# Spontaneous synchronisation between repetitive speech and rhythmic gesture

*Gregory Zelic, Jeesun Kim, Christopher Davis*

The MARCS Institute, University of Western Sydney, Sydney, NSW, Australia

G.Zelic@uws.edu.au, J.Kim@uws.edu.au, Chris.Davis@uws.edu.au

## Abstract

Although studies have described how motion in diverse biological systems may spontaneously synchronize it is not known whether speech and gesture exhibit such a property. Previous research on the coordination of speech and gesture has focused on pointing or tapping tasks, the structure of which may regulate speech and gesture dynamics. Here we examined whether synchronies might arise between a repetitive utterance and rhythmic finger movement oscillations in a non-intentional paradigm. Participants were instructed to repeatedly utter /ba/ or /sa/ syllables with/without vocalizing, while continuously moving their right index finger in flexion/extension. No instructions about synchronization were given; participants were only told to adopt the most comfortable motions. We expected that the larger amplitude of face motion for /ba/ syllables and vocalized speech would lead to greater influence on the gesture. In contrast, the results showed more synchronization for /sa/ and when syllables were articulated silently. Less perceptive feedback may lead to a reduction in the robustness of the speech component, making it more susceptible to gesture influence.

**Index Terms**: non-intentional synchronization, speech, gesture

## 1. Introduction

Many biological systems in nature, from fireflies to neuronal networks, spontaneously synchronise in space and/or in time [1], [2]. Human behaviour can also exhibit non-intentional synchrony (e.g., when clapping [3], walking, or just standing face to face [4]) and people adopt preferential coordination patterns (e.g., in bimanual coordination tasks [5], [6]; in postural tasks [7], [8]).

Such synchronies have been understood by considering coupled non-linear oscillators, and strongly depend upon the properties of each oscillator and of the coupling established between them [9], [10]. That is, synchronisation emerges from the specific interactions (the coupling) spontaneously established between the components of the system in response to the constraints of the environment and with regards to their own properties/capacities. The nature of this coupling can be based on the continuous perception of cues, e.g., visual and auditory ones when people applause, tactile and kinesthetic cues when walking hand in hand, as well as on neurophysiological determinants when intrapersonal coordination patterns emerge between limbs.

In the present study we sought to determine whether and how such non-intentional and spontaneous synchronisation might emerge between speech and gesture. Before considering any unconstrained coordination between speech and manual gesture, it should be pointed out that there is considerable evidence for a mutual influence between speech and beat gestures that occurs when modulating the intensity/amplitude of performances [11-16] as well as a strong anchoring of deixis in the pointing gesture, especially with the indicative aspect of the gesture [17-19].

However, in these cases there seems to be an underlying constraint that likely would have established cooperation between speech and beat or deictic gesture, i.e., in the case of speech and beat gestures both have a frequency dependency and on in the case of deixis, there is a common functionality, i.e., showing/pointing to something,

In the current study, the issue was whether synchronisation between speech and gesture would emerge spontaneous even when there was no constraint shaping cooperation. This is an important issue since skilled learning (in this case, speech articulation) occurs with respect to a background of pre-existing coordination tendencies and so it is instructive to examine the degree to which skilled motion displays coupling with other motion in a situation when both can run freely. This issue has also potential practical consequences since in many of our daily activities and the non-intentional synchronisation of speech and gesture could lead to undesirable side effects. For example, it has been suggested that speaking on a hands-free mobile phone degrades driving performance due to an unintentional influence from speech to gesture performance [20].

In this regard a number of questions arise. For example, what is the strength of the coupling, interactions, or influences between speech and gesture? Are such influences dependent on the nature of the motor and/or the speech task?

One property of speech is that articulators as the tongue, the jaw or the lips are differently recruited and combined according to the syllable produced [21]. Such differences in speech production may occasion different perceptive feedback depending on which syllable is produced. In the present experiment, we tested whether and how more or less sensory information feedback about the speech can infer a simultaneously performed motor task. We expected that more perceptive information received about the speech would lead to greater influences from the speech to the gesture performance, and be reflected in a stronger coupling (synchronisation) between them.

## 2. Method

### 2.1. Participants

Nine right-handed volunteers took part in the experiment (6 men, 3 women, 29.89 ± 5.01 yrs) and were tested in accordance with the Declaration of Helsinki, and approved by the Ethics committee of the University of Western Sydney.

### 2.2. Procedure and experimental conditions

Participants had to repeatedly utter a /CV/ syllable while continuously executing oscillatory movement, about the metacarpophalangeal joint, with their right index finger.

Participants were instructed to perform both tasks in whatever way was the most comfortable and to allow any changes in behaviour that y may feel coming (a “do not intervene” paradigm; [6], [9]). Finally, we asked them to close their eyes during trials to avoid any effects due to visual processing.

