Gait alignment in mobile phone conversations

Roderick Murray-Smith,[1,2]
Andrew Ramsay,[2]
[1]Hamilton Institute
NUI Maynooth, Ireland.
rod,adr@dcs.gla.ac.uk

Simon Garrod,[3]
Melissa Jackson[3]
[2]Dept. of Comp. Sci.
[3]Dept. of Psychology
University of Glasgow
Scotland
s.garrod@psy.gla.ac.uk

Bojan Musizza
Dept. Syst. & Control,
Institut Jozef Stefan
Jamova 39, 1000 Ljubljana
Slovenia
bojan.musizza@ijs.si

ABSTRACT

Conversation partners on mobile phones can align their walking gait without physical proximity or visual feedback. We investigate gait synchronization, measured by accelerometers while users converse via mobile phones. Hilbert transforms are used to infer gait phase angle, and techniques from synchronization theory are used to infer level of alignment. Experimental conditions include the use of vibrotactile feedback to make one conversation partner aware of the other's footsteps. Three modes of interaction are tested: reading a script, discussing a shared image and spontaneous conversation. The vibrotactile feedback loop on its own is sufficient to create synchronization, but there are complex interference effects when users converse spontaneously. Even without vibration crosstalk, synchronisation appeared for long periods in the spontaneous speech condition, indicating that users were aligning their walking behaviour from audible cues alone.

Author Keywords

Accelerometer, alignment, synchronization, mobile devices, rhythmic interaction, gait alignment, instrumented usability.

ACM Classification Keywords

Usability testing and evaluation, handheld devices and mobile computing, empirical methods quantitative, user studies

INTRODUCTION

Conversation is a joint activity, and people tend to use a variety of cues to subconsciously create a rapport with their conversation partners. When talking on a mobile phone with a remote conversation partner, many of these nonverbal communication channels are lost. It is, however, possible to augment the sensory capabilities of a mobile device, and communicate information via other channels. Such augmented communication allows both an increase in explicit communication and the potential to improve unconscious communication. This introduces interesting possibilities for imitation, modulation of meaning, and can be used to punctuate a converstation, entraining the timing of a discussion. The augmented information can also increase the mutual situation awareness in terms of changes to the environment a conversation partner is in, which can have positive benefits in terms of safety, as well as improving communication.

This paper investigates two major themes:

1. Exploring mechanisms for measuring the level of alignment in mobile conversation throughout a given experimental condition. We measure the synchronization of gait phase angle in walking interlocuters, and measure the gap between turns in conversation.
2. Using these measurements of alignment to evaluate the changes on the level of coupling achieved between two interlocuters. Specifically, we investigate the consequences of communicating step data sensed by accelerometers in each hand-held device with vibration feedback in the conversation partner's device.

Previous research has suggested that being in the same physical space as your conversational partner could be important for synchrony. Studies have shown that if participants are in separate rooms conversation becomes more formal – making imitation less likely [17]. The distributions of pauses and speech-overlaps in telephone and face-to-face conversations have different statistical properties [19]. We predicted that subjects would be more likely to converge to the same walking rhythm when they feel the footsteps of their interlocutor, opposed to their own. We converge when walking beside whoever we are engaged in conversation with, but does it extend to over a mobile phone?

If a relationship between alignment and conversational engagement levels can be demonstrated, then measures of alignment derived from synchronization theory [15] could be used as a proxy for a measure of quality of communication. This would provide an objective, rapidly observable measure for use in mobile scenarios, where otherwise more variable subjective approaches have to be taken, which are less straightforward to measure, and which have less temporal accuracy.

We use these methods in this experiment to try and infer whether, when engaged in conversation (but not walking in the same immediate space), interlocutors will converge in walking rhythm when speaking via a mobile phone while walking. This could further highlight that cooperative conversation could be driving interpersonal synchronization. Furthermore, subjects will be using a PDA during the experiment. In one condition they will feel vibrations on this from their own footsteps and for the other condition they will feel the vibrations of their conversational partner's footsteps.

BACKGROUND LITERATURE

Alignment

Conversation is a joint activity, so interlocutors involved must work together to try and facilitate understanding and the flow of conversation. Working together towards understanding is known as 'interactive alignment'. Conversational partners use the same words, sounds and facial expressions in order to become aligned. The process involved to achieve this goes as follows: speakers use the same representations as one another, so if speaker 1 describes something in a particular way speaker 2 will have speaker 1's representation primed so will also use it, thus, speaker 1 and speaker 2 are now aligned. This ensures conversational partners are operating at the same level, so dialogue will be fluid [8].

