and in line with previous, synchronization-oriented results, we find an effect of overlap on corrugator activity; listening to speakers who overlap resulted in slightly (1%) more frowning activity (relative to the activity in the pre-window of that particular condition) $[\beta = 1.21, 95\%\ C.I.(0.30,\ 2.16), S.E. = 0.47, t = 2.59]$.

Table 4.3

Fitted estimates and bootstrapped 95% confidence intervals (C.I.), explaining corrugator activity in response to our overlap manipulation, based on model m2. The intercept refers to the no overlap condition. Corrugator activity is expressed as a percentage of the activity in the baseline-window of that particular condition.

Fixed effects:	Estimate (S.E.)	95% C.I.	t value
Intercept	102.58 (0.45)	(101.74, 103.42)	225.89
Overlap	1.21 (0.47)	(0.30, 2.16)	2.59

Random effects	S.D.	95% C.I.	n
Subjects (intercept)	2.41	(1.60, 3.05)	60
Residual	21.25	(20.89,21.59)	8270

4.3.2 Affiliation ratings

The different models explaining the affiliation ratings are displayed in Table 4.8 (at the end of this chapter).

Random Part Adding item (mini-dialogue) as a random intercept improved the model fit (m0b). To account for individual differences in tempo on evaluated affiliation, tempo was added as a random slope at the subject level (m0c). As expected, this did not improve the model fit. Overlap was then added as a random slope at the subject level (m0d), which improved the model fit. However, inspection of the random part of model m0d suggests that it may be spurious, as indicated by the perfect correlation between the intercept and the subject slope of overlap (Baayen et al., 2008), and by the bootstrapping confidence intervals of this correlation $[r = 1.00, 95\%\ C.I.(-1.00, 1.00)]$. Eliminating item as a random intercept gave the same spurious results. Therefore, we decided to discard overlap as a random slope at the subject level.

We then checked for individual differences in response to our synchronization manipulation on evaluated affiliation (m0e). Adding synchronization as a random slope at the subject level did not improve the model fit, indicating that our participants did not display significant individual differences in evaluated affiliation in response to our synchronization manipulation.

Fixed Part To define the fixed part of our model, tempo (m1) was then added as a fixed predictor. Even though we did find effects of tempo in our previous studies, the addition of tempo did not improve the model fit here. We then added overlap (m2) as a fixed predictor, again improving the model fit, indicating that the presence or absence of overlap did indeed influence the affiliation ratings. We then checked for effects of synchronization (m3), which did improve the model fit as well. More complex models, with overlap, synchronization, and trial (m4), with tempo (m5), or with all experimental factors, trial, and their interactions (m6) were not better than the simpler model m3. Because containing our main factors of interest, i.e., tempo, overlap, and synchronization, model m5 is the optimal model for the current analysis.

A simpler model (m7) with only main effects of synchronization, overlap and tempo, and random intercepts of subjects and items proved to be an equally good fit of the data compared to model m5 $[\chi^2(7) = 4.51, p = 0.72]$. Here, we will present the results of this simpler model m7. The summary of model m7 can be found in Table 4.4, the fitted means of this model are displayed in Table 4.10 (at the end of this chapter) and the fitted averages of model m7 are shown in Figure 4.5.

Tempo We did not replicate the previous effect of speech tempo on evaluated affiliation. Slow speech did not increase evaluated affiliation $[\beta = -0.02, 95\%\ C.I.(-0.11,\ -0.07), S.E. = 0.05, t = -0.34]$.

Overlap As expected, and in line with the corrugator results of the overlap-oriented-analysis (CSP 2), we found a main effect of overlap on evaluated affiliation. Rated affiliation was lower when speakers produced overlap $[\beta = -0.23, 95\%\ C.I.(-0.33,\ -0.14), S.E. = 0.05, t = -4.79]$.

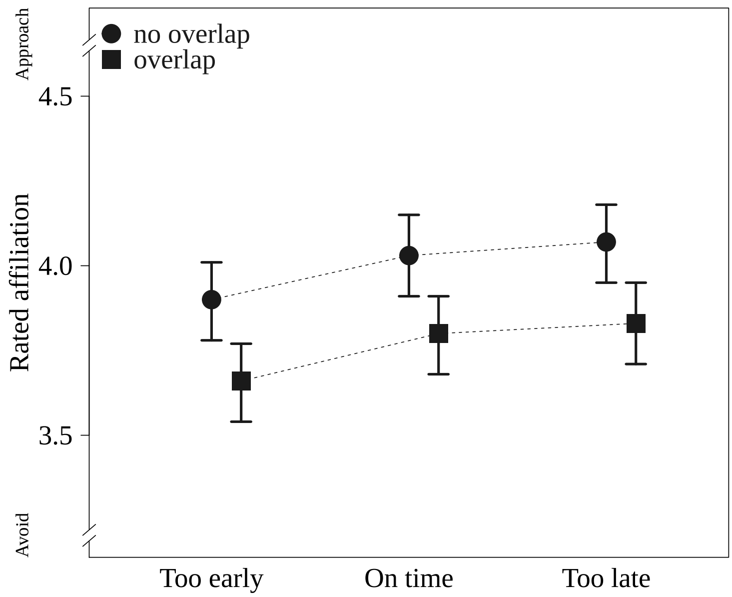

Figure 4.5. Bootstrapped (fitted) means and 95% confidence intervals of evaluated affiliation, based on model m7, broken down by synchronization and overlap. The round symbols represent the bootstrapped fitted means for the non-overlapping conditions, the square symbols the bootstrapped fitted means for the overlapping conditions. The bootstrapped 95% confidence intervals of each condition are represented by the vertical error bars, which may be slightly asymmetrical due to the bootstrapping outcomes.

Synchronization Critically, we also found an effect of synchronization on evaluated affiliation, independent of overlap and tempo. Affiliation ratings were lower when speakers responded 'too early' $[\beta = -0.14, 95\%\ C.I.(-0.26,\ -0.02), S.E. = 0.06, t = -2.32]$. However, evaluated affiliation did not lower in response to speakers being 'too late' $[\beta = 0.03, 95\%\ C.I.(-0.08,\ 0.15), S.E. = 0.06, t = 0.58]$.

Table 4.4

Fitted estimates and bootstrapped 95% confidence intervals (C.I.) of evaluated affiliation, explaining affiliation ratings in response to our overlap, synchronization and tempo manipulation, based on model m7. The intercept refers to the on time, no overlap, fast condition. Evaluated affiliation ranges from 1 (avoidance) to 7 (approach).

Fixed effects:	Estimate (S.E.)	95% C.I.	t-value
Intercept	4.04 (0.06)	(3.92, 4.16)	67.46
Slow	-0.02 (0.05)	(-0.11, 0.07)	-0.34
Overlap	-0.23 (0.05)	(-0.33,-0.14)	-4.79
Too early	-0.14 (0.06)	(-0.26,-0.02)	-2.33
Too late	0.03 (0.06)	(-0.08, 0.15)	0.58

Random effects	S.D.	95% C.I.	n
Subjects (intercept)	0.17	(0.08, 0.23)	60
Items	0.10	(0.00, 0.17)	72
Residual	1.60	(1.56, 1.63)	4320

Summary of the results

In line with our predictions, we found clear effects of overlap; when listening to speakers who overlapped with the turn of the preceding speakers, we found that corrugator activity increased and that affiliation ratings were lowered, compared to non-overlapping speakers.

Against our predictions, we did not find the expected effect of synchronization on corrugator activity; our participants did not frown more (or less) when listening to speakers who did or who did not synchronize to each others speech rhythm. However, for the affiliation ratings, we did find the expected effect of synchronization. As in Study 1, responding ‘too early’ led to significantly lower ratings of affiliation, but no effect on evaluated affiliation was found for speakers responding ‘too late’.

As expected, corrugator activity was not affected by tempo. The effects of tempo on evaluated affiliation found in our previous studies were not replicated; slow and fast speech were rated as equally affiliative.

4.4 Discussion

In this chapter we sought to unravel the affective impact of speech rhythm synchronization between speakers in conversation. In the two previous studies (Chapter 2 and 3) we found conflicting evidence for the hypothesis under investigation. These previous two studies exclusively used post-hoc ratings to assess affiliation. These ratings may not be sensitive enough to detect the subtle effects of synchronization. In addition to these affiliation ratings, the current study therefore investigated the affective impact of our synchronization hypothesis with a different measure, namely by recording participants' corrugator, or 'frowning', activity. Corrugator activity is a reliable and fast indicator of the affective valence of a given stimulus (Larsen et al., 2003; Van Boxtel, 2010). Measuring corrugator activity during the experiment allowed us to track the instantaneous, affective response of our participants to our critical manipulations of synchronization, tempo, and overlap, which may increase our understanding of how the impression of affiliation may come about, since we assumed that these impressions are primarily driven by the affective system of the listener.

4.4.1 The presence of overlap

We predicted that listening to overlap between speakers would lead to *lower* ratings of affiliation and that listening to these speakers would *increase* corrugator activity, compared to listening to non-overlapping speakers. Both predictions were corroborated: overhearing overlapping talk between speakers lowered perceived speaker affiliation and increased corrugator activity.

First of all, these findings indicate that both our measures, i.e. affiliation ratings as well as corrugator activity, were sensitive enough to pick up our temporal manipulations, or at least sensitive enough to pick up the presence of an overlap in conversation, even if this conversation is not experienced at first hand. These results thus serve as a sanity check for our paradigm.

Affiliation ratings

These findings of overlap on evaluated affiliation closely resemble the finding of overlap found in Study 1. Moreover, our findings corroborate the research about overlap-induced speaker attributions, in which overlaps are generally perceived as rude and disrespectful (Beňuš et al., 2011; Goldberg, 1990; Maat et al., 2010). This indicates that the presence of overlap in conversation can indeed covey disaffiliation between speakers in conversation.

Corrugator activity

In addition to the *impression* of disaffiliation, the presence of overlap in conversation seems to trigger negative emotions as well, as indicated by the increased corrugator activity of our participants when listening to overlapping compared to non-overlapping speakers. In other words, participants seem to really 'feel' the presence of an overlap when listening to overlapping speakers.

However, the results of the corrugator muscle could also be attributable to the fact that listening to overlapping talk may be more challenging than listening to non-overlapping talk. Detecting pitch tones in auditorily challenging environments increases corrugator activity, compared to the same task under less challenging conditions (Cohen, Davidson, Senulis, Saron, & Weisman, 1992). In other words, increased corrugator activity may reflect mental effort too, besides being an indicator of the affective valence of the presence of an overlap. Nevertheless, overlap also affected the *impression* of interpersonal affect, as indicated by the results of our affiliation ratings.

If the results of the corrugator should indeed be explained in terms of mental effort, then this explanation would entail, contrary to our predictions, that the processes guiding the recorded affective response (corrugator activity) and the impression of affiliation (ratings) should be be disconnected. If this were the case, then participants somehow 'knew' instead of 'felt' that overlaps should be interpreted as a sign of lack of affiliation, without being affectively moved. The results of the corrugator may then be explained in terms of mental effort. This option does not seem parsimonious given that it implies the separation of two processes which we assumed to be inherently connected. We assumed that our temporal manipulations would trigger the affective system of the listener, giving rise to emotions, either experienced consciously or not, leading to the impression of affiliation. In addition, the separation of these two processes entails that our listeners would somehow rely on learned patterns to assess affiliation, while such patterns can only be learned through *affective* processes, since they involve the affective evaluation of human interactions, either experienced indirectly or at first hand (Abdai & Miklóosi, 2016; Hamlin, 2015).

The more parsimonious explanation would be that overlap strongly conveys interpersonal negativity, impacting both the rated *impression* of affiliation as well as the recorded affective response of our overhearing listeners. This idea is reinforced by the fact that the results of overlap on the impression of affiliation are in line with earlier research on the social implications of overlap (Beňuš et al., 2011; Goldberg, 1990; Maat et al., 2010).

In any case, as reflected in the affiliation ratings, the presence of overlap in conversation appears to robustly convey interpersonal negativity, impacting the *impression* of affiliation, and possibly leading to negative affect as well, even for overhearing listeners.

4.4.2 Synchronization

The second prediction was that listening to speakers who do *not* synchronize their speech rhythm to that of the previous speaker will be rated as *less* affiliative and will *increase* participants corrugator activity, compared to listening to speakers who do synchronize their speech rhythm to that of the previous speaker.

Affiliation ratings

As expected, perceived interpersonal affect was influenced by speech rhythm synchronization between speakers. Irrespective of the presence or absence of overlapping talk and independent of speech tempo, speakers who responded ‘too early’ were evaluated as *less* affiliative than speakers who responded ‘on time’. As before, however, perceived affiliation between speakers who responded ‘on time’ and ‘too late’ did not differ.