We manipulated the amount of perceptive feedback received from the speech following a 2 X 2 factorial design. Participants had to utter /ba/ or /sa/ syllables with the assumption that the former will mobilize more face motion (jaw and lips), that is to say will generate more sensorimotor feedback from the speech, than the latter [21]. We also controlled the amount of auditory speech feedback available by having participants vocalize (Sound condition) or silently articulate (No-sound condition).

Each of the four experimental conditions was performed in trials of one minute duration, three per condition (12 trials per participant).

### 2.3. Apparatus and analysis

Face motion (jaw and lips motion) was captured using an Optotrak motion recording system (sample frequency at 150 Hz), which was synchronised with a Minibird system used to record the motion of the finger (sample frequency around 100 Hz).

Face motion and finger motion time-series were first dual passed through a second order Butterworth filter (fourth-order) with low pass cut-off frequency at 6 Hz and reported in a reference frame relative to the head (rigid face motion).

Three types of information were extracted from the face motion signals and the finger motion signal:

1. Speech dynamics (amplitude and variability of the face motion signals)

2. Finger dynamics (amplitude and variability of the finger motion signal)

3. The degree of synchrony between 1 and 2 (above). The first step here was to detect the onset of the speech as well as the flexion and extension maxima respectively from the time-series of the face (jaw and lips) and of the finger. Then, a pointwise relative phase ($\phi$) expressing the latency ($\Delta t$) of the speech event relative to the current finger cycle duration (T) was computed such as: $\phi = 2\pi * \Delta t/T$ ([6], see also Figure 1). We finally submitted the resulting distribution of relative phase to the Kuiper’s test, which gives two outputs: the probability for the tested distribution to be different from a randomized one, and a score reflecting how important this difference is. For present purposes, the latter score is the more important one as it represents the compactness of the relative phase distribution; the more compact, the more synchronized are speech and gesture. This measure was used as a synchronisation index (normalized for each participant).

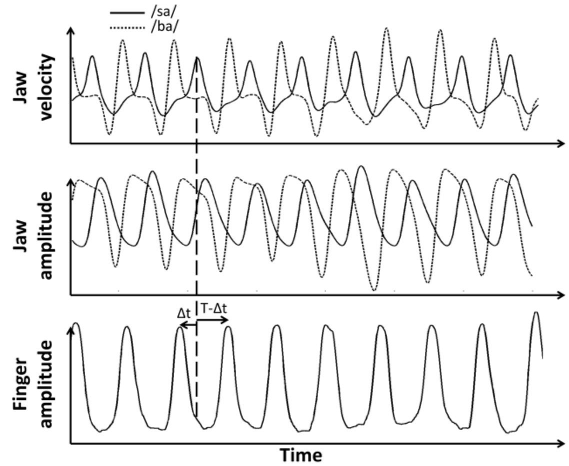


Figure 1: *Typical time-series of jaw amplitude and velocity for /ba/ (dashed line) and /sa/ syllables are illustrated. Below, the finger motion for the representative /sa/ trial is also given. The vertical dashed line indicates the onset of speech corresponding to a jaw velocity peak, reported in the finger time-series (see text for details).*

An analysis of variance with repeated measures with two factors (ANOVA, Syllable X Vocalization conditions), each one with two levels, was finally run.

## 3. Results

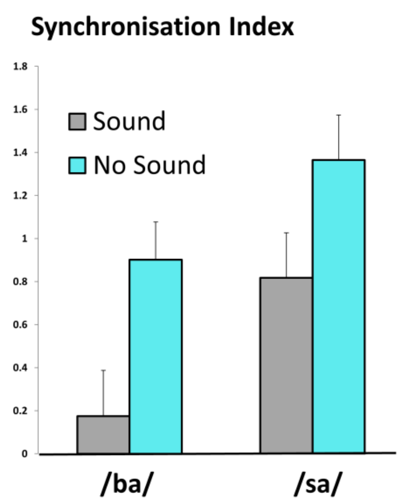


Figure 2: *Mean synchronization score (SEM) between speech and gesture (opening of the jaw onset relative to finger motion) for /ba/ and /ba/ syllables in the Sound and No Sound (silent articulation of speech) conditions.*

There was greater synchronisation between the face motion and the finger movements in the No sound than in the Sound condition ($F = 5.64$, $p = 0.023$; Figure 2), and for /sa/ syllables than for /ba/ syllables ($F = 4.25$, $p = 0.047$; Figure 2).

Amplitude of face motion ($F = 4.6$, $p = 0.039$) and velocity of jaw opening ($F = 7.07$, $p = 0.012$) were found to be larger for /ba/ syllables rather than /sa/ syllables.