Synchronization theory

Since first being observed by Christiaan Huygens in 1673, the study of synchronization has been applied in physical, physiological and social situations. A standard introduction to the field, [15], describes synchronization as *"an adjustment of rhythms of oscillating objects due to their weak interaction"*. Other examples of synchronization effects include the rhythmically synchronized flashing of fireflies and the menstrual synchrony between women in close social contact. The suggestion that interlocutors' breathing patterns converge when they are engaged in certain tasks is made by McFarland in [13], conversational partners' breathing patterns were monitored during: quiet breathing, reading aloud, spontaneous monologue, scripted dialogue and spontaneous conversation. It was hoped the study might provide insight into aspects of conversational exchange and interactional synchrony between interlocutors. The results showed some selected, informal examples of synchronization of breathing patterns of interlocutors at turn-taking boundaries and simultaneous vocalizations e.g. laughter. Breathing patterns during listening differ considerably from quiet breathing – instead, respiratory patterns during listening resemble breathing during speech, and speaker and listener converged to the same breathing pattern. We applied synchronization theory in HCI contexts in earlier work, including analysis of tapping accuracy while walking, in [5]. It can also provide a framework for analysis and design of rhythmic gesture recognition systems, as developed in [12].

Synchronization in Cooperative Conversation

It has been suggested that interactional synchrony serves as a coordination device, which helps us to achieve mutual or complementary goals [2]. The fact that conversation is a joint activity could account for the imitation between interlocutors. Synchronous activity between conversational partners has also been observed in several further ways, as speakers have been found to converge in speaking rate , vocal intensity and the amount they pause. Listeners will imitate the posture of the speaker if they find them engaging [7]. Furthermore, listeners have been found to move in time with the speaker's speech [14]. The coordinating influence of tempo in music and speech is discussed in [11]. Wilson and Wilson [20] propose that endogenous oscillators in the speaker and listener become mutually entrained on the basis of the speaker's rate of syllable production.

Shockley, Santana and Fowler [18], investigated whether interlocutors mimic one another's posture when engaged in cooperative conversation. They varied whether the conversation was cooperative or not. In the cooperation condition, verbal communication was required and the participants were either facing one another, or facing in the opposite direction. The results showed that verbal communication alone was enough to lead to convergence of the interlocutor's postural movements. Even when participants were not facing one another imitation happens due to the cooperative nature of task. This has important implications for the cooperative nature of conversation. Giles [9] found that listeners were more likely to imitate speakers if they were in a cooperative environment, rather than a non-cooperative environment. This imitation between interlocutors is a nonconscious process. This was highlighted by [1] in their paper investigating the Chameleon Effect for facial expressions, where a person changes their behaviour to suit their environment. Perceiving a behaviour performed by someone in your environment will subconsciously make you more likely to also perform that behaviour, e.g. walking beside our conversational partner in perfect synchrony of footsteps. Such behavioural synchrony suggests that we automatically adapt to fit the social environment we are in, imitating whoever we are interacting with at the time. Participants were more likely to report increased liking with confederates who imitated their behaviour, suggesting it did aid the flow of the conversation and allow a rapport to develop.

HARDWARE AND SOFTWARE

The inertial sensing equipment used in the experiment consists of an HP iPAQ 5550 equipped with a MESH [10] inertial navigation system (INS) backpack consisting of 3 Analog Devices ±2g dual-axis ADXL202JE accelerometers, 3 Analog Devices ±300deg/s Single chip gyroscopes, 3 Honeywell HMC1053 magnetometers and a vibrotactile transducer, used for feedback purposes. The system can also record location – its GPS unit is a Trimble Lassen Sq module for mobile devices, and is also built-in as part of MESH (see figure 1). This module provides us with a 9m resolution with up to 6m resolution around 50% of the time it is used. It also provides us with velocity resolution of 0.06m/s and an 18m altitude resolution. Data were sampled at 90Hz.

ALGORITHMS

Step detection

The step detection algorithm used in this study examines the vertical oscillation of the device. While walking, one vertical oscillation occurs for each step (one sinusoid per step). In order to extract information from this oscillation, the raw acceleration data must be processed. The algorithm first ensures the oscillation is centred around zero by subtracting

Figure 1. Left: MESH device alone and attached to an HP5550 Pocket PC. Right: The MESH circuit board showing the main components.

the mean of the previous one second's worth of data from all incoming values. This ensures that the system is robust to changes in user posture during the study. These zero-centred oscillatory data are then low-pass filtered to produce a smooth sinusoidal wave. In this procedure care was taken to avoid any introduction of phase lags, by time reversing the filtered signal and sending it through the same filter again. The step detection algorithm then extracts peaks and troughs from this filtered signal. A step may be detected only when a peak in this signal is detected. However, as not all oscillations will be due to the user walking, further conditions must be met before a step is detected. The detected peak and previous trough must occur within a certain time of each other (> 0.1s and < 0.7s) and must be above a certain magnitude of signal before the oscillation is detected as a step. These timing and signal magnitude thresholds were set through an iterative trial and error process.

Synchronization detection

How do we detect synchronization? The oscillations involved are often irregular, ruling out simple strategies. In some cases, such as respiratory examples, or electrocardiogram data, there are clear marked events with pronounced peaks in the time-series which can be manually annotated, or automatically detected. One practical advantage of the use of synchronization theory is that often we have a quite complex nonlinear oscillation, which might be sensed via a large number of sensors. The phase angle ϕ of that oscillation is however a simple scalar value, so if we are investigating the synchronization effects in two complex systems, the analysis can sometimes be a single value, the relative phase angle $\phi_2 - \phi_1$.