These results were very similar to the results for non-overlapping speakers of Study 1. There we hypothesized that this asymmetry between being ‘too early’ and ‘too late’ might be attributable to the fact that being ‘too late’ might not be as socially disruptive as being ‘too early’. As pointed out in Study 1, being ‘too late’ may also be interpreted, by the listener, as a cue for other speaker variables, such as problems with comprehension (Beňuš et al., 2011), planning difficulties (Bull & Aylett, 1998), fluency (Bosker, 2014), or word frequency (Hartsuiker & Notebaert, 2010). It may be that participants took these possibilities into account when assessing affiliation between speakers, thereby attenuating the synchronization effect of ‘too late’.

In contrast to Study 1, no interaction between synchronization and overlap was found, suggesting that the assumed effect of the amount of overlap might not be as robust as initially hypothesized.

Corrugator activity

Unexpectedly, synchronization did not influence the activity of the corrugator muscle of our participants. Given the findings on the ratings of affiliation, one would expect participants to frown more when listening to speakers responding ‘too early’ than when listening to speakers responding ‘on time’. Two accounts may explain the lack of agreement between the affiliation ratings and corrugator activity of our participants.

Under the first account, synchronization functions as a ‘cold’ cue. It might be that participants did not affectively respond to either of the speakers, instead they use synchronization ‘coldly’, somehow basing their rating of affiliation on learned patterns, i.e., acquired correlations between particular interpersonal timing relations and particular degrees of affiliation between people, instead of basing their judgment on their own feelings or emotions. Even though this account may seem implausible, the data compel us to at least consider it as an option.

Another explanation for the lack of agreement between the affiliation ratings and corrugator activity is that here the low power of our study may have led to a false negative.

The presence or absence of overlap lowered evaluated affiliation by 7% (Table 5.3), while the presence of overlap only increased corrugator activity by 1% (Table 4.3, Chapter 4). The impact of being 'too early' on evaluated affiliation was even smaller (compared to the effect of overlap); responding 'too early' lowered evaluated affiliation with only 3% (compared to 7% for overlap) in the joint analysis, see Table 5.3. If corrugator activity in response to being 'too early' would increase to the same extent as it does for responding with an overlap, then corrugator activity will only increase with $(3 * (1/7) =)$ 0.4%. The impact of synchronization on corrugator activity may thus have been too limited to result in a substantial increase in corrugator activity, suggesting that corrugator activity may be less sensitive to our temporal manipulations than our rating of affiliation, even though we hypothesized the opposite.

A possible explanation for the small impact of synchronization in general, i.e., on both corrugator activity as well as evaluated affiliation, may be the fact we took our overhearing participant as a proxy for the interlocutor in conversation. It might be that taking the overhearer as a proxy for the interlocutor is just one step too far in the case of the already subtle effect of synchronization. Since not in conversation themselves, the subtle cue of synchronization between speakers might not be of sufficient emotional salience or relevance to lead to a substantial increase in participants' corrugator activity.

Overlap may be of greater social significance than synchronization, since it involves breaking into someone's talk. This may give rise to the impression that the overlapper does not respect the other person, or his or her personal space. Synchronization, on the other hand, 'only' involves the non-alignment of speech rhythms and may as such be less socially disruptive than breaking into someone's turn. The difference between overlap and being 'too early' may also be explained in terms of acoustic complexity. While being too early relies on the *interplay* of two acoustic cues, i.e., the interacting rhythms of the turn-taking speakers, overlap relies solely on the absence or presence of simultaneous speech. This suggests that overlap may constitute a more salient cue, both socially as well as acoustically, than synchronization, thereby influencing corrugator activity to a greater extent than synchronization does.

Taken together, the absence of an effect of being ‘too early’ on corrugator activity may have been the result of an accumulation of factors. The fact that the corrugator muscle seemed to be less sensitive to our temporal manipulations than the affiliation ratings, the use of overhearing listeners instead of participating interlocutors, and the subtle social or acoustic nature of synchronization may all have contributed to to this null result. Still, the subtle between speaker cues of being ‘too early’, gave rise to, consciously or not, impressions of disaffiliation. This indicates that, even though listeners seem insensitive to synchronization as an *affective* cue, synchronization, as in the case of being ‘too early’, somehow does impact the *affective impression* of the push-and-pull in the dynamics of interpersonal communication.

4.4.3 Tempo

The tempo effects on evaluated affiliation of Study 1 and 2 were not replicated; slow speech was rated as equally affiliative as fast speech. Moreover, and in line with our predictions, corrugator activity was not affected by speech tempo. Together, this indicates that the effect of tempo on evaluated affiliation is not very stable across studies, suggesting that we should interpret these tempo effect with caution. The next chapter (Chapter 5) will continue the discussion about tempo.

4.4.4 Summary

Although there are still many uncertainties considering the affective impact of synchronization in conversation, we do know that the presence of overlap, at least in our paradigm, robustly conveys negativity in the affective dynamics of interpersonal communication; participants frown more when they hear speakers overlap, and deem those speakers as less affiliative than when those speakers do not overlap. Synchronization, as operationalised in our studies, appears to be a more subtle and volatile phenomenon only affecting the impression of affiliation under restricted circumstances. In the next chapter we will reconsider our synchronization hypothesis in more detail, and elaborate on its nature and existence in conversation, as well as its hypothesized affective impact, on interlocutors and overhearers alike.

Table 4.5

Models explaining corrugator activity segmented on the first beat of S2 in order to assess effects of synchronization.

Name	Model	R^2_{marg}	R^2_{cond}	Comparison	-2LogLik	Chi-squared
m0a	$1 + (1\|subj)$	0	0.016		-37272	
m0b	$1 + (1\|subj) + (1\|item)$	0	0.017	m0a vs m0b	-37272	$\chi^2(1) = 0.11, p = 0.74$
m0c	$1 + (1 + temp\|subj)$	0		m0a vs m0c		Failed to converge
m0d	$1 + (1 + ovl\|subj)$	0	0.020	m0a vs m0d	-37268	$\chi^2(2) = 8.21, p = 0.02$
m0d	$1 + (1 + sync + ovl\|subj)$	0	0.036	m0d vs m0e	-37236	$\chi^2(7) = 63.20, p < 0.001$

Note: see Table 2.2 (Chapter 2) and Bates et al. (2015) regarding the format of the models. R^2_{marg} and R^2_{cond} are calculated based on Nakagawa and Schielzeth (2013).

Table 4.6

Models explaining corrugator activity segmented on the first beat of S2 in order to assess effects of synchronization with one participant excluded.

Name	Model	R^2_{marg}	R^2_{cond}	Comparison	-2LogLik	Chi-squared
m0a	$1 + (1\vert subj)$	0	0.019		-35752	
m0b	$1 + (1\vert subj) + (1\vert item)$	0	0.021	m0a vs m0b	-35752	$\chi^2(1) = 0.49, p = 0.48$
m0c	$1 + (1 + temp\vert subj)$	0	0.019	m0a vs m0c	-35752	$\chi^2(2) = 0.01, p = 0.99$
m0d	$1 + (1 + ovl\vert subj)$	0	0.025	m0a vs m0d	-35742	$\chi^2(2) = 19.34, p<0.001$
m0d	$1 + (1 + sync + ovl\vert subj)$	0				Failed to converge
m1	$1 + temp + (1\vert subj)$	0.000	0.020	m0a vs m1	-35752	$\chi^2(1) = 0.50, p = 0.48$
m2	$1 + ovl + (1\vert subj)$	0.002	0.022	m0a vs m2	-35742	$\chi^2(1) = 20.39, p = <0.001$
m3	$1 + sync * ovl + (1\vert subj)$	0.003	0.022	m2 vs m3	-35740	$\chi^2(4) = 3.10, p = 0.54$
m4	$1 + trial * sync * ovl + (1\vert subj)$	0.004	0.023	m2 vs m4	-35736	$\chi^2(10) = 10.81, p = 0.37$
m5	$1 + temp * trial * sync * ovl + (1\vert subj)$	0.006	0.028	m2 vs m5	-35728	$\chi^2(22) = 27.02, p = 0.21$
m6	$1 + sync + ovl + (1\vert subj)$	0.003	0.022	m2 vs m6	-35740	$\chi^2(2) = 1.01, p = 0.60$

Note: see Table 2.2 (Chapter 2) and Bates et al. (2015) regarding the format of the models. R^2_{marg} and R^2_{cond} are calculated based on Nakagawa and Schielzeth (2013).

Table 4.7

Models explaining corrugator activity segmented on the onset of S2 in order to assess effects of overlap.

Name	Model	R^2_{marg}	R^2_{cond}	Comparison	-2LogLik	Chi-squared
m0a	$1 + (1\|subj)$	0	0.013		-37045	
m0b	$1 + (1\|subj) + (1\|item)$	0	0.013	m0a vs m0b	-37045	$\chi^2(1) = 0.00, p = 1.00$
m0c	$1 + (1 + temp\|subj)$	0	0.013	m0a vs m0c	-37044	$\chi^2(1) = 0.99, p = 0.61$
m0d	$1 + (1 + ovl\|subj)$					Failed to converge
m1	$1 + temp + (1\|subj)$	0.000	0.013	m0a vs m1	-37045	$\chi^2(1) = 0.10, p = 0.75$
m2	$1 + ovl + (1\|subj)$	0.001	0.013	m0a vs m2	-37042	$\chi^2(1) = 6.72, p = 0.01$
m3	$1 + trial * ovl + (1\|subj)$	0.001	0.014	m2 vs m3	-37040	$\chi^2(2) = 2.56, p = 0.28$
m4	$1 + temp * trial * ovl + (1\|subj)$	0.001	0.014	m2 vs m4	-37039	$\chi^2(6) = 4.68, p = 0.59$

Note: see Table 2.2 (Chapter 2) and Bates et al. (2015) regarding the format of the models. R^2_{marg} and R^2_{cond} are calculated based on Nakagawa and Schielzeth (2013).

Table 4.8

Models explaining the affiliation ratings of Study 3

Name	Model	R^2_{marg}	R^2_{cond}	Comparison	-2LogLik	Chi-squared
m0a	$1 + (1\|subj)$	0	0.011		-8201	
m0b	$1 + (1\|subj) + (1\|item)$	0	0.015	m0a vs m0b	-8200	$\chi^2(1) = 1.76p = 0.18$
m0c	$1 + (1 + temp\|subj) + (1\|item)$	0	0.021	m0b vs m0c	-8197	$\chi^2(2) = 5.37, p = 0.07$
m0d	$1 + (1 + ovl\|subj) + (1\|item)$	0	0.021	m0b vs m0d	-8196	$\chi^2(2) = 7.46, p = 0.02$
m0e	$1 + (1 + sync\|subj) + (1\|item)$	0	0.017	m0b vs m0e	-8198	$\chi^2(5) = 3.52, p = 0.62$
m1	$1 + temp + (1\|subj) + (1\|item)$	0.000	0.015	m0b vs m1	-8200	$\chi^2(1) = 0.12, p = 0.73$
m2	$1 + ovl + (1\|subj) + (1\|item)$	0.005	0.020	m0b vs m2	-8189	$\chi^2(1) = 22.87, p<0.001$
m3	$1 + sync * ovl + (1\|subj) + (1\|item)$	0.008	0.023	m2 vs m3	-8182	$\chi^2(4) = 12.31, p = 0.02$
m4	$1 + trial * sync * ovl + (1\|subj) + (1\|item)$	0.009	0.023	m3 vs m4	-8181	$\chi^2(6) = 3.58, p = 0.73$
m5	$1 + temp * sync * ovl + (1\|subj) + (1\|item)$	0.008	0.024	m3 vs m5	-8182	$\chi^2(6) = 1.75, p = 0.94$
m6	$1 + trial * temp * sync * ovl + (1\|subj) + (1\|item)$	0.012	0.026	m3 vs m6	-8175	$\chi^2(18) = 15.81, p = 0.61$
m7	$1 + temp + sync + ovl + (1\|subj) + (1\|item)$	0.007	0.022	m5 vs m7	-8182	$\chi^2(7) = 4.51, p = 0.72$

Note: see Table 2.2 (Chapter 2) and Bates et al. (2015) regarding the format of the models. R^2_{marg} and R^2_{cond} are calculated based on Nakagawa and Schielzeth (2013).

Table 4.9

Bootstrapped (fitted) means and 95% confidence intervals (C.I.) of corrugator activity based on model m6, broken down by synchronization and overlap. Corrugator activity is expressed as a percentage of the activity in the baseline-window of that particular condition.

