Measuring physiological speech entrainment with electromagnetic articulography (EMA)

Phil Hoole & Jonathan Harrington

IPS Munich

Setting up a dual EMA configuration for studying speaker interaction

Some EMA background

- A set of transmitter coils generates an alternating electromagnetic field at frequencies of about 10kHz
- ===> Induction of a position- and orientation-dependent signal in small sensors attached to the articulators
- Captures data for both visible (including head) and invisible articulators in readily analyzable form
- Good temporal resolution (samplerate typically > 200Hz)

First approach: Edinburgh Speech Production Facility

Two Carstens AG500 machines

Main drawback: Both systems use the same transmitter frequencies

==> interference ==> speakers must be at least 6m apart

==> in practice, in separate rooms (connected by CCTV)

Geng, C. et al. (2013) *Recording Speech Articulation in Dialogue: Evaluating a synchronized double Electromagnetic Articulography Setup*. J. Phonetics 41(6): 421-431.

Second Approach

1x Carstens AG500 with 1x NDI Wave

No mutual interference of transmitter frequencies

Speakers can be face-to-face, about 2m apart