The Hilbert Transform

How do we find the phase angle from the data? A common approach is to use the Hilbert transform introduced by Gabor in 1946, which gives the *instantaneous phase and amplitude* of a signal $s(t)$ [15]. The Hilbert transform signal $s_H(t)$ allows you to construct the complex signal

$$\zeta(t) = s(t) + is_H(t) = A(t)e^{i\phi(t)} \tag{1}$$

where $\phi(t)$ is the phase at time t, and $A(t)$ is the amplitude of the signal at time t. The Hilbert transform of $s(t)$ is

$$s_H(t) = \frac{1}{\pi} \lim_{T\to\infty} \int_{-T}^{T} \frac{s(\tau)}{t-\tau} d\tau \tag{2}$$

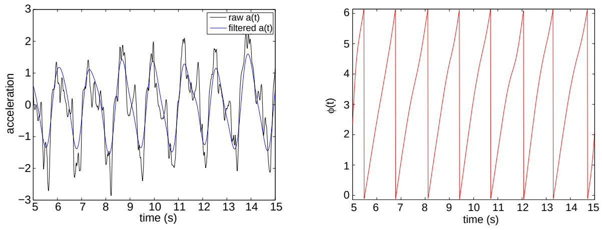

Figure 2. Generating the phase angle $\phi(t)$ from observed acceleration data $a(t)$ from a user walking

Although $A(t)$ and $\phi(t)$ can be computed for an arbitrary $s(t)$ they are only physically meaningful if $s(t)$ is a narrow-band signal. For the gait analysis, we therefore filter the data to create a signal with a single main peak in the frequency spectrum around the typical walking pace (below 4Hz).

Synchrograms and synchronisation index

In order to visualise the changes in relative phase in the experimental data, we use the standard tool of a *synchrogram*. This is a stroboscopic technique, where we plot a point at the value of phase $\psi(t_k)$ in the first oscillator, $\psi(t_k) = (\phi_1(t_k) \mod 2\pi)$, when the second one completes a cycle.

Note that synchronization does not need to be a 1:1 synchronisation. We can also detect $n : m$ synchronisation by viewing sychrograms. If there is visible structure to the points in the synchrogram then that suggests synchronization is occurring. In order to generate a synchronization index for a particular synchronization ratio $n : m$, we can use a local sum over the last M_l cycles:

$$\Lambda_l = \left| \frac{\sum_k^{M_l} e^{\imath\eta_{k,l}}}{M_l} \right| \tag{3}$$

where $\eta_{k,l} = \phi_2 \mod 2\pi n|_{\phi_1 \bmod 2\pi m = \theta_l}$, where θ_l is the 'strobe' point on the interval [0, $2\pi m$] at which the ϕ_2 is recorded. Finally, we average over all N points θ_l to obtain a synchronization index: $\lambda_{n,m} = \frac{1}{N}\sum_l^N \Lambda_l$, which varies from 0 to 1, where 0 indicates no synchronization at the $n : m$ ratio, and 1 is full synchronization

Example of synchronization via footstep feedback

To illustrate the basic synchronization detection, we illustrate detection of synchronization in the situation where two people are walking and talking to each other and can feel each other's footsteps via vibration feedback. The participants do not talk to each other, but have 4 stages: 3 minutes of feeling their partner's footsteps, 3 minutes of their own footsteps, 3 minutes of their partner's, and 3 minutes of their own.

The synchrogram and $\lambda(t)$ for this data is shown in Figure 3. The crosstalk between users was active in minutes 1-3, and 6-9. In the first 3 minutes only sporadic bursts of synchronisation appear, possibly due to the participants familiarising themselves with the experiment, and walking somewhat uneasily. In period 6-9, however, there is strong evidence of

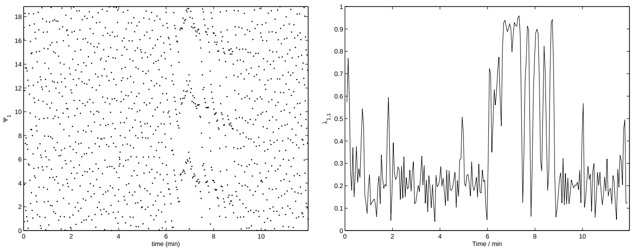

Figure 3. The left figure is the synchrogram of two time-series of acceleration data from participants walking while holding a PocketPC. The right plot shows $\lambda(t)$ values summarising the level of synchronisation at each time point.

synchronized walking behaviour, as can be seen in the structured form of the synchrogram, and the high level of $\lambda(t)$. Note how rapidly synchronization is lost at 9 minutes, when the vibration is continued, but linked to the users' own footsteps, rather than their partners'. This initial dataset already shows that remote participants who cannot see each other can synchronise their gait.

EXPERIMENT

Method

Participants

22 undergraduate or graduate students (11 pairs). The walking rhythm data from 1 pair was not recorded, so the speech data from this trial was not included in the analysis, leaving 10 pairs. The pairs of subjects had never met one another before and were either both male or both female. There were no mixed pairs. Each participant provided informed consent and had the opportunity to withdraw from the experiment at any time. They all participated on a voluntary basis.