	No overlap		Overlap	
	Mean	95% C.I.	Mean	95% C.I.
Too early	102.51	(101.35, 103.66)	104.47	(103.37, 105.58)
On time	101.82	(100.74, 102.93)	103.79	(102.68, 104.94)
Too late	101.82	(100.71, 102.92)	104.48	(102.87, 105.97)

Table 4.10

Bootstrapped (fitted) means and 95% confidence intervals (C.I.) of the affiliation ratings based on model m7. Evaluated affiliation ranges from 1 (avoidance) to 7 (approach).

	No overlap		Overlap	
	Mean	95% C.I.	Mean	95% C.I.
Too early	3.90	(3.78, 4.01)	3.66	(3.54, 3.77)
On time	4.03	(3.91, 4.15)	3.80	(3.68, 3.91)
Too late	4.07	(3.95, 4.18)	3.83	(3.71, 3.95)

CHAPTER 5

Discussion and conclusion

5.1 Overview

In this dissertation we sought to unravel whether certain temporal patterns between turn-taking speakers convey interpersonal affect between those speakers, on top of the established effects of silent gap and overlapping talk between them. Particularly, we were interested in the social impact of two specific types of temporal patterns between speakers, namely speech rhythm (de)synchronization (Chapter 2, 3, and 4) and tempo (mis)matching (Chapter 3). Speech rhythm *synchronization* occurs when a speaker adjusts his or her speech rhythm to the (extrapolated) speech rhythm of the previous speaker. *Tempo matching*, on the other hand, occurs when a speaker adjusts his or her speech tempo to match the tempo of the previous speaker.

To explore whether (de)synchronization and/or tempo (mis)-matching between speakers in conversation may indeed convey (a lack of) affiliation between those speakers, we presented listeners with short, synthesized fragments of dialogues after which participants were asked to rate the perceived degree of affiliation between the speakers in dialogue. Additionally, in Study 3 we also recorded participants' instantaneous affective response, by measuring their corrugator activity while they listened to these dialogues.

The dialogues between speakers consisted of short fragments between two speakers in which we manipulated the temporal patterns between, as well as within speaking turns.

- (Study 1, 2 and 3) Speakers either synchronized or did not synchronize their speech rhythms;
- (Study 2) Speakers either matched or mismatched their speech tempos;
- (Study 1 and 3) Speakers either overlapped or did not overlap with the previous speaker;
- (Study 1, 2 and 3) And, as a control factor, speakers either spoke slowly or fast.

The core hypothesis was that, on top of the established effects of overlapping talk and silent gap, synchronization and tempo matching would impact perceived affiliation (ratings) as well as the affective system (corrugator activity) of our participants. Specifically, we hypothesized the following:

Affiliation ratings

H1a Speakers who produce overlapping talk will be rated as *less* affiliative than speakers who do not overlap.

H2 Independent of tempo, speakers who do *not* match their speech tempo to that of the previous speaker will be rated as *less* affiliative than speakers who do match their speech tempo to that of the previous speaker.

H3a Independent of overlap and tempo, speakers who do *not* synchronize their speech rhythm to that of the previous speaker will be rated as *less* affiliative than speakers who do synchronize their speech rhythm to that of the previous speaker.

Corrugator activity

H1b Listening to speakers who produce overlapping talk will elicit *more* corrugator activity than listening speakers who do not overlap.

H3b Independent of overlap and tempo, listening to speakers who do *not* synchronize their speech rhythm to that of the previous speaker will elicit *more* corrugator activity than listening to speakers who do synchronize their speech rhythm to that of the previous speaker.

Table 5.1 presents an overview of the findings of each of the three experimental studies.

Table 5.1

Overview of the effects on evaluated affiliation per experimental factor and per study.

Experimental factor	Ratings Study 1		Ratings Study 2		Ratings Study 3	
1. Overlap	Overlap	*	-	-	Overlap	*
2. Synchronization	Too early, n.o.	*	Too early, n.o.	n.s.	Too early, n.o.	*
	Too early, ov.	*	-	-	Too early, ov.	*
	Too late, n.o.	n.s.	Too-late, n.o.	n.s.	Too late, n.o.	n.s.
	Too late, ov.	?*	-	-	Too late, ov.	n.s.
3. Tempo matching	-	-	Tempo matching	n.s.	-	-
4. Tempo	Slow tempo	?*	Slow tempo	?*	Slow tempo	n.s.
5. Tempo interactions	Interactions	n.s.	Interactions	n.s.	Interactions	n.s.

*Note:** significant, expected; ?* significant, unexpected; n.s. non significant; - not assessed; n.o. no overlap condition; ov. overlap condition.

As can be seen from Table 5.1, the effects were not always consistent across studies. Looking at Study 1, we found that being ‘too late’ for overlapping speakers (Too late, ov.) unexpectedly resulted in *higher* ratings of affiliation (as indicated with ?*). In Study 3 however, the effect of being ‘too late’ in the same context was not significantly different from being ‘on time’. Moreover, in Study 1 and 2 tempo unexpectedly influenced evaluated affiliation, while Study 3 did not reveal tempo effects.

Combined with the fact that all our effects were small[1], these inconsistencies across studies suggests that our studies may suffer from low statistical power. Besides reducing the likelihood of detecting a true effect, low power also reduces the likelihood that a statistically significant result reflects a true effect (Ioannidis, 2005). Low power may have led to spurious effects in our data, obscuring the real effects of our experimental factors on perceived affiliation. Power could be increased by combining the results of Study 1 and 3[2], thereby increasing the chances that the observed effects are non-spurious. Such a joint analysis may help us to decide which of the effects should be interpreted with caution, and which of the effect are robust, at least under the comparable circumstances of Study 1 and 3.

Thus, before we turn to the interpretation of the findings of our studies (section 5.3), we first present a joint analysis of Study 1 and 3 in order to gain a better understanding of the real effects of synchronization and tempo on rated affiliation. The results of this joint analysis will be discussed in the next section (5.2).

5.2 Joint analysis

For the analysis of the results of the joint analysis of Study 1 and 3 we followed the same optimal model procedure as before (cf. Study 1 and 3, see Table 5.2). The optimal model m8 of the joint analysis consists of the random intercepts of item (mini-dialogue), random slopes at the subject level of overlap, without the random intercepts of subject, and with the fixed factors tempo, overlap, and synchronization. The estimates of model m8 are presented in Table 5.3. The fitted averages are displayed in Table 5.4 and visualized in Figure 5.1.

[1] The estimated effects are not presented in this chapter, but see Table 2.1 (Study 1), Table 3.1 (Study 2), and Table 4.4 (Study 3).

[2] The results of Study 2 cannot be included in this joint analysis because the experimental design of Study 2 was different to that of Study 1 and 3.

Table 5.2

Models explaining the affiliation ratings of the joint analysis of Study 1 and 3.

Name	Model	R^2_{marg}	R^2_{cond}	Comparison	-2LogLik	Chi-squared
m0a	$1 + (1 \mid subj)$	0	0.015		-16114	
m0b	$1 + (1 \mid subj) + (1 \mid item)$	0	0.020	m0a vs m0b	-16108	$\chi^2(1) = 11.48, p < 0.001$
m0c	$1 + (1 + temp \mid subj) + (1 \mid item)$	0	0.024	m0b vs m0c	-16105	$\chi^2(2) = 4.74, p = 0.09$
m0d	$1 + (1 + ovl \mid subj) + (1 \mid item)$	0	0.047	m0b vs m0d	-16051	$\chi^2(2) = 112.89, p < 0.001$
m0e	$1 + (1 + sync + ovl \mid subj) + (1 \mid item)$					Failed to converge
m0f	$1 + (1 + sync + ovl \mid subj)$	0	0.045	m0d vs m0f	-16050	$\chi^2(6) = 1.58, p = 0.95$
m0g	$1 + (1 + ovl \mid subj) + (1 \mid item)$	0	0.047	m0b vs m0g	-16051	$\chi^2(2) = 112.89, p < 0.001$
m0h	$1 + (0 + ovl \mid subj) + (1 \mid item)$	0	0.047	m0b vs m0h	-16051	$\chi^2(2) = 112.89, p < 0.001$
m1	$1 + temp + (0 + ovl \mid subj) + (1 \mid item)$	0.000	0.047	m0d vs m1	-16050	$\chi^2(1) = 1.95, p = 0.16$
m2	$1 + ovl + (0 + ovl \mid subj) + (1 \mid item)$	0.009	0.046	m0d vs m2	-16036	$\chi^2(1) = 30.53, p < 0.001$
m3	$1 + sync * ovl + (0 + ovl \mid subj) + (1 \mid item)$	0.012	0.050	m2 vs m3	-16022	$\chi^2(4) = 28.01, p < 0.001$
m4	$1 + tempo * sync * ovl + (0 + ovl \mid subj) + (1 \mid item)$	0.013	0.050	m3 vs m4	-16018	$\chi^2(6) = 7.86, p = 0.25$
m5	$1 + exp * sync * ovl + (0 + ovl \mid subj) + (1 \mid item)$	0.013	0.050	m3 vs m5	-16020	$\chi^2(6) = 4.06, p = 0.67$
m6	$1 + trial * sync * ovl + (0 + ovl \mid subj) + (1 \mid item)$	0.012	0.051	m3 vs m6	-16020	$\chi^2(6) = 5.09, p = 0.53$
m7	$1 + tempo * exp * trial * sync * ovl + (0 + ovl \mid subj) + (1 \mid item)$	0.018	0.049	m3 vs m7	-16003	$\chi^2(42) = 37.67, p = 0.66$
m8	$1 + tempo + sync + ovl + (0 + ovl \mid subj) + (1 \mid item)$	0.011	0.049	m3 vs m8	-16018	$\chi^2(7) = 13.09, p = 0.07$

Note: see Table 2.2 (Chapter 2) and Bates et al. (2015) regarding the format of the models. R^2_{marg} and R^2_{cond} are calculated based on Nakagawa and Schielzeth (2013).

Table 5.3

Fitted estimates and bootstrapped 95% confidence intervals of evaluated affiliation of the joint analysis of Study 1 and 3 based on model m8. The intercept refers to the on time, no overlap, fast condition. Evaluated affiliation ranges from 1 (avoidance) to 7 (approach).

Fixed effects	Estimate (S.E.)	95% C.I.	t-value
Intercept	4.04 (0.04)	(3.95, 4.12)	94.98
Slow	0.05 (0.03)	(-0.02, 0.11)	1.38
Overlap	-0.29 (0.05)	(-0.39,-0.19)	-5.90
Too early	-0.14 (0.04)	(-0.22,-0.07)	-3.56
Too late	0.03 (0.04)	(-0.05, 0.10)	0.71

Random effects	Std.Dev.	95% C.I.	Corr.	95 %C.I.	n
No overlap (subj.)	0.18	(0.09, 0.24)			120
Overlap (subj.)	0.36	(0.28, 0.42)	-0.06	(-0.51, 0.37)	
Items	0.12	(0.05, 0.16)			72
Residual	1.53	(1.50, 1.55)			8640

Table 5.4

Bootstrapped (fitted) means and 95% confidence intervals (C.I.) of the affiliation ratings of the joint analysis of Study 1 and 3, based on model m8. Evaluated affiliation ranges from 1 (avoidance) to 7 (approach).

	No overlap		Overlap	
Synchronization	Mean	95% C.I.	Mean	95% C.I.
Too early	3.91	(3.83, 4.00)	3.62	(3.52, 3.72)
On time	4.06	(3.97, 4.14)	3.77	(3.67, 3.86)
Too late	4.09	(4.00, 4.17)	3.79	(3.69, 3.90)

Tempo

In the joint analysis of Study 1 and 3, the unexpected effect of tempo on evaluated affiliation of Study 1 was not replicated. Slow speech neither increased nor decreased evaluated affiliation [$\beta = 0.05, 95\%\ C.I.\ (-0.02, 0.11), S.E. = 0.03, t = 1.38$]. We therefore suggest that the effect of tempo of Study 1 may be interpreted as a type I error.