Finally, we observed that there was larger variability of face motion ($F = 11.2$, $p = 0.002$) as well as of velocity peaks ($F = 10.75$, $p = 0.002$) in the No Sound condition than in the Sound condition.

No effects were found for the finger motion across the experimental conditions

# 4. Discussion

There are underlying constraints on temporal and spatial movement parameters that give rise to the coordination of limb movements, e.g., [22], [23]. Indeed, it has been suggested that synchronization can be seen as the central nervous system's default mode of operation [24]. Yet, many activities require that such a tendency for synchronisation be subdued. In the current study we looked at whether the motion involved in uttering repetitive syllables would show any tendency for synchrony with finger flexion/extension in a free performance setting.

We found evidence of different synchronisation levels between speech and gesture that depended on the characteristic of the speech. Contrary to our expectations, it turned out that less auditory as well as less sensorimotor feedback about the speech lead to greater synchronisation (greater influence) between speech and gesture.

## 4.1. Stronger non-intentional synchronisation in the No-Sound condition

We contrasted speech spoken aloud (Sound condition) with silently articulated speech (No-sound condition). We expected that less spoken auditory feedback should lead to a weaker influence from speech to the gesture and hence weaker synchronisation between the two. We found however that there was greater synchronisation in the No Sound condition.

In the No-sound condition, the results showed there was an increased variability in face motion. Variability of movements is usually related to their amplitude: the greater the amplitude, the more variable the motion will be ([25] for eyes saccades). Interestingly, in the case of the No sound condition, the greater variability in face motion was not associated with larger movements which may indicate other causes.

The relationship between stability and variability have been widely discussed and debated. In brief, it appears that high variability may point to difficulties in maintaining the robustness of a performance and that this might be a marker of system greater flexibility; something that in turn can facilitate the discovery of new stable solutions (see [9],[26] and [27] for a discussion about variability and stability of gait in elderly people).

In the present experiment then, the increase in the variability of the face motion in the No-sound condition might reflect the loss of robustness and a subsequent gain in the flexibility of the speech component. Such a reduction in robustness may have lead to a concomitant loss of independence of this component and an effective enhancement of the sensitivity of the forcing from the gesture component. On this view, the gain in synchronisation observed between speech and gesture would be the result of the stronger influences from the gesture to the speech, allowed by the loss of robustness of the speech component.

## 4.2. Stronger non-intentional synchronisation for /sa/ than for /ba/

The /sa/ and /ba/ syllables were selected based on the assumption that the production of /ba/ would involve more face motion than the production of /sa/. In turn, we expected that this greater recruitment of face motion would lead to a stronger influence of speech and hence reinforce the forcing from speech to the gesture component, i.e. to a greater synchronisation.

As expected, the amplitude of face motion was larger for the production of /ba/ compared to /sa/. However, there was greater synchronisation between speech and gesture for /sa/ than for /ba/. This greater synchronisation cannot be explained in the way to that was used to explain the greater synchronisation for the No-sound condition (discussed above). That is, unlike the No-sound condition, in which we argued that the increased variability of face motion was the marker of an increased sensitivity to the gesture component, there was no evidence of such increased variability for /sa/.

One possibility for why there was greater speech gesture synchrony for the /sa/ syllable is that the production of /sa/ also involves the active participation of the tongue articulator [21]. We propose that the behaviour of the tongue in the /sa/ utterance may reveal coupling between speech and gesture that was not reflected by the motion of the jaw. Further experiments are needed to test this hypothesis.

## 4.3. Less perceptive feedback about the speech leads to stronger synchronisation between speech and gesture

In summary, we manipulated two factors that relate to the amount of perceptive (auditory and sensorimotor) feedback in speech production. In a non-linear oscillators system, coupling is expressed as the mutual forcing of components and is increased by the reinforcement of one or the other of the forcing terms. We expected that a gain in perceptive feedback from speech would lead to a greater forcing from speech to the gesture and hence a greater synchronisation.

Our results indicated that different non intentional synchronisations were achieved between speech and gesture according to the characteristics of the speech. These were not related to speech contents or signification, but appeared to reflect the amount of perceptive feedback experienced. Contrary to our expectation, it appeared that less perceptive feedback from speech made this component more sensitive to gesture influence. We suggest that this makes sense by considering the bi-directionality of the influence of speech and gesture.

# 5. References


[1] Strogatz, S. H., and Stewart, I., "Coupled oscillators and biological synchronization", Scientific American, 269(6): 102-109, 1993.

[2] Strogatz, S. H., "Sync: The emerging science of spontaneous order", Hyperion, 2003.

[3] Néda, Z., Ravasz, E., Vicsek, T., Brechet, Y., and Barabási, A. L., "Physics of the rhythmic applause", Physical Review E, 61(6): 6987, 2000.