Apparatus

Each participant's speech was recorded and saved via a laptop computer, which was held in a rucksack and carried by each participant during the experiment. The also wore headphones with a microphone attached, to actually record the speech. There were two mobile phones (Nokia 6680s), one for each participant. Users were holding the PDA (HP Ipaq 5550) which provided vibration feedback, logged the acceleration data and showed experimental instructions.

The PDAs and laptops communicated via an *ad hoc* wireless network (i.e. a network between devices with no access point involved). This meant that communication between devices was direct, rather than going through an access point. Also we used UDP to send and receive the footstep events between each PDA rather than TCP. UDP involves less overhead than TCP since it has no need to set up a connection, acknowledge packets, and handle errors. We did check for missed packets at application level, but found almost no packets were dropped anyway, expected behaviour when all devices are in a relatively small area.

Using the *ad hoc* network and UDP resulted in an improvement in reduction of Round-trip-time (RTT) latency and variability. The bulk of the RTTs tended to fall in the 10-30ms range, meaning one way trip times of 5-15ms, and the few outliers were all under 100ms. The mean round-trip-time for the vibration communication of each partner's footsteps was less than 20ms. For conversation, each participant was provided with a Nokia LDW1 Bluetooth headset, which allowed them to carry the mobile phone in their pockets, out of the way.

Line: 1-2/12

A customer enters a pet shop.

'Ello, I wish to register a complaint' .... 'Ello, Miss?'

Figure 4. Screen shots of the PocketPC display during the experiment. Left: script. Right: image.

Materials

The PDAs held the script, which was a Monty Python sketch (the "Dead Parrot" sketch) and the cartoon images. The PDA has a 3.8" TFT active matrix screen, with a resolution of 240 × 320. There were 48 images – participants were instructed to work through them at their own pace, so they may only have described a few of the 48 images in the 5 minutes provided. All 48 were Gary Larson cartoons.

Experiment

The aim of the experiment is to determine whether synchronisation of utterances and walking rhythms will happen between interlocutors while conversing via mobile phones. Dyads are walking with the device throughout the experiment. The task involves each pair reading a Monty Python script (the 'Dead Parrot' sketch), describing a series of cartoon pictures and holding spontaneous conversations with their interlocutor – all via the mobile phone. The different tasks vary in their degree of cooperation needed between interlocutors. As Fowler *et al.* [6] suggests, a task that involves more cooperation between speakers should produce more synchrony in their speech and gait behaviour. For half of the experiment they could feel their own footsteps, and for the other half they could feel their partner's, displayed as vibrations on their PDA.

Procedure

Each pair of participants completed the three different tasks, which individually lasted 5 minutes. They also had to complete each of the three tasks twice; as once they felt the vibrations from their own walking rhythm (no crosstalk) and the second time they felt the vibrations from their partner's footsteps (with crosstalk). The pairs of participants were given a 5-minute lesson on how to use the PDA and then they took part in a demonstration of the experiment, to make them more comfortable with the device and to view what the experiment would entail.

They put on the rucksacks, put in the Bluetooth earpiece and put the mobile phone device in their pockets. The actual experiment took place in a park area, just outside the university building. This allowed behavioural synchrony and interpersonal conversation to be monitored in a more natural setting. Participants were instructed to circle the park at opposite ends, so they would only be able to hear one another via the mobile phone. Each of the tasks lasted 5-minutes and all six were presented by the PDA in an order determined by the balanced experiment design.

During the spontaneous speech task, participants could speak freely to one another on any topic of their choice. The PDA was timed to end the task after 5-minutes, but participants had to occasionally monitor the screen, as they had to watch for the end of the 5-minutes. During the scripted task, one participant was given the role of speaker 1 and the other participant was named speaker 2. This remained the same when the task was repeated, so both participants read the same part in the script twice. Finally, when describing the images, participants were viewing the same images at the same time; they simply had to briefly describe the cartoons to one another before moving onto the next one.

Data analysis

The speech (separate channels for each speaker) and acceleration data from each person were stored separately.

Speech data

The conversation during the spontaneous speech task was annotated to analyse whether interlocutors' speech patterns converged. The spontaneous speech task was analysed, as this is where subjects will have the greatest opportunity to develop a rapport with their conversational partner - thus, speech will be most natural. We used Sound Edit, a sound wave analysis package, to mark when each speaker started and stopped speaking. Feedback was marked separately from speech, i.e. every time the listener provided feedback to the speaker's speech it was to be coded 'feedback', rather than 'speech'. For example, if the speaker was telling a story and the listener offered agreement by saying 'yes', or 'uhuh' etc., this would be coded as 'feedback'. This would allow us to explore whether pairs converged to the same speaking patterns, e.g. similar amount of feedback given by both interlocutors and similar length of speech in each turn. This provides observations allowing investigation of whether there was synchrony in speech patterns, as well as analysing synchronization of gait. The expectation was that there should be more imitation in speech patterns during the crosstalk condition, as participants should be more connected from feeling their conversational partner's footsteps. The means were calculated for length of feedback and length of speech turn, both in the presence and absence of crosstalk.