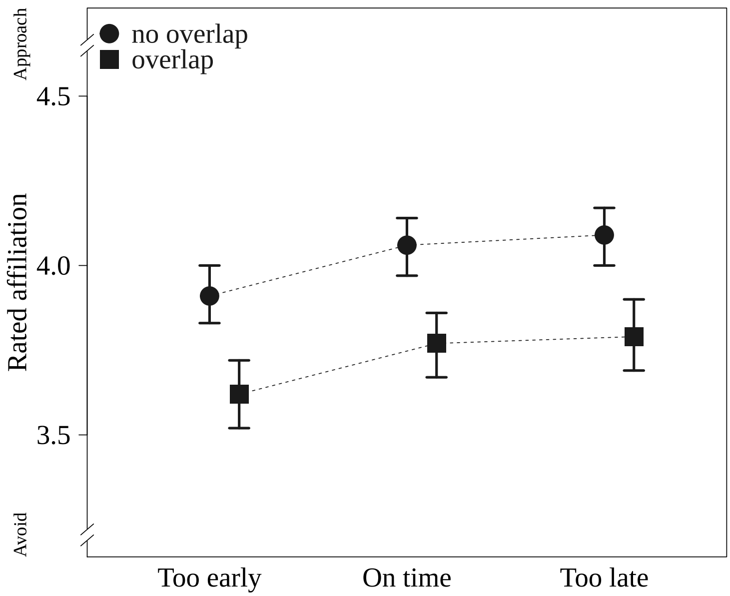

Figure 5.1. Bootstrapped (fitted) means and 95% confidence intervals of evaluated affiliation of the joint analysis of Study 1 and 3 broken down by synchronization and overlap. The round symbols represent the bootstrapped fitted means for the non-overlapping speakers, square symbols the bootstrapped fitted means for the overlapping speakers. The bootstrapped 95% confidence intervals of each condition are represented by the vertical error bars. Evaluated affiliation ranges from 1 (avoidance) to 7 (approach).

Overlap

The combined analysis revealed a main effect of overlap on evaluated affiliation. Evaluations of affiliation were lower when speakers produced overlapping talk [$\beta = -0.29, 95\%\, C.I.\ (-0.39, -0.19), S.E. = 0.05$, $t = -5.90$].

Moreover, overlap also played a role in the individual differences of rated affiliation. When listening to overlapping speakers, evaluated affiliation differed more substantially between participants than when listening to non-overlapping speakers. This indicates that participants display more variability when it comes to the evaluation of speakers who produce overlapping talk, compared to non-overlapping speakers in conversation.

Synchronization

Crucially, the joint analysis also revealed a main effect of synchronization on the ratings of affiliation, independent of overlap and tempo. Affiliation ratings were lower when speakers responded ‘too early’ [$\beta = -0.14, 95\%\ C.I.\ (-0.22, -0.07), S.E. = 0.04, t = -3.56$]. Evaluated affiliation was not affected for speakers responding ‘too late’ [$\beta = 0.03$, $95\%\ C.I.\ (-0.05, 0.10), S.E. = 0.04, t = 0.71$]. The results of the model comparisons do not show an interaction between overlap and synchronization.

Thus, in contrast to Study 1, the joint analysis did not reveal an interaction between overlap and synchronization; being ‘too late’ for overlapping speakers was not rated as more affiliative. We therefore suspect that the effect of being ‘too late’ for overlapping speakers of Study 1 may be interpreted as a type I error as well.

Summary of the results

Taken together, the results of the joint analysis of Study 1 and 3 revealed main effects of overlap and synchronization. The presence of an overlap, resulted in lower ratings of affiliation, as did responding ‘too early’, but not ’too late’. The two unexpected effects of Study 1 were not corroborated by the joint analysis, suggesting that they may be taken as false positives which we should interpreted with caution.

Table 5.5 again shows the results of rated affiliation of our three experimental studies (upper panel, which is a copy of Table 5.1). In addition to the findings of rated affiliation per study, the lower panel of Table 5.5 also presents the results of rated affiliation of the joint analysis of Study 1 and 3, as well as the results of corrugator activity of Study 3. The findings depicted in the lower panel of Table 5.5 will guide our discussion.

Table 5.5

Overview of the effects on the affiliation ratings per experimental factor and study. The upper paned displays the effects per experimental factor on rated affiliation of Study 1, 2, and 3. The lower panel displays the effects on rated affiliation of the joint analysis (Study 1 and 3) and the effect on corrugator activity of Study 3.

Experimental factor	Ratings Study 1		Ratings Study 2		Ratings Study 3	
1. Overlap	Overlap	*	-	-	Overlap	*
2. Synchronization	Too early, n.o.	*	Too early, n.o.	n.s.	Too early, n.o.	*
	Too early, ov.	*	-	-	Too early, ov.	*
	Too late, n.o.	n.s.	Too-late, n.o.	n.s.	Too late, n.o.	n.s.
	Too late, ov.	?*	-	-	Too late, ov.	n.s.
3. Tempo matching	-	-	Tempo matching	n.s.	-	-
4. Tempo	Slow tempo	?*	Slow tempo	?*	Slow tempo	n.s.
5. Tempo interactions	Interactions	n.s.	Interactions	n.s.	Interactions	n.s.

	Ratings joint analysis Study 1 and 3		Corrugator activity Study 3	
1. Overlap	Overlap	*	Overlap	*
2. Synchronization	Too early, n.o.	*	Too early, n.o.	n.s.
	Too early, ov.	*	Too early, ov.	n.s.
	Too late, n.o.	n.s.	Too late, n.o.	n.s.
	Too late, ov.	n.s.	Too late, ov.	n.s.
3. Tempo matching	–	-	–	-
4. Tempo	Slow tempo	n.s.	Slow tempo	n.s.
5. Tempo interactions	Interactions	n.s.	Interactions	n.s.

Note: * significant, expected; ?* significant, unexpected; n.s. non significant; - not assessed; n.o. no overlap condition; ov. overlap condition

5.3 Interpretation of the results

In this section we will discuss the findings of our experimental studies. As can be seen from Table 5.5 (lower panel), we find robust evidence for effects of overlap on both perceived affiliation as well as on corrugator activity in our paradigm. For the effects of synchronization, however, the results were not so clear cut.

Only being ‘too early’, and not ‘too late’ influenced perceived affilliation, while corrugator activity appeared to be insensitive to synchronization altogether. Moreover, the predictions of tempo matching were not confirmed; against expectations, tempo matching did not influence perceived affiliation. Tempo alone, on the other hand, did unexpectedly affect perceived affiliation, albeit only in Study 1 and 2, and not in Study 3 or the joint analysis. While tempo did affect perceived affiliation in Study 1 and 2, it did not affect corrugator activity.

In the following sections we will present a coherent interpretation of these seemingly complex results. Even though this interpretation is tentative, it is supported by our data and our conservative statistical approach. We will argue that, in the case of synchronization, some specific temporal patterns between speakers do indeed convey affiliation, and that, although we hypothesized otherwise, some others do not. More specifically, we will argue that, in addition to the presence of an overlap between speakers, responding too early given the beat of the previous speaker indeed conveys lack of affiliation. For tempo matching we will argue that, at this point, we cannot confirm nor reject our initial hypothesis.

5.3.1 Presence of overlapping talk

We predicted that the presence of overlapping talk between turn-taking speakers would impact perceived interpersonal affect between those speakers. As can be seen in Table 5.5, this prediction is confirmed by the affiliation rating of our two studies investigating the social impact of overlap (Study 1 and 3) and their joint analysis, as well as by the results of corrugator activity. This is a robust indication that the presence of an overlap between speakers in conversation indeed conveys negative interpersonal affect between those speakers, even in the absence of prosodic or lexical-semantic information.

Moreover, the robustness and stability of the effect of overlap on evaluated affiliation as well as the effect of overlap on corrugator activity attested to the validity of our paradigm, indicate that both our measures were valid and that they were adequately sensitive to our temporal manipulations.

Affiliation ratings

Speakers who overlapped were rated as less affiliative than speakers who did not produce overlap. These results are in line with earlier research on the social implications of overlap during turn timing (Beňuš et al., 2011; Goldberg, 1990; Maat et al., 2010) which showed that overlapping speakers are generally perceived as more assertive and dominant (Beňuš et al., 2011), and more rude and disrespectful (Goldberg, 1990) than speakers who do not produce overlap. Presumably, these attributions arise from the impression that somehow the overlapping speaker is unwilling or unable to wait for the other speaker, thereby providing evidence that the overlapping speaker is indifferent, or insensitive, to the other speaker. The overlapper may be insensitive to, among others, the other speaker in general, to his (expressed) thoughts, feelings, or face needs (Brown & Levinson, 1987). For these reasons overlaps are often evaluated as a sign of disrespect, and often linked to interpersonal processes such as power, dominance and control (Goldberg, 1990). It is most likely that the overlaps in our studies were also taken as signs of disrespect, leading to impressions of negative affiliation between speakers.

As briefly discussed in Chapter 1, under different circumstances overlaps may convey rapport, cooperation and camaraderie as well (Goldberg, 1990). The author therefore proposes two types of overlaps ('interruptions'): power-oriented interruptions, and rapport-oriented interruptions. "Power-oriented interruptions are generally heard as rude, impolite, intrusive and inappropriate; conveying the interruptor's antipathy, aggression, hostility, dislike, disdain, apathy, etc. towards the interrupted speaker and/or the talk at hand." (p. 890). "Rapport-oriented interruptions, on the other hand, are generally understood as expressions of open empathy, affection, solidarity, interest, concern, etc." (p. 890).

Such distinction raises the question of what makes a given instance of an overlap, specifically in our paradigm, a sign of disrespect, instead of a sign of cooperation. Goldberg (1990) proposes a system to differentiate between these two types of overlaps on the basis of the 'fit' in terms of cohesion and coherence between the interupting and interrupted utterance. Such an analysis cannot be applied to our paradigm because our speech is unintelligible and the context of the dialogues remains unspecified to our overhearing participants. Other researchers have tried to make this distinction on the basis of the number of words or syllables, i.e., the duration of the overlap, but without success (Murray, 1985). Murray (1985) claims that there are no absolute syntactical or acoustical criteria available for recognizing an occurrence as a power-oriented interruption, neither to those involved in a speech event nor to analysts. The authors conclude that "to understand either how smooth conversation is done [rapport-oriented interruption] or what sometimes goes wrong [power-oriented interruption] requires a fuller analysis of the context than counting syllables of simultaneous speech." (p. 38).

For our participants, the necessary context to be able to interpret the overlap as either power- or rapport-oriented was not available, or at least it was not explicitly defined to them. In our paradigm, participants received neither explicit information about the speakers, nor any context or cover story about the speakers' situation. When listening to the dialogues, the speakers' voices were unfamiliar, they sounded muffled and they were unintelligible. Under these undefined circumstances, it may well be that listeners fall back on their default interpretation of an overlap in conversation. On average, overlaps appear to be taken as conveying negative affiliation, at least for the Dutch participants of our studies.

However, our participants also displayed individual differences to our overlap manipulation. The individual differences to overlapping speakers were much larger than those in response to non-overlapping speakers. This suggests that an overlap, at least in our paradigm, may have different connotations for different listeners, even though the presence of an overlap was, on average, clearly linked to negative interpersonal attributions. The individual differences moreover suggest that not all participants interpreted the presence of an overlap as equally disruptive. Some listeners may evaluate overlapping speakers as very disrespectful, others evaluate the overlapper as only mildy disaffiliative, while others even evaluate the overlapper as positively affiliating. These results suggest that the default interpretation of an overlap in conversation may be different for different listeners. In general however, our findings suggests that the presence of an overlap in conversation robustly conveys negative interpersonal affect in the social dynamics of turn-taking.

Corrugator activity

In addition to leading to negative interpersonal *impressions*, the presence of an overlap in conversation appeared to directly trigger the *affective system* of third-party listeners as well. When overhearing conversations between overlapping speakers, listeners frowned more than when overhearing conversations between speakers who did not overlap. As previously discussed, frowning or corrugator activity is a reliable and fast indicator of the affective valence of a given stimulus (Larsen et al., 2003; Van Boxtel, 2010). Together with the affiliation ratings, this suggests that the mere presence of overlap in conversation not only robustly influences the *impression* of interpersonal affect, but that it also leads to negative emotions, even for third-party listeners.

As pointed out in Chapter 4, the results of the corrugator muscle could also be attributable to the fact that listening to overlapping talk may be more challenging than listening to non-overlapping talk. We argued that this option does not seem parsimonious given that it implies that the processes giving rise to the recorded affective response (corrugator activity) and the impression of affiliation (ratings) should be be disconnected, and that our listeners would somehow rely on learned patterns to assess affiliation, while such patterns can only be learned through *affective* processes.

The more parsimonious explanation would be that overlap strongly conveys interpersonal negativity, impacting both the rated *impression* of affiliation as well as the recorded affective response of our overhearing listeners.

In all, the presence of an overlap in conversation appears to shape listeners' intuitions about interpersonal affect. Listening to speakers who produce overlapping talk in conversation led to the impression that these speakers were less socially close or less connected than when those speakers did not produce overlap. Moreover, merely observing overlapping speakers led to an increase in corrugator activity. This indicates that the presence of overlap not only conveys disaffiliation between speakers, but that it may lead to negative emotions as well, even for mere observers of a conversation.