[4] Varlet, M., Marin, L., Lagarde, J., and Bardy, B. G., "Social postural coordination", Journal of Experimental Psychology-Human Perception and Performance, 37(2): 473, 2011.

[5] Kelso, J. S., Holt, K. G., Rubin, P., and Kugler, P. N., "Patterns of human interlimb coordination emerge from the properties of non-

linear, limit cycle oscillatory processes: Theory and data", Journal of motor behavior, 13(4): 226-261, 1981.
[6] Haken, H., Kelso, J. S., and Bunz, H., "A theoretical model of phase transitions in human hand movements", Biological cybernetics, 51(5): 347-356, 1985.
[7] Bardy, B. G., Marin, L., Stoffregen, T. A., & Bootsma, R. J., "Postural Coordination Modes Considered as Emergent Phenomena", Journal of Experimental Psychology Human Perception and Performance, 25: 1284-1301, 1999.
[8] Oullier, O., de Guzman, G. C., Jantzen, K. J., Lagarde, J., and Kelso, J. A. S., "Spontaneous interpersonal synchronization", In European workshop on movement sciences: Mechanics-Physiology-Psychology, 34-35, 2005
[9] Kelso, J. A. S., "Dynamic Patterns: the Self-Organization of Brain and Behavior", MIT Press, Cambridge, MA, 1995.
[10] Pikovsky, A., Rosenblum, M., & Kurths, J., "Synchronization: a universal concept in nonlinear sciences", Cambridge university press, 12, 2003.
[11] Kelso, J. S., Tuller, B., and Harris, K. S., "A "dynamic pattern" perspective on the control and coordination of movement", In The production of speech, Springer New York, 137-173, 1983.
[12] Smith, A., Mcfarland, D. H., and Weber, C. M., "Interactions between speech and finger movements: An exploration of the dynamic pattern perspective", Journal of Speech, Language and Hearing Research, 29(4): 471, 1986.
[13] Franz, E. A., Zelaznik, H. N., and Smith, A., "Evidence of common timing processes in the control of manual, orofacial, and speech movements", Journal of Motor behavior, 24(3): 281-287, 1992.
[14] Treffner, P., and Peter, M., "Intentional and attentional dynamics of speech–hand coordination", Human Movement Science, 21(5): 641-697, 2002.
[15] Parrell, B., Goldstein, L., Lee, S., and Byrd, D., "Articulatory evidence for functional coupling of speech and non-speech motor tasks", Proceedings of Speech Prosody V, 2010.
[16] Parrell, B., Goldstein, L., Lee, S., and Byrd, D., "Temporal Coupling Between Speech and Manual Motor Actions", 9th International Seminar on Speech Production, 2011.
[17] Treffner, P., Peter, M., and Kleidon, M., "Gestures and phases: The dynamics of speech-hand communication", Ecological Psychology, 20(1): 32-64, 2008.
[18] Rochet-Capellan, A., Laboissière, R., Galvan, A., and Schwartz, J. L., "The speech focus position effect on jaw-finger coordination in a pointing task" Journal of Speech, Language and Hearing Research, 51(6): 1507, 2008.
[19] Leonard, T., and Cummins, F., "The temporal relation between beat gestures and speech", Language and Cognitive Processes, 26: 1295-1309, 2010.
[20] Treffner, P. J., and Barrett, R., "Hands-free mobile phone speech while driving degrades coordination and control", Transportation Research Part F: Traffic Psychology and Behaviour, 7(4): 229-246, 2004.
[21] Mooshammer, C., Hoole, P., and Geumann, A., "Interarticulator cohesion within coronal consonant production", The Journal of the Acoustical Society of America, 120: 1028, 2006.
[22] Kelso, J. A. S., Southard, D. L. and Goodman, D., "On the nature of human interlimb coordination", Science, 203: 1029–1031, 1979.
[23] Marteniuk, R. G., MacKenzie, C. L. and Baba, D. M., "Bimanual movement control: information processing and interaction effects", Q. J. Exp. Psychol. A, 36: 335–365, 1984.
[24] Varela, F., Lachaux, J. P., Rodriguez, E. and Martinerie, J., "The brainweb: phase synchronization and large-scale integration", Nature Rev. Neurosci., 2: 229–239, 2001.
[25] van Beers, R. J., "The sources of variability in saccadic eye movements", The Journal of Neuroscience, 27(33): 8757-8770, 2007.
[26] Schöner, G., Haken, H., and Kelso, J. A. S., "A stochastic theory of phase transitions in human hand movement", Biological cybernetics, 53(4): 247-257, 1986.
[27] Toebes, M. J., Hoozemans, M. J., Furrer, R., Dekker, J., and van Dieën, J. H., "Local dynamic stability and variability of gait are associated with fall history in elderly subjects", Gait & Posture, 2012.