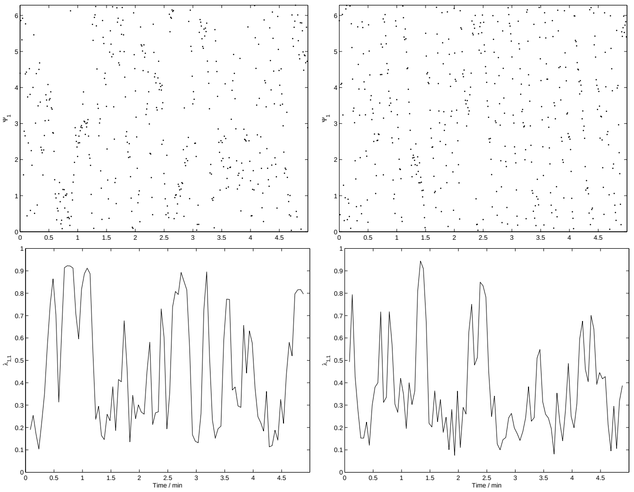

Figure 5. The top figures are the synchrograms from Pair 4 reading the script while walking. The lower plots shows $\lambda(t)$ values summarising the level of synchronisation at each time point. Left plots are crosstalk condition, Right plots are no crosstalk. The crosstalk case shows longer periods of synchronisation than the no-crosstalk case.

The spontaneous speech task allowed interlocutors to speak freely on any topic they desired. Conditions of the task were natural, giving dyads the opportunity to develop a rapport. For this reason, as mentioned, this is the task that was explored for the convergence of speech patterns. Pairs 1, 4, 9 and 10 were selected for analysis, as there were technical difficulties with recorded speech from pairs 2-8, limiting the usefulness, as one participant's speech in each of these trials was inaudible.

Average speech length in a turn was very similar between some pairs – usually an indicator that speech was interactive. For example, mean speech length for pair 9, speaker 1, was 3.64 seconds and for speaker 2 average speech length was, 3.93 seconds. Further, the average length of feedback given also highlights a convergence of speech patterns: speaker 1 was 0.87 seconds and speaker 2 was 0.97 seconds. This was during the crosstalk condition – when interlocutors could feel the vibrations of their partner's footsteps. This was not the case in every pair that was annotated, as pair 10 converged to a more similar speech length in the absence of crosstalk. Average speech length for speaker 1 was 3.66 and for speaker 2 it was 4.56. During the crosstalk condition, pair 10 seem to have diverged from the same speech patterns, as average speech length in a turn was, 1.96 seconds for speaker 1 and 4.80 seconds for speaker 2. As speech was lost from one subject in pair 4, data here could not be annotated in the same way. This is unfortunate, as this was where most synchronization of gait was observed.

Gait synchronization

All three different communication tasks: scripted, images and spontaneous/free speech were predicted to produce different levels of synchronization of gait. Free speech, as discussed, was predicted to elicit the most convergence of walking rhythms, as this task yielded very natural conversation.

In many cases, conversation between dyads was very enthusiastic and after the experiment had ended, participants often resumed telling a story they had begun during the 5-minute free speech task, but had not had the opportunity to finish.

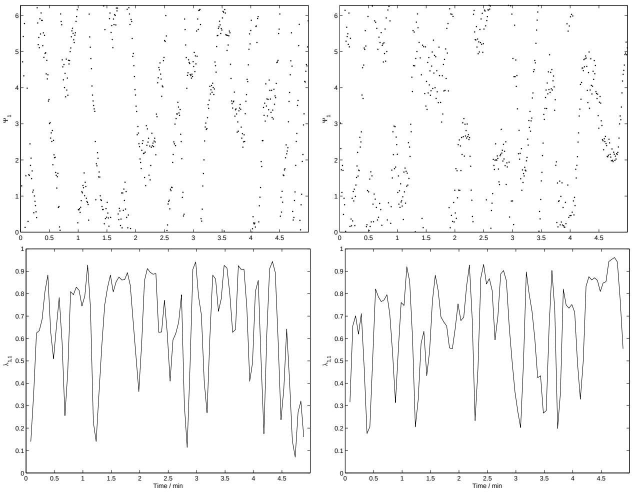

Figure 6. The top figures are the synchrograms from Pair 4 conversing spontaneously while walking. The lower plot shows $\lambda(t)$ values summarising the level of synchronisation at each time point. Left plots are crosstalk condition, Right plots are no crosstalk. In both cases there are significant periods of synchronisation in both cases, as shown by the structured form of the synchrogram, and the high levels of $\lambda(t)$

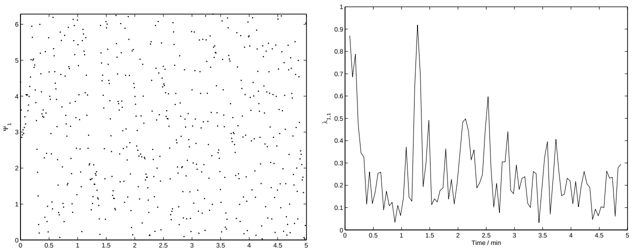

Figure 7. The left figure is the synchrogram from Pair 10 reading the script, while walking in the no-crosstalk condition. The right plot shows $\lambda(t)$ values summarising the level of synchronisation at each time point. Note the negligible amount of synchronisation in this case.