5.3.2 Synchronization

We predicted that synchronization between turn-taking speakers would impact interpersonal affect between those speakers, on top of the established effects of overlap and gap between them. Specifically, we hypothesized that desynchronization, as in the case of placing the first stressed syllable of a speaking turn too early, or too late, given the previously established rhythm, would lead to the impression of a lack of affiliation between those speakers. Additionally, we hypothesized that listening to desynchronizing speakers would lead to negative emotions as well, since we assumed that the impression of affiliation relies on the affective system of the listener.

Affiliation ratings

In the section that follows we will elaborate on the findings of synchronization on rated affiliation. We will first discuss the effects of being ‘too early’, before discussing the effects of being ‘too late’.

Too early Crucially and in line with our predictions, responding ‘too early’ given the previously established beat did influence perceived interpersonal affect, independent of speech tempo and the presence of overlap. For Study 1 and 3 (Table, 5.5, upper panel) as well as for the joint analysis of Study 1 and 3 (Table 5.5, lower panel), responding ‘too early’ was evaluated as less affiliative than responding ‘on the beat’ (on time), both for non-overlapping (n.o.) as well as for overlapping speakers (ov.). Although the effect of ‘too early’ on evaluated affiliation was very small, it was robust and replicable. For Study 2, however, we did not find effects on perceived affiliation of being ‘too early’. We suspect that the cause for this null result was the fact that in Study 2 speech tempo was presented in a mixed instead of blocked fashion, as in Study 1 and 3. Section 1.3.2 will discuss the issue of why being ‘too early’ in Study 2 did not influence rated affiliation in more detail.

One may wonder to what extent the effect of being ‘too early’ is due to the co-varying gap or overlap duration resulting from our synchronization manipulation, instead of to a true synchronization effect. In our paradigm, dialogues in which speakers respond ‘too early’ have largest overlaps and smallest gaps, compared to dialogues in which speakers respond ‘on time’ or ‘too late’. This explanation seems implausible given that, in our paradigm, gap and overlap durations are within normal range for all synchronization conditions (Heldner & Edlund, 2010; Stivers et al., 2009). Moreover, we did not find any interactions with tempo in any of our studies, as can be seen in Table 5.5. The absence of such interactions with tempo indicate that the duration of the gap or overlap cannot account for the affiliation ratings alone. Finally, being ‘too late’, even in the absence of tempo interactions, was not rated as structurally more affiliative, as one would expect if the effect of being ‘too early’ would be due to gap or overlap duration.

Together this suggests that the effect of being ‘too early’ can indeed be attributed to the lack of speech rhythm synchronization between speakers. Responding ‘off the beat’ in the case of responding ‘too early’ really conveys a lack of affiliation between speakers in conversation.

Too late In contrast to responding ‘too early, responding ‘too late’ did not influence evaluated affiliation as expected. The joint analysis revealed that responding ‘too late’ was rated as equally affiliative as responding ‘on time’, even though we initially hypothesized that responding ‘too late’ (as well as ‘too early’) would be rated as less affiliative than responding ‘on time’.

A possible explanation for the asymmetry between the effects of being ‘too early’ and ‘too late’ on rated affiliation might be that being ‘too late’ is not as socially disruptive as being ‘too early’. Responding too slow, as in the case of a (long) pause between speaking turns can, besides social dimensions, also indicate cognitive variables, such as problems with comprehension (Beňuš et al., 2011), planning difficulties (Bull & Aylett, 1998), fluency (Bosker, 2014), or word frequency (Hartsuiker & Notebaert, 2010). Due to such cognitive challenges, speakers may be unable to respond ‘on the beat’, and instead their stressed syllable may occur slightly later than initially intended. Listeners may well take these possibilities into account when assessing interpersonal dimensions on the basis of temporal patterns in conversation. Being ‘too late’ may thus be inherently ambiguous for the listener, thereby attenuating the effect on rated affiliation of ‘too late’ turns.

Corrugator activity

We found no effects of synchronization on corrugator activity. Independent of overlap and tempo, the activity of the corrugator muscle of our participants did not differ between listening to speakers who synchronized or who did not synchronize their speech rhythms. Given the findings on the ratings of affiliation, one would expect participants to frown more when listening to speakers responding ‘too early’ than when listening to speakers responding ‘on time’. As discussed in Study 3, two accounts may explain the lack of agreement between the affiliation ratings and corrugator activity of our participants.

Under the first account, synchronization functions as a ‘cold’ cue, that is, participants may not have responded *affectively* when listening to speakers being ‘too early’, even though we hypothesized otherwise. While this option may not seem very parsimonious, it should be considered seriously given our data.

Under the second account, synchronization does function as an affective cue. However, due to an accumulation of factors the effect of synchronization on corrugation may not have been substantial enough to reach significance. As discussed in Study 3, the fact that the corrugator muscle seemed to be less sensitive to our temporal manipulations than the affiliation ratings, the use of overhearing listeners instead of participating interlocutors, and the subtle social or acoustic nature of synchronization may all have contributed to to this null result.

Nevertheless, even though listeners seem insensitive to synchronization as an *affective* cue, synchronization, being ‘too early’ somehow does impact the *affective impression* of interpersonal affect, suggesting that responding off the beat, as in the case of being ‘too early’ is a real sign of disaffiliation between speakers in conversation.

5.3.3 Tempo matching

Against expectations, tempo matching did not influence evaluated affiliation. We predicted that speakers who matched their tempo with that of the preceding speaker would be rated as more affiliative than speakers who did not match their speech tempo during conversation. This prediction was not borne out.

As a result of our tempo matching factor, the tempos of both speakers in Study 2 were presented in a mixed fashion. In Study 1 and 3, however, tempo and overlap were both presented in a blocked fashion, allowing our participants to be maximally sensitive to our synchronization manipulation. It may well be that under these less restricted circumstances of varying tempo, listeners were unable to notice the subtle interplay between the two speakers, i.e., they may not have detected our tempo matching and synchronization manipulations altogether. This explanation is reinforced by the fact that, in Study 2, synchronization did not influence the ratings of affiliation either, as previously pointed out and shown in Table 5.5. In our already challenging paradigm, it seems that the experimental design of Study 2 involved too many attenuating factors, preventing our participants to be sensitive to the subtle manipulations of tempo matching and synchronization between speakers. We are therefore unable to confirm or reject our initial hypothesis about tempo matching.

5.3.4 Tempo

Speech tempo alone unexpectedly influenced evaluated affiliation. As previously discussed, we included speech tempo in our experimental design to be able to deconfound the effects of synchronization from gap (or overlap) duration. In our paradigm, fast speech inherently has smaller gap durations and larger overlap durations than slow speech, and, as a result of our synchronization manipulation, gap (and overlap) durations also co-varied with our levels of synchronization. The presence of any interaction with tempo would be an indication that gap (or overlap) duration, and not synchronization, affected evaluated affiliation. Crucially, as can be seen in Table 5.5, neither the findings of the individual studies, nor the results of the joint analysis revealed any interactions with tempo. The absence of such tempo interactions suggests that the ratings of affiliation cannot be explained by the duration of the silent gap or overlap alone.

Even though we had no a priori theoretical expectations about the effects of speech tempo on evaluated affiliation, Study 1 and 2 revealed that speakers speaking slowly were rated as more affiliative than speakers speaking fast. As already pointed out, we suggested that the effect of tempo of Study 1 may be attributed to the low statistical power of that study, giving rise to spurious effects. This idea was corroborated in Study 3 as well as in the joint analysis, where we found no effects of tempo; neither evaluated affiliation nor corrugator activity was affected by the speech tempo of the speakers.

However, the tempo effect of Study 2 should be interpreted differently, since the presentation of tempo in this study was different from that of Study 1. As previously discussed, the tempo of both speakers was presented in a mixed fashion in Study 2, instead of in a blocked fashion, as in Study 1 and 3. As a result of the varying tempo, listeners appear to somehow only be able to focus on characteristics of the last speaker, as if that is the only cue available for their evaluation of affiliation.

Together this suggests that speech tempo alone does not, or only very marginally, influence the impression of affiliation in real-life interaction.

5.3.5 Other issues

Gap and overlap duration

In Study 1 we suggested that overlap *duration* between turn-taking speakers might also influence interpersonal dimensions, on top of the effect of the presence of overlap alone. This idea resulted from the finding that, for overlapping speakers in Study 1, being 'too late' was rated as relatively more affilitive than being 'on time' (see Table 5.5). As previously pointed out, the amount of overlap (and gap) co-varied with our synchronization manipulation. We therefore assumed that this 'too late' effect for overlapping speakers may in fact be an effect of the amount of overlap, instead of a true synchronization effect. This 'overlap duration' effect however was neither replicated in Study 3, nor corroborated by our joint analysis of Study 1 and 3, nor visible as an interaction with tempo in any of our studies. This suggests that the unexpected effect of being 'too late' in Study 1 may be interpreted as a false negative, attributable to the low statistical power of that study, instead of being interpreted as resulting from the gap or overlap overlap durations, as hypothesized in the discussion of Study 1.

In our paradigm, both gap, as well as overlap durations were well within normal range (Heldner & Edlund, 2010; Stivers et al., 2009). It may well be that outside this range, or under different circumstances, overlap (or gap) duration affects interpersonal dimensions during turn-taking, on top of the effects of the presence of overlap alone.

Conversational tendencies

Even though the relative impact of overlap and responding ‘too early’ may not be the same, as previously pointed out, both responding ‘too early’ as well as responding with an overlap seem to involve a violation of some sort of conversational ‘norm’ or tendency. When producing overlap, speakers may violate the assumption or norm that you will only start speaking when the other speaker has finished. This is often referred to as the so-called ‘no-gap-no-overlap’ tendency during turn-taking (Sacks et al., 1974; Stivers et al., 2009). For synchronization, responding ‘too early’ may violate the norm that your speech rhythm has to coincide with the rhythm of the previous speaker.

Wilson and Wilson (2005) argue that this may indeed be the case for speech rhythm during turn-taking. The authors claim that this rhythmic coupling governs when a speaker, or listener, as potential next speaker, initiates speech at any given moment. Presumably, such conversational norms or tendencies facilitate smooth alignment between speakers during turn-taking (Sacks et al., 1974; Stivers et al., 2009; Wilson & Wilson, 2005). The results of our studies indicate that deviations from these conversational norms indeed have communicative consequences.

Moreover, our findings provide further evidence for the communication accommodation theory (CAT), as discussed in Chapter 1. CAT claims that ‘between-speaker accommodation’, i.e., coordination, helps to facilitate coherent interaction and, importantly, it allows interactants to manage social distance between one another (Chartrand & van Baaren, 2009). The more speakers wish to affiliate, the more they accommodate. Our findings reveal that this may indeed be the case; when smooth conversation is hindered, as in the case of overlapping talk or being ‘too early’, negative attributions arise.

Ecological validity

In our quest to isolate the ‘pure’ effects of synchronization, we have obtained our effect of synchronization with a slightly unnatural paradigm, which may raise questions about the ecological validity of our findings.

We used synthesized speech to be able to very precisely control the micro-timing between speakers, low-pass filtered speech to avoid lexical-semantic interference during listening, and we counterbalanced intonation and speaker voice in order to control for their possible influence on the impression of affiliation. This approach ensured that the observed effects of synchronization could only be attributed to our synchronization manipulation. However, this approach also resulted in a slightly unnatural paradigm, which makes it difficult to estimate the actual effects of synchronization in real life.

On the one hand, our paradigm may have led to an overestimation of the real life effects of synchronization. In Study 2, where we did not find effects of synchronization, the tempo of both speakers was presented in a mixed fashion, instead of in a blocked fashion, as in Study 1 and 3. The restriction of varying tempo could be taken to suggest that social-affective effects of synchronization are not of the same importance in the inherently more variable environments of real life interactions.

On the other hand, the restriction of varying tempo may also be an indication of the limits of our unnatural experimental paradigm. Due to the high degree of control over the acoustic parameters, the nature of the dialogues was rather static, making it difficult for the participants to be truly engaged with the materials. Moreover, by using overhearers rather than interactants it may have been even harder for our participants to be really involved. In Study 2, where tempo was presented in a mixed fashion, the task of the participants was even more challenging, which may have attenuated the effect of synchronization even more. Our highly controlled, but slightly unnatural paradigm, may have attenuated the effects of synchronization. In real life interactions, being ‘too early’ may therefore be of greater influence than observed in our studies.