We generated synchrograms for each time-series, and generated histograms for $\lambda_{n:m}$ values for each minute of the 5 minute long experiment conditions. This gave us an indication of the level of alignment in gait during that period. Level of synchronization goes from 0-1, $\lambda(t) \geq 0.5$ was defined for the purposes of this analysis to be synchronous behaviour.

Investigating the results from Pair 4, as shown in Figure 6, we see a sudden increase in synchrony around 1.25-1.75 minutes. Here, the conversation was highly interactive and a lot of feedback was given. There was another rise in synchronous behaviour between 2.75 and 3.20 minutes. At this point, interlocutors were returning to a subject they had discussed in the previous free speech task, in the presence of crosstalk. The rise in synchronous activity around 4.20-4.30 minutes was associated with highly interactive conversation. For example, the listener seemed to interrupt the speaker several times to add a point to the conversation or to ask a question.

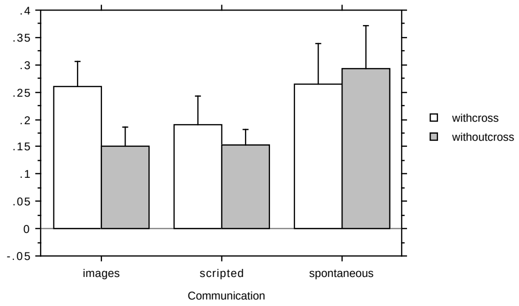

Figure 8. The interaction bar chart highlights the average proportion of time gait was synchronised in each communication task, in the presence and absence of crosstalk, with standard error bars ($\sigma/\sqrt{n}$). Free speech yields most synchronisation ($\lambda(t) > 0.5$ 29% of time).

Did the presence of crosstalk lead to more synchronization of gait as predicted (i.e. when interlocutors could feel vibrations of their partners footsteps via the PDA), and was there a task effect, i.e. which of the tasks – scripted, images or spontaneous speech – would elicit the greatest amount of synchrony? In order to analyse when most synchronization of gait was happening, a 2-way repeated measures analysis of variance (ANOVA) was performed on the communication tasks and crosstalk – these results are shown in Figure 8. The ANOVA shows no main effects. The combination of different tasks and the presence or absence of crosstalk led to varying degrees of synchronization, so in some conditions more synchronization of gait was observed than in others. The spontaneous speech task, without crosstalk, produced the most synchrony of footsteps between conversational partners (mean = 0.294). In the absence of crosstalk, there was a significant difference observed between the images task and the spontaneous speech, $F(1, 18) = 7.7, p < 0.05$ and between the scripted task and the spontaneous speech, $F(1, 18) = 7.3, p < 0.05$. The spontaneous speech task, in the presence of crosstalk, produced the second greatest amount of synchronous behaviour (mean = 0.265). There was no significant difference between the spontaneous speech task and the scripted task in the crosstalk case.

Although the data generally supports the prediction that during the free speech task the most synchronous behaviour would be observed, it does not support the hypothesis that feeling your partners' footsteps, *while in conversation*, would lead to an increase in synchronization of gait, as more syncronization was observed in the absence of crosstalk during free speech. The third most synchronous condition was the image task in the presence of crosstalk (mean = 0.260), higher than the images task in the absence of crosstalk. In the scripted task, marginally more synchronous activity was also observed in the presence of crosstalk (see Figure 5).

DISCUSSION OF RESULTS

Shockley *et al.* [18] suggested that it was speech that was responsible for coordinating participants. They stated this was, *"evidence of an effect of cooperative speaking on movement entrainment"*. Our findings are compatible with this, as dyads were not in face-to-face conversation, yet alignment of walking rhythm was still evident, even in no-crosstalk cases.

The least synchronization of gait was observed in the image task, when crosstalk was absent (mean = 0.150), with scripted slightly more synchronized (mean = 1.54). The small level of difference between the scripted task and the image task was somewhat unexpected, as it was predicted that as the script condition entailed interlocutors working together to complete the task, more synchrony should have been evident than for the image task. One possible explanation for the higher than expected levels of sync for the images was that some participants were almost treating this task like free speech. For example, they would engage in joint laughter and make jokes regarding the images, potentially explaining this result. This more free-flowing conversation occurred occasionally in other cases as well, when technical problems with the communication arose – typically when they had trouble hearing via their Bluetooth headsets.

It is possible that feeling the vibrations of your partner's footsteps leads to increased synchronization in these tasks, as participants were concentrating on the task at hand, therefore the vibrations could have been subconsciously entraining them. Conversely, during the free speech task the participants were free to interpret this as they wished, as far as task instructions are concerned. It is possible that the vibrations were more distracting, as the participants are no longer focused on completing a task. In a task involving rhythmic limb movements and conversation, Richardson *et al.* [16] noted *"conversational behaviour may undermine the stability of the unintentional synchrony brought about by vision"*.

An alternative suggestion for the synchronization of footsteps observed, is due to the rhythmical nature of speech i.e. dyads will converge during their conversation to the same rhythmical speech patterns and this will in turn lead to entrained behavioural rhythms. This seems like a viable explanation, as [4] found that interlocutors become synchronized to their conversational partner's speaking rhythms e.g., in turn taking, the new speaker picks up on the beat of the last speaker's rhythm. Once a rhythm is created this can potentially be entrained to other rhythms, which could account for synchronization of gait observed in this study. Interlocutors breathing patterns could have become entrained, thus entraining partner's gait i.e., rhythmically coincident breathing patterns could in turn be causing synchronization of footsteps. The least synchronization of gait was observed in the image task, when crosstalk was absent (mean = 0.150), with scripted slightly more synchronized (mean = 0.154). It was predicted that the least sync would happen during the image task in the absence of crosstalk. The scripted task was predicted to produce more synchrony, as the completion of the task depends on dyads working together. The marginal difference between the image and scripted cases may be due to participant appearing to treat the image task more like free conversation. This occurred in other cases as well, when technical problems with the communication arose – typically when they had trouble hearing via their Bluetooth headsets.

More detailed analysis of the speech in some of the pairs supported the theory that speech and behavioural rhythms entrain, as a result of a rapport having developed between interlocutors. Pair 4 showed the greatest level of synchronization of gait during the free speech task. Exploration of the conversation here yielded some interesting findings on what could be driving synchrony. Firstly, interlocutors were discussing shared interests and experiences when increased levels of synchrony occurred. Also sync increased when subjects were discussing something they had previously discovered about one another, in the spontaneous conversation task in the presence of crosstalk. Pair 4 seemed to find one another engaging and conversation was highly interactive, compared to other pairs where the speech was annotated ([7] noted that we are more likely to imitate if we find our partner engaging). The interlocutors had already completed a 5-minute free speech conversation (with crosstalk), therefore they may have begun to create a mutual rhythm during this task, which was fully developed and strengthened by their next free speech trial (without crosstalk). In pair 4 sync increased when the listener interrupted the speaker to add something to the conversation e.g., feedback or asking a question. This has implications for the interactive nature of conversation, i.e. the more interactive; the more we will synchronize with our conversational partner.

Because the WLAN range of 100m constrained the distance apart during the experiments, dyads were instructed to walk in a circle, on grass, at opposite ends of a park to ensure they could only hear one another via the mobile phone device. However, there is the possibility that the proximity could have had an effect on walking behaviour, although most participants said that they had not been aware of the walking behaviour of the other person, and did not appear to be observing each other. Poorer synchronization in the scripted and image cases might be due to the interaction between visual attention to the PDA screen and gait.

CONCLUSIONS

This study is an initial exploration of synchronization in mobile settings, which has provided objective data supporting the surprising observation that it is possible for remote mobile participants to synchronize their walking behaviour based on the voice channel alone. On the methodological side, this paper has introduced novel techniques to the HCI and speech communities, which can help measure the consequences of augmenting voice communication with other modalities. We found that in speech-free cases it *was* possible to create aligned walking behaviour just by feeding back the conversation partner's footsteps to the vibrator in the pocketPC, but that when combined with speech input, the effect on synchronization was less clear, suggesting the possibility of interference among oscillators. Even more interesting perhaps, though, was the observation of synchronization in the speech-only cases, *without any vibration crosstalk*, and without colocation of the participants.

Implications for understanding conversation

The results presented here are interesting exploratory results in a novel style of experiment. Until now, linking the analysis of walking behaviour in a realistic setting with alignment and engagement levels in conversation would have provided a significant challenge to speech researchers. Observational procedures were used, such as hand scoring videotapes of listeners movements and speech [3]. This is time-consuming, error-prone and not open to online experimental control. Recent rapid developments in mobile device capacity, and compact inertial sensing, coupled with the use of the tools from synchronization theory, have opened up a new way of investigating alignment effects in conversation, and their effect on quality of communication. The method used in this study to code rhythmic behaviour and imitation is automatic, robust, as the accelerometer monitors walking patterns throughout the experiment, and can be used to control experimental stimuli, providing a more stringent method of exploring synchronization during interpersonal interaction.

Implications for technology

This paper is not about the particular application of feeling each others' footsteps – it is an example of a way of assessing the consequences of any augmentation of the voice communication between two mobile users. Various augmentations of basic communication technologies can be imagined. In this paper, an augmented mapping from acceleration signals to vibration was used, but a great variety of mappings between sensors and displays of the interlocuters could be imagined. The alignment approach, with tools from synchronization theory, provides an objective technique which gives insight into the effects of a new communication configuration, without requiring subjective feedback from users.