At this point, we have no independent evidence that unequivocally indicates one of the two scenarios as the most plausible. It is also possible that they both contributed to the observed effects, even though their relative impact is still unknown. In any case, it may be worth investigating the affective impact of synchronization with a more natural paradigm, for example by using corpus-based, instead of synthesized speech.

Nevertheless, even with this paradigm, many listeners expressed judgments such as 'it sounded Italian' or 'sometimes they sounded really agitated', indicating that they did experience the materials as natural conversational speech. It therefore seems reasonable to assume that speech rhythms synchronization has similar social connotations in everyday conversations.

5.4 Conclusion

In this dissertation we sought to unravel whether specific temporal patterns between speakers during turn-taking may convey interpersonal affect between those speakers, on top of the established effects of a silent gap or the presence of an overlap between them. According to our findings, this is truly the case, albeit in a limited way.

By integrating different perspectives from the literature about joint action, phonetics, and emotion we reveal that, like temporal coordination in the non-verbal domain, certain kinds of temporal coordination in the *verbal* domain also convey affiliation. Our findings moreover testify to the social importance of accommodation (as proposed in CAT, e.g., Giles, 2016), and point to hitherto overlooked new dimensions of accommodation.

In addition to the presence of overlapping talk, responding too early given the beat of the previous speaker conveys disaffiliation. So it does not only matter what you say and whether you interrupt the preceding speaker, it sometimes also matters whether you say it on the beat of the preceding rhythm. 'Being in sync' is not just a figure of speech, but a real sign of affiliation in spoken dialogue.

References

Abdai, J., & Miklóosi, Á. (2016). The origin of social evaluation, social eavesdropping, reputation formation, image scoring or what you will. *Frontiers in Psychology*, *7*, 1772.

Aubanel, V., Davis, C., & Kim, J. (2016). Exploring the role of brain oscillations in speech perception in noise: Intelligibility of isochronously retimed speech. *Frontiers in Human Neuroscience*, *10*, 430.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390-412.

Babel, M. (2010). Dialect divergence and convergence in New Zealand English. *Language in Society*, *39*(4), 437-456.

Babel, M. (2012). Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics*, *40*(1), 177-189.

Barnes, R., & Jones, M. R. (2000). Expectancy, attention, and time. *Cognitive psychology*, *41*(3), 254-311.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1-48.

Bavelas, J. B., Black, A., Chovil, N., Lemery, C. R., & Mullett, J. (1986). I show how you feel - motor mimicry as a communicative act. *Journal of personality and social psychology*, *50*(2), 322-329.

Bavelas, J. B., Black, A., Chovil, N., Lemery, C. R., & Mullett, J.

(1988). Form and function in motor mimicry - topographic evidence that the primary function is communicative. *Human Communication Research*, *14*(3), 275-299.

Beňuš, Š. (2009). Are we 'in sync': Turn-taking in collaborative dialogues. In *Proceedings of the 10th INTERSPEECH* (p. 2167-2170).

Beňuš, Š., Gravano, A., & Hirschberg, J. (2011). Pragmatic aspects of temporal accommodation in turn-taking. *Journal of Pragmatics*, *43*(12), 3001-3027.

Bergeson, T., & Trehub, S. (2002). Absolute pitch and tempo in mothers' songs to infants. *Psychological Science*, *13*(1), 72-75.

Bernieri, F. J., & Rosenthal, R. (1991). Interpersonal coordination: Behavior matching and interactional synchrony. In R. S. Feldman & B. Rime (Eds.), *Fundamentals of nonverbal behavior* (p. 401-432). Cambridge University Press.

Bertinetto, P. (1989). Reflections on the dichotomy "stress" vs. "syllable-timing.". *Revue de Phontique Applique*, 91-93:99-130.

Bögels, S., & Torreira, F. (2015). Listeners use intonational phrase boundaries to project turn ends in spoken interaction. *Journal of Phonetics*, *52*, 46 - 57.

Boltz, M. G., & Miller, A. R. (2010). Jo are you lying to me? temporal cues for deception. *Journal of Language and Social Psychology*, *29*(4), 458-466.

Bosker, H. R. (2014). *The processing and evaluation of fluency in native and non-native speech.* Utrecht University, Utrecht: LOT. (doctoral disseration)

Brokx, J. P. L., & Nooteboom, S. G. (1982). Intonation and the perceptual separation of simultaneous voices. *Journal of Phonetics*, *10*, 23-36.

Brown, P., & Levinson, C. (1987). *Politeness: some universals of language usage*. Cambridge University Press.

Bull, M. C., & Aylett, M. P. (1998). An analysis of the timing of turn-taking in a corpus of goal-orientated dialogue. In *Proceedings of the 5th International Conference on Spoken Language Processing, ICSLP 1998* (p. 1175-1178).

Buller, D. B., & Aune, R. K. (1992). The effects of speech rate similarity on compliance: Application of communication accommodation theory. *Western Journal of Communication (includes Communication Reports)*, *56*(1), 37-53.

Campbell, J., Cothren, D., & Burg, A. (2010). The perpetuation of entrained behavior during computer-mediated communication. *Social Influence*, *5*(1), 59-73.

Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception-behavior link and social interaction. *Journal of personality and social psychology*, *76*(6), 893-910.

Chartrand, T. L., & van Baaren, R. (2009). Human mimicry. *Advances in Experimental Social Psychology*, *41*, 219-274.

Chow, I., Belyk, M., Tran, V., & Brown, S. (2015). Syllable synchronization and the P-center in Cantonese. *Journal of Phonetics*, *49*, 55-66.

Cirelli, L. K. (2014). Interpersonal synchrony increases prosocial behavior in infants. *Developmental science*, *17*(6), 1003-3011.

Clark, H. (1996). *Using language*. Cambridge University Press.

Cohen, B. H., Davidson, R. J., Senulis, J. A., Saron, C. D., & Weisman, D. R. (1992). Muscle tension patterns during auditory attention. *Biological psychology*, *33*(2-3), 133-156.

Crystal, T. H., & House, A. S. (1990). Articulation rate and the duration of syllables and stress groups in connected speech. *Journal of the Acoustical Society of America*, *88*(1), 101-112.

Cummins, F. (2009a). Rhythm as an affordance for the entrainment of movement. *Phonetica*, *66*(1-2), 15-28.

Cummins, F. (2009b). Rhythm as entrainment: The case of synchronous speech. *Journal of Phonetics*, *37*(1), 16-28.

Cummins, F. (2015). Rhythm and speech. In M. A. Redford (Ed.), *Handbook of speech production* (p. 158-177). Wiley-Blackwell.

Cummins, F., & Port, R. (1998). Rhythmic constraints on stress timing in English. *Journal of Phonetics*, *26*(2), 145-171.

Cutler, A., & Mehler, J. (1993). The periodicity bias. *Journal of Phonetics*, *21*, 103-108.

Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics*, *11*(1), 51-62.

Davis, M. H. (1983). Measuring individual-differences in empathy – evidence for a multidimensional approach. *Journal of personality and social psychology*, *44*(1), 113-126.

De Jong, K. J. (1994). The correlation of P-Center adjustments with articulatory and acoustic events. *Perception and psychophysics*, *56*(4), 447-460.

Decety, J., & Cowell, J. M. (2014). Friends or foes: Is empathy necessary for moral behavior? *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *9*(4), 525-537.

Demos, A. P., Chaffin, R., Begosh, K. T., Daniels, J. R., & Marsh, K. L. (2012). Rocking to the beat: Effects of music and partner's movements on spontaneous interpersonal coordination. *Journal of Experimental Psychology–General*, *141*(1), 49-53.

Derksen, A., & Menert, L. (2013). *Fluency TTS 6. 0 (customized) [computer program].*

Dimberg, U., & Thunberg, M. (1998). Rapid facial reactions to emotional facial expressions. *Scandinavian Journal of Psychology*, *39*(1), 39-45.

Donovan, A., & Darwin, C. J. (1979). The perceived rhythm of speech. In *Proceedings of the 9th International Congres of Phonetic Sciences, ICPhS, 1979* (Vol. 2, p. 268-274).

Fischer, A. H., & Manstead, A. S. (2016). Social functions of emotion and emotion regulation. In L. F. Barrett, M. Lewis, & J. M. Haviland-Jones (Eds.), *Handbook of emotions* (p. 424-439). Guilford.

Fischer, A. H., Rotteveel, M., Evers, C., & Manstead, A. S. R. (2004). Emotional assimilation: How we are influenced by others' emotions. *Cahier de Psychologie Cognitive*, *22*(2), 223-245.

Frijda, N. H. (2009). Emotion experience and its varieties. *Emotion Review*, *1*(3), 264-271.

Garrod, S., & Pickering, M. J. (2009). Joint action, interactive alignment, and dialog. *Topics in Cognitive Science*, *1*(2), 292-304.

Giles, H. (Ed.). (2016). *Communication accommodation theory: negotiating personal relationships and social identities across contexts*. Cambridge University Press.

Giles, H., Coupland, J., & Coupland, N. (Eds.). (1991). *Contexts of accommodation*. Cambridge University Press.

Goldberg, J. A. (1990). Interrupting the discourse on interruptions - an analysis in terms of relationally neutral, power-oriented and rapport-oriented acts. *Journal of Pragmatics*, *14*(6), 883-903.

Goldman-Eisler, F. (1968). *Psycholinguistics: Experiments in spontaneous speech*. Academic Press.

Hamlin, J. K. (2015). The case for social evaluation in preverbal infants: gazing toward one's goal drives infants' preferences for helpers over hinderers in the hill paradigm. *Frontiers in Psychology*, *5*, 1563.

Hartsuiker, R. J., & Notebaert, L. (2010). Lexical access problems lead to disfluencies in speech. *Experimental Psychology*, *57*(3), 169-177.

Hatfield, E., Cacioppo, J., & Rapson, R. L. (1994). *Emotional contagion*. Cambridge Univesity Press.

Hawkins, S. (2014). Situational influences on rhythmicity in speech, music, and their interaction. *Philosophical Transactions of the Royal Society B–Biological Sciences*, *369*(1658), 20130398.

Heldner, M., & Edlund, J. (2010). Pauses, gaps and overlaps in conversations. *Journal of Phonetics*, *38*(4), 555-568.

Hess, U., & Philippot, P. (Eds.). (2007). *Group dynamics and emotional expression*. Cambridge Univesity Press.

Hove, M. J., & Risen, J. L. (2009). It's all in the timing: Interpersonal synchrony increases affiliation. *Social Cognition*, *27*(6), 949-960.

Ilie, G., & Thompson, W. (2006). A comparison of acoustic cues in music and speech for three dimensions of affect. *Music Perception*, *23*(4), 319-329.

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLOS Medicine*, *2*(8), e124.

Kiesling, S. F. (2011). Stance in context: Affect, alignment and investment in the analysis of stancetaking. Bristol, UK: Paper presented at the iMean conference.

Konvalinka, I., Vuust, P., Roepstorff, A., & Frith, C. D. (2010). Follow you, follow me: Continuous mutual prediction and adaptation in joint tapping. *Quarterly Journal of Experimental Psychology*, *63*(11), 2220-2230.

Kuiper, K. (1992). The oral tradition in auction speech. *American Speech*, *67*(3), 279-289.

Kurtić, E., Brown, G. J., & Wells, B. (2013). Resources for turn competition in overlapping talk. *Speech Communication*, *55*(5), 721-743.

Lakens, D. (2010). Movement synchrony and perceived entitativity. *Journal of experimental social psychology*, *46*(5), 701-708.

Lakens, D., & Stel, M. (2011). If they move in sync, they must feel in sync: Movement synchrony leads to attributions of rapport and entitativity. *Social Cognition*, *29*(1), 1-14.

Lakin, J. L., & Chartrand, T. L. (2003). Using nonconscious behavioral mimicry to create affiliation and rapport. *Psychological Science*, *14*(4), 334-339.

Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological review*, *106*(1), 119-159.

Larsen, J. T., Norris, C. J., & Cacioppo, J. T. (2003). Effects of positive and negative affect on electromyographic activity over zygomaticus major and corrugator supercilii. *Psychophysiology*, *40*(5), 776-785.

Lehiste, I. (1977). Isochrony reconsidered. *Journal of Phonetics*, *5*(3), 253-263.

Levinson, S. C., & Enfield, N. J. (Eds.). (2006). *Roots of human sociality*. Berg Publishers.

Lidji, P., Palmer, C., Peretz, I., & Morningstar, M. (2011). Listeners feel the beat: Entrainment to English and French speech rhythms. *Psychonomic bulletin and review*, *18*(6), 1035-1041.

Maat, M. T., Truong, K. P., & Heylen, D. (2010). How turn-taking strategies influence users' impressions of an agent. In *Intelligent Virtual Agents: Proceedings of the 10th International Conference, IVA 2010* (p. 441-453).

Manson, J. H., Bryant, G. A., Gervais, M. M., & Kline, M. A. (2013). Convergence of speech rate in conversation predicts cooperation. *Evolution and Human Behavior*, *34*(6), 419-426.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing type i and power in linear mixed models. *Journal of Memory and Language*, *94*, 305-315.

McGarva, A. R., & Warner, R. (2003). Attraction and social coordination: Mutual entrainment of vocal activity rhythms. *Journal of psycholinguistic research*, *32*(3), 335-354.

Miles, L. K., Nind, L. K., & Macrae, C. N. (2009). The rhythm of rapport: Interpersonal synchrony and social perception. *Journal of experimental social psychology*, *45*(3), 585-589.

Morgan, J. L. (1996). A rhythmic bias in preverbal speech segmentation. *Journal of Memory and Language*, *35*(5), 666-688.

Morton, J. M., & Marcus, F. C. (1976). Perceptual centers (P-centrers). *Psychological review*, *83*(5), 405-408.

Murray, S. O. (1985). Toward a model of members' methods for recognizing interruptions. *Language in Society*, *14*(1), 31-40.

Myers-Schulz, B., Pujara, M., Wolf, R. C., & Koenigs, M. (2013). Inherent emotional quality of human speech sounds. *Cognition and Emotion*, *27*(6), 1105-1113.

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining r2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*(2), 133-142.

Niederhoffer, K. G., & Pennebaker, J. W. (2002). Linguistic style matching in social interaction. *Journal of Language and Social Psychology*, *21*(4), 337-360.

Ogden, R., & Hawkins, S. (2015). Entrainment as a basis for coordinated actions in speech. In *Proceedings of the 18th International Congres of Phonetic Sciences, ICPhS, 2015* (p. 599:1-5).

Panksepp, J., & Biven, L. (2012). *The archaeology of mind: Neuroevolutionary origins of human emotions*. W. W. Norton & Company.

Pardo, J. S., Gibbons, R., Suppes, A., & Krauss, R. M. (2012). Phonetic convergence in college roommates. *Journal of phonetics*, *40*(1), 190-197.

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS*. Springer.

Port, R. F. (2003). Meter and speech. *Journal of phonetics*, *31*(3-4), 599-611.

Quené, H. (2008). Multilevel modeling of between-speaker and within-speaker variation in spontaneous speech tempo. *Journal of the Acoustical Society of America*, *123*(2), 1104-1113.

Quené, H., & Port, R. F. (2005). Effects of timing regularity and metrical expectancy on spoken-word perception. *Phonetica*, *62*(1), 1-13.

Quené, H., & Van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language*, *59*(4), 413-425.

R Core Team. (2015). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. Retrieved from `http://www.R-project.org/`

Ramig, L. A. (1983). Effects of physiological aging on speaking and reading rates. *Journal of communication disorders*, *16*(3), 217-226.

Repp, B. H., & Su, Y. (2013). Sensorimotor synchronization: A review of recent research (2006-2012). *Psychonomic bulletin and review*, *20*(3), 403-452.

Richardson, M. J., Marsh, K. L., & Schmidt, R. C. (2005). Effects of visual and verbal interaction on unintentional interpersonal coordination. *Journal of Experimental Psychology–Human Perception and Performance*, *31*(1), 62-79.

Roberts, F., Margutti, P., & Takano, S. (2011). Judgments concerning the valence of inter-turn silence across speakers of American English, Italian, and Japanese. *Discourse processes*, *48*(5), 331-354.

Robinson, L. F., & Reis, H. T. (1989). The effects of interruption, gender, and status on interpersonal perceptions. *Journal of Nonverbal Behavior*, *13*(3), 141-153.

Ryan, W. J. (1972). Acoustic aspects of the aging voice. *Journal of gerontology*, *27*(2), 265-268.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, *50*(4), 696735.

Scherer, K. R. (2005). "what are emotions? and how can they be measured?". *Social Science Information*, *44*(4), 693727.

Schultz, B. G., O'Brien, I., Phillips, N., Mcfarland, D. H., Titone, D., & Palmer, C. (2016). Speech rates converge in scripted turn-taking conversations. *Applied Psycholinguistics*, *37*(5), 1201-1220.

Scott, S. K. (1993). *P-centers in speech: An acoustic analysis.* University College London, Londen. (doctoral disseration)

Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: bodies and minds moving together. *Trends in cognitive sciences*, *10*(2), 70-76.

Shockley, K., Baker, A. A., Richardson, M. J., & Fowler, C. A. (2007). Articulatory constraints on interpersonal postural coordination. *Journal of Experimental Psychology–Human Perception and Performance*, *33*(1), 201-208.

Shockley, K., Santana, M. V., & Fowler, C. A. (2003). Mutual interpersonal postural constraints are involved in cooperative conversation. *Journal of Experimental Psychology-Human Perception and Performance*, *29*(2), 326-332.

Stetson, R. (1951). *Motor phonetics: A study of speech movements in action*. Springer.

Stivers, T. (2008). Stance, alignment, and affiliation during storytelling: When nodding is a token of affiliation. *Research on Language and Social Interaction*, *41*(1), 31-57.

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., . . . Levinson, S. C. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(26), 10587-10592.

Street, R. L. (1984). Speech convergence and speech evaluation in fact-finding interviews. *Human Communication Research*, *11*(2), 139-169.

Tassinary, L. G., Cacioppo, J. T., & Vanman, E. J. (2007). The skeletomotor system: Surface electromyography. In G. G. B. J. T. Cacioppo L. G. Tassinary (Ed.), *Handbook of psychophysiology* (p. 267-302). Cambridge University Press.

Ten Bosch, L., Oostdijk, N., & de Ruiter, J. P. (2004). Durational aspects of turn-taking in spontaneous face-to-face and telephone dialogues. In *Text, Speech and Dialogue: Proceedings of the 7th International Conference TSD 2004* (p. 563-570). Springer.

't Hart, J. T., Collier, R., & Cohen, A. (1990). *A perceptual study of intonation: An experimental-phonetic approach to speech melody*. Cambridge University Press.

Tomasello, M. (2008). *Origins of human communication*. Cambridge MA: MIT Press.

Valdesolo, P., & DeSteno, D. (2011). Synchrony and the social tuning of compassion. *Emotion*, *11*(2), 262-266.

Van Berkum, J. J. A. (in press). Language comprehension, emotion, and sociality: aren't we missing something? In S. A. Rueschemeyer & M. G. Gaskell (Eds.), *Oxford handbook of psycholinguistics.* Oxford: OUP.

Van Boxtel, T. (2010). Facial EMG as a tool for inferring affective states. In A. J. Spink, F. Grieco, O. E. Krips, L. W. S. Loijens, L. P. J. J. Noldus, & P. H. Zimmerman (Eds.), *Measuring Behavior 2010, Proceedings of the 7th International Conference on Methods and Techniques in Behavioral Research* (p. 104-108).

van Beinum, F. K. K., & van Donzel, M. E. (1996). Relationship between discourse structure and dynamic speech rate. In *The 4th International Conference on Spoken Language Processing, ICSLP 1996* (Vol. 3, p. 1727-1731).

Verhoeven, J., Pauw, G. D., & Kloots, H. (2004). Speech rate in a pluricentric language: A comparison between Dutch in Belgium and the Netherlands. *Language and speech*, *47*(3), 297-308.

Vicaria, I. M., & Dickens, L. (2016). Meta-analyses of the intra- and interpersonal outcomes of interpersonal coordination. *Journal of Nonverbal Behavior*, *40*(4), 335-361.

Wilson, M., & Wilson, T. P. (2005). An oscillator model of the timing of turn-taking. *Psychonomic bulletin & review*, *12*(6), 957-968.

Wlodarczak, M., Šimko, J., & Wagner, P. (2012). Syllable boundary effect: Temporal entrainment in overlapped speech. In *Proceedings of the 6th International Conference on Speech Prosody* (p. 611-614). Tongji University Press.

Yuan, J., Liberman, M., & Cieri, C. (2006). Towards an integrated understanding of speaking rate in conversation. In *Proceedings of the 9th International Conference on Spoken Language Processing, ICSLP 2006* (p. x. 1-4).

Zuckerman, M., & Driver, R. E. (1989). What sounds beautiful is good - the vocal attractiveness stereotype. *Journal of Nonverbal Behaviour*, *13*(2), 67-82.

Zuckerman, M., Hodgins, H., & Miyake, K. (1990). The vocal attractiveness stereotype - replication and elaboration. *Journal of Nonverbal Behaviour*, *14*(2), 97-112.

Zwaan, R. A. (1999). Situation models. *Current Directions in Psychological Science*, *8*(1), 15-18.

Samenvatting in het Nederlands

Dit proefschrift gaat over de sociale effecten van bepaalde temporele patronen tussen sprekers in conversatie.

Een temporele relatie tussen twee sprekers kan op verschillende manieren worden uitgedrukt. Zo kan deze relatie bijvoorbeeld worden beschreven als de afstand in tijd tussen de twee sprekers, zoals de duur van de stilte, of de hoeveelheid overlappende spraak tussen hen. Onderzoek laat zien dat gesprekspartners die langere stiltes tussen hun spreekbeurten hebben als minder (wederzijds) verbonden worden beschouwd dan sprekers die kortere stiltes laten vallen. Daarnaast is het zo dat sprekers die spreken voordat de vorige spreker is uitgesproken als onaangenamer en dominanter worden beschouwd dan sprekers die geen overlap produceren.

Een temporele relatie tussen twee sprekers kan echter ook worden beschreven in termen van spraakritme of spraaktempo. Zo kunnen sprekers bijvoorbeeld hun spraaktempo aanpassen, om zo het tempo van de vorige spreker te evenaren ('tempomatching'), of kunnen ze hun eerste 'beat' zodanig plaatsen dat deze samenvalt met het spraakritme van de vorige spreker ('synchronisatie'). Onderzoek naar de sociale impact van *non-verbale* interpersoonlijke coördinatie, zoals samen lopen, klappen, of tikken, laat zien dat synchronisatie en matching in het non-verbale domein sociale verschillen vermindert en wederzijdse verbondenheid ('affiliatie') vergroot. Ook het enkel *zien* van gesynchroniseerde interacties leidt tot impressies van affiliatie tussen de geobserveerde individuen. Interpersoonlijke coördinatie in het non-verbale domein lijkt dus op een of andere manier te fungeren als een 'sociaal bindmiddel', dat affiliatie creeërt en reflecteert.

In dit proefschrift zijn we geïnteresseerd naar de affectieve impact van synchronisatie en tempomatching in *gesproken* interactie en proberen we te ontrafelen of deze specifieke vormen van interpersoonlijke coördinatie ook wederzijdse verbondenheid reflecteren, net zoals dat het geval is in het non-verbale domein.

Synchronisatie en tempomatching Hoewel spontane spraak vaak niet strikt ritmisch is, wordt ze vaak wel als ritmisch ervaren, en wordt ze, onder bepaalde omstandigheden, ritmisch geproduceerd. Onderzoek naar vraag-antwoord paren laat bijvoorbeeld zien dat sprekers hun spraakritme kunnen synchroniseren, maar wel alleen *lokaal*, namelijk alleen rondom de beurtwisseling van de ene spreker naar de volgende. Het akoestische piekmoment (‘P-center’ of ‘beat’) van het eerste geluid dat de antwoorder maakt valt vaak samen, of is gesynchroniseerd, met de laatste twee of drie beats van de voorafgaande vraag. Onduidelijk is echter of dit type synchronisatie van invloed is op de mate van affiliatie tussen sprekers, zoals dat wel het geval is in het non-verbale domein.

Voor spraaktempo geldt dat tijdens conversatie het tempo van beide sprekers over het algemeen steeds meer gelijk wordt. Net zoals in non-verbale matching wordt deze vorm van matching geassocieerd met positieve affiliatie.

Affectieve impact van synchronisatie en tempomatching

In dit proefschrift proberen we te ontrafelen of synchronisatie van spraakritme en/of tempomatching tussen sprekers wederzijdse verbondenheid (‘affiliatie’) uitdrukt.

In drie experimenten onderzochten we deze vraag. In Studie 1 onderzochten we de impact van synchronisatie op geëvalueerde affiliatie. In Studie 2 onderzochten we de impact van tempomatching en synchronisatie op geëvalueerde affiliatie. In Studie 3 onderzochten we opnieuw onze synchronisatiehypothese. Dit deden we zoals in Studie 1, maar deze keer hadden we zowel een offline (geëvalueerde affiliatie) als een online (fronsspieractivatie) maat van affiliatie.