Predictions/Future work

This paper represents the first stage of a research programme, providing an initial exploration of whether interlocutors align. The next stage is to characterise the link between alignment and measures of rapport empirically. This would provide the mobile HCI community with new tools to assess such systems, acting as a proxy for subjective measures such as quality of communication or level of engagement. Is the synchronisation an indicator of successful, engaging conversation, or can we improve the quality of the conversation by increasing the level of synchronisation via other channels? Can this contribute to a better understanding of why mobile phone users often pace back and forth while talking on the phone? The interaction between motion and conversational behaviour in mobile communication remains a relatively unexplored area. With instrumented devices, we could monitor user movements while on the phone, and relate activity in one person to activity in the conversation partner. The measures of alignment could potentially predict the expected remaining duration of the call, and the quality of the communication at each moment in time. Commercially available phones such as the Nokia 5500 include accelerometers, so it is now straightforward to log user behaviour in normal phone use, while walking in a natural setting. This allows us now to look at correlations in physical movements, and link them to the quality and duration of the communication, and investigate the consequences of technology or interface mediated interruptions on walking and communication behaviour.

ACKNOWLEDGEMENTS

RM-S is grateful for the support of SFI grants 00/PI.1/C067, 00/RFP06/CMS052, EC funded project OpenInterface, and the HEA funded *Body Space project*, and a Nokia donation of equipment and funding. RM-S & BM are grateful for support of EPSRC grants GR/R98105/01, EP/E042740/1, and RM-S & SB BRG project *Continuous Gestural Interaction with Mobile devices*.

REFERENCES

1. Chartrand, T. L., and Bargh, J. A. The chameleon effect: The perception-behaviour link and social interaction. *Journal of Personality and Social Psychology 76* (1999), 893–910.
2. Clark, H. H. *Using language*. New York: Cambridge University Press., 1996.
3. Condon, W., and Ogston, W. Speech and body motion synchrony of speaker-hearer*The perception of language*, D. Horton and J. Jenkins, Eds. Columbus, OH: Charles E. Merrill, 1971, 150–184.
4. Couper-Kuhlen, E. *English speech rhythms*. Amsterdam: John Benjamins, 1993.
5. Crossan, A., Murray-Smith, R., Brewster, S., Kelly, J., and Musizza, B. Gait phase effects in mobile interaction. *ACM SIG CHI 2005, Portland*. 1312–1315.
6. Fowler, C, A., Brown, J, M., Sabadini, L., and Weihing, J. Rapid access to speech gestures in perception: Evidence from choice and simple response time tasks. *Journal of Memory and Language 49*, 3 (2003), 396–413.
7. France, M. L. Posture mirroring and rapport*Interaction Rhythms: Periodicity in communicative behaviour*, M. Davis, Ed. New York: Human Sciences Press, 1982, 1–68.
8. Garrod, S., and Pickering, M. J. Why is conversation so easy? *Trends in Cognitive Sciences 8*, 1 (2004), 8–11.
9. Giles, H. Accent mobility: A model and some data. *Anthropological Linguistics 15* (1973), 87–105.
10. Hughes, S., Oakley, I., and O'Modhrain, S. MESH: Supporting mobile multi-modal interfaces. *ACM UIST'04*. ACM (2004).
11. Jungers, M. K., Palmer, C., and Speer, S. R. Time after time: The coordinating influence of tempo in music and speech. *Cognitive Processing 1-2* (2002), 21–35.
12. Lantz, V., and Murray-Smith, R. Rhythmic interaction with a mobile device. *NordiCHI '04, Tampere, Finland*. ACM (2004), 97–100.
13. McFarland, D. H. Respiratory markers of conversational interaction. *Journal of Hearing and Language Research 44* (2001), 128–143.
14. Newtson, D. The perception and coupling of behaviour waves*Dynamical systems in social psychology*, R. Vallacher and A. Nowak, Eds. San Diego, CA: Academic Press, 1994, 139–167.
15. Pikovsky, A., Rosenblum, M., and Kurths, J. *Synchronization: A universal concept in nonlinear sciences*. Cambridge University Press, 2001.
16. Richardson, M. J., Marsh, K. L., and Schmidt, R. C. Effects of visual and verbal interaction on unintentional interpersonal coordination. *Journal of Experimental Psychology: Human Perception and Performance 31*, 1 (2005), 62–79.
17. Sellen, A. J. Remote conversations: The effects of mediating talk with technology. *Human-Computer Interaction 10* (1995), 401–444.
18. Shockley, K., Santana, M.-V., and Fowler, C. A. Mutual interpersonal postural constraints are involved in cooperative conversation. *Journal of Experimental Psychology: Human Perception and Performance 29*, 2 (2003), 326–332.
19. ten Bosch, L., Oostdijk, N., and de Ruiter, J. P. Durational aspects of turn-taking in spontaneous face-to-face and telephone dialogues. *TSD 2004*. Springer-Verlag (2004), K. P. P. Sojka, I. Kopecek, Ed., LNAI 3206, 563–570.
20. Wilson, M., and Wilson, T. P. An oscillator model of the timing of turn-taking. *Psychonomic Bulletin & Review 12*, 6 (2005), 957–968.