Methode

In alledrie de studies hebben we de deelnemers naar gespreksfragmenten laten luisteren en vervolgens gevraagd om de mate van affiliatie tussen de sprekers te evalueren. De exacte vraag die de proefpersonen kregen was: *In de reactie van de tweede spreker hoorde ik:*, en daar onder, als de uiteinden van de 7-puntschaal: *toenadering* (‘approach’) en *verwijdering* (‘avoidance’). We kozen deze bewoordingen omdat deze schaal nauw verbonden is met de approach–avoidance-dimensie die centraal staat in algemene theorieën over emotie.

In Studie 3 hebben we naast deze evaluatie ook de fronsspieractivatie van de deelnemers gemeten tijdens het luisteren naar deze fragmenten. Fronsspieractivatie heeft een zeer betrouwbare correspondentie met affectieve valentie, d.w.z. fronsspieractivatie neemt toe in reactie op negatieve stimuli, en af in reactie op positieve stimuli. Op deze manier hoopten we een beter beeld te krijgen van de, wellicht onbewuste, affective impact van onze temporele manipulaties.

Manipulaties De fragmenten die de deelnemers gepresenteerd kregen, bestonden uit gesynthetiseerde minidialogen tussen twee sprekers waarin de temporale relatie tussen en binnen elke spreker gemanipuleerd was.

Synchronisatie werd gemanipuleerd in termen van de locatie van de onset van de eerste beklemtoonde klinker (‘beat’) van de tweede spreker. De eerste spreker had altijd regelmatige intervallen tussen de onsets van beklemtoonde klinkers, waardoor er een regelmatig ritme werd gecreëerd. De eerste beklemtoonde syllabe van de tweede spreker kwam dus ‘op the beat’ van het ritme van de eerste spreker, of deze kwam ‘off beat’, d.w.z. te vroeg of te laat, ten opzichte van het ritme van de eerste spreker. *Tempomatching* werd gemanipuleerd door het spraaktempo van beide sprekers gelijk (‘match’) of ongelijk (‘mismatch’) te maken. *Overlap* werd gemanipuleerd door het begin van de spreekbeurt van de tweede spreker zodanig te plaatsen dat deze overlapte (‘overlap’) of niet overlapte (’no overlap’) met de spreekbeurt van de eerste spreker. Tenslotte werden er twee verschillende spraak*tempo*’s opgenomen om te kunnen controleren voor de wisselende hoeveelheid stilte en overlap tussen sprekers.

Predicties We verwachtten i) dat sprekers die overlap produceren als minder affiliatief geëvalueerd worden dan sprekers die geen overlap produceren; ii) dat sprekers die hun tempo niet matchen met dat van de vorige spreker als minder affiliatief geëvalueerd worden dan sprekers die hun tempo wel matchen met dat van de vorige spreker; iii) en dat, onafhankelijk van tempo en overlap, sprekers die hun spraakritme niet synchroniseren met dat van de vorige spreker als minder affiliatief geëvalueerd worden dan sprekers die hun spraakritme wel synchroniseren met het ritme van de vorige spreker.

Daarnaast verwachten we, m.b.t. Studie 3, i) dat fronsspieractivatie zal toenemen bij het luisteren naar overlappende sprekers in vergelijking met het luisteren naar niet-overlappende sprekers; ii) en dat fronsspieractivatie zal toenemen bij het luisteren naar sprekers die hun spraakritme niet synchroniseren in vergelijking met het luisteren naar sprekers die hun spraakritme wel synchroniseren.

Hierna volgen voor elke studie de kernresultaten.

Studie 1

In Studie 1 testten we in hoeverre overlap, stilte, en synchronisatie (op tijd, te vroeg, of te laat) van invloed zijn op de mate van geëvalueerde affiliatie tussen sprekers.

Zoals verwacht werden sprekers die overlap produceren als minder affiliatief geëvalueerd dan sprekers die geen overlap produceren. Daarnaast werden sprekers die hun spraakritme niet synchroniseerden met dat van de vorige spreker ook als minder affiliatief geëvalueerd. Deze laatste bevinding gold echter alleen als de tweede spreker te vroeg reageerde met betrekking tot het ritme van de eerste spreker. Was de tweede spreker te laat én produceerde deze geen overlap, dan werd deze spreker als even affiliatief geëvalueerd als een spreker die op tijd reageerde (zonder overlap). Was de tweede spreker te laat én produceerde deze wel overlap, dan werd deze spreker als affiliatiever geëvalueerd dan een spreker die op tijd reageerde (met overlap).

Deze verwarrende resultaten met betrekking tot synchronisatie zouden te maken kunnen hebben met de invloed van overlap op geëvalueerde affiliatie. Overlap is wellicht saillanter dan synchronisatie, waardoor de effecten van synchronisatie op geëvalueerde affiliatie in de overlappende dialogen niet, of in veel mindere mate, aanwezig zijn. De ‘synchronisatie’ effecten in de overlappende dialogen zou dan verklaard kunnen worden door de hoeveelheid overlap: te vroege sprekers hebben het meeste overlap (minst affiliatief), terwijl te late sprekers het minste overlap hebben (meest affiliatief). Deze lezing van de synchronisatieresultaten in de overlappende dialogen konden we wegens gebrek aan power niet uitsluiten, noch bevestigen. Om deze reden testten we onze synchronisatiehypothese opnieuw in Studie 2 (voor niet-overlappende sprekers) en in Studie 3 (zowel voor overlappende als voor niet-overlappende sprekers).

Studie 2

In Studie 2 testten we in hoeverre tempomatching van invloed is op de mate van geëvalueerde affiliatie tussen sprekers. Daarnaast testten we onze synchronisatiehypothese opnieuw, dit om onze eerdere synchronisatiebevindingen voor niet-overlappende sprekers te bekrachtigen. Omdat de aanwezigheid van overlap de effecten van synchronisatie wellicht beïnvloed, onderzochten we dit maal alleen sprekers die geen overlap produceerden.

Tegen onze verwachtingen in vonden we geen aanwijzingen voor de invloed van synchronisatie of tempomachting op geëvalueerde affiliatie. Wel werden langzame sprekers als meer affiliatief geëvalueerd dan snelle sprekers.

In deze studie werd synchronisatie, tempomatching en tempo geheel gemixt aangeboden. Dat wil zeggen dat onze deelnemers steeds verschillende dialogen te horen kregen. Bijvoorbeeld: eerst kregen ze een dialoog waarin de tempo’s snel en gematched waren, en waarin de spraakritmes gesynchroniseerd waren en vervolgens kregen de deelnemers een dialoog met twee verschillende tempo’s (mismatch), waarin de ritmes niet gesynchroniseerd waren. De volgorde van de verschillende condities was dus compleet willekeurig in deze studie.

In Studie 1 was dit echter geblokt. Daar werd synchronisatie gemixt, maar tempo en overlap geblokt aangeboden. Dat wil zeggen dat onze deelnemers in een blok bijvoorbeeld álleen overlappende en snelle dialogen hoorden, waarin synchronisatie vervolgens wel willekeurig was. Na ieder blok kregen de deelnemers een korte pauze, om vervolgens het volgende blok te doen. Bijvoorbeeld: eerst kregen ze een blok met alleen snelle, en overlappende sprekers. Binnen dat blok wordt synchronisatie wel compleet willekeurig aangeboden. Na een korte pauze beginnen de deelnemers dan met een nieuw blok, bijvoorbeeld een blok met alleen langzame en niet-overlappende spraak.

Wellicht zorgde het steeds wisselende tempo in deze studie ervoor dat de luisteraars alleen op het tempo van de sprekers letten, en niet op de subtiele temporele relatie tussen de twee sprekers, i.e., op onze synchronisatie- en tempomatching-manipulatie.

Studie 3

Om meer duidelijkheid te krijgen in de effecten van synchronisatie en overlap op geëvalueerde affiliatie testten we in Studie 3 opnieuw onze synchronisatiehypothese. Dit gebeurde zoals in Studie 1. Dit maal echter gebruikten we zowel een online als een offline maat. Naast geëvalueerde affiliatie maten we ook de fronsspieractivatie van de deelnemers tijdens het luisteren naar de fragmenten. Zoals hierboven beschreven heeft fronsspieractivatie een zeer betrouwbare correspondentie met affectieve valentie, dat wil zeggen dat fronsspieractivatie toe neemt in reactie op negatieve stimuli, en af in reactie op positieve stimuli. Door fronsspieractivatie mee te nemen hoopten we een beter beeld te krijgen van subtiele, en wellicht onbewuste, effecten van synchronisatie.

Zoals verwacht nam fronsspieractivatie inderdaad toe tijdens het luisteren naar sprekers die overlap produceerden. Ook nam, zoals in Studie 1, geëvalueerde affiliatie af in de aanwezigheid van overlap. Echter, onze predicties met betrekking tot synchronisatie en fronsspieractivatie werden niet bevestigd. Fronsspieractivatie werd niet beïnvloed door onze synchronisatiemanipulaties. Geëvalueerde affiliatie echter werd, zoals in Studie 1, wel beïnvloed door synchronisatie. Sprekers die hun spraakritme niet synchroniseerden met dat van de vorige spreker werden als minder affiliatief geëvalueerd. Deze laatste bevinding goldt echter alleen als de tweede spreker te vroeg reageerde met betrekking tot het ritme van de eerste spreker, onafhankelijk van de aanwezigheid van overlap. Was de tweede spreker echter te laat, dan werd deze spreker als even affiliatief geëvalueerd als een spreker die op tijd reageerde.

Gezamenlijke analyse

We voerden een gezamenlijke analyse uit met de affiliatie ratings van Studie 1 en 3 om onze power te vergroten en zo een beter inzicht te krijgen in de werkelijke effecten van onze temporele manipulaties. We gaan ervan uit dat deze gezamenlijke analyse een beter beeld geeft dan de afzonderlijke resultaten van de twee studies. Hieronder bespreken we de hoofdbevindingen van deze gezamenlijke analyse.

De aanwezigheid van overlap verminderde geëvalueerde affiliatie. Gecombineerd met de resultaten van fronsspieractivatie impliceert dit dat de aanwezigheid van overlap niet alleen negatieve affiliatie tussen sprekers reflecteert, maar dat de aanwezigheid van overlappende spraak ook kan leiden tot negatieve emoties (bewust of onbewust), zelfs bij luisteraars die niet meedoen aan het gesprek.

Daarnaast werden sprekers die niet synchroniseerden met het ritme van de vorige spreker als minder affiliatief geëvalueerd dan sprekers die hun spraakritme wel synchroniseerden met het ritme van de voorafgaande spreker. Deze bevinding gold echter alleen als de tweede spreker te vroeg reageerde met betrekking tot het ritme van de eerste spreker. Was de tweede spreker te laat dan werd deze spreker als even affiliatief geëvalueerd als een spreker die op tijd reageerde.

Fronsspieractivatie echter werd niet beïnvloed door synchronisatie. Dit impliceert dat synchronisatie niet direct van invloed is op het affectieve systeem van de luisteraars, i.e., synchronisatie beïnvloedt niet, zoals de aanwezigheid van overlap, de (bewuste of onbewuste) gevoelswereld van de (mee)luisteraar. Echter, synchronisatie, in het geval van de eerste beat te vroeg plaatsen gezien het ritme van de voorafgaande spreker, beïnvloed wel degelijk de affectieve *impressie* van een gesprek.

Conclusie

In dit proefschrift trachtten we te ontrafelen of specifieke temporele patronen wederzijdse verbondenheid uitdrukt. Volgens onze bevindingen is dit inderdaad het geval, zij het op een beperkte manier. We laten zien dat, zoals temporale cordinatie in het niet-verbale domein, bepaalde soorten temporale cordinatie in het *verbale* domein ook wederzijdse verbondenheid kan uitdrukken. Het maakt niet alleen uit wát je zegt, en of je de voorgaande spreker onderbreekt, soms is het ook van belang of je iets zegt op de ‘beat’ van de voorafgaande spreker. Met iemand ‘in sync’ zijn is dus niet zomaar een uitdrukking, het is echt een teken van wederzijdse verbondenheid in gesproken conversatie.

Curriculum Vitae

Anne van Leeuwen was born on May 27th, 1984 in Nijmegen, the Netherlands. After obtaining her VWO dimploma, she studied Dutch language and culture at Utrecht University. In 2009, she graduated with a Bachelor's degree, specializing in Linguistics. She went on to study Linguistics at Utrecht University, obtaining a Master's degree (cum laude) in 2011.

This dissertation is the result of work carried out between 2012 and 2017 as a PhD researcher at Utrecht University, within the NWO-funded project 'Moving the language user – Affect and perspective in discourse processing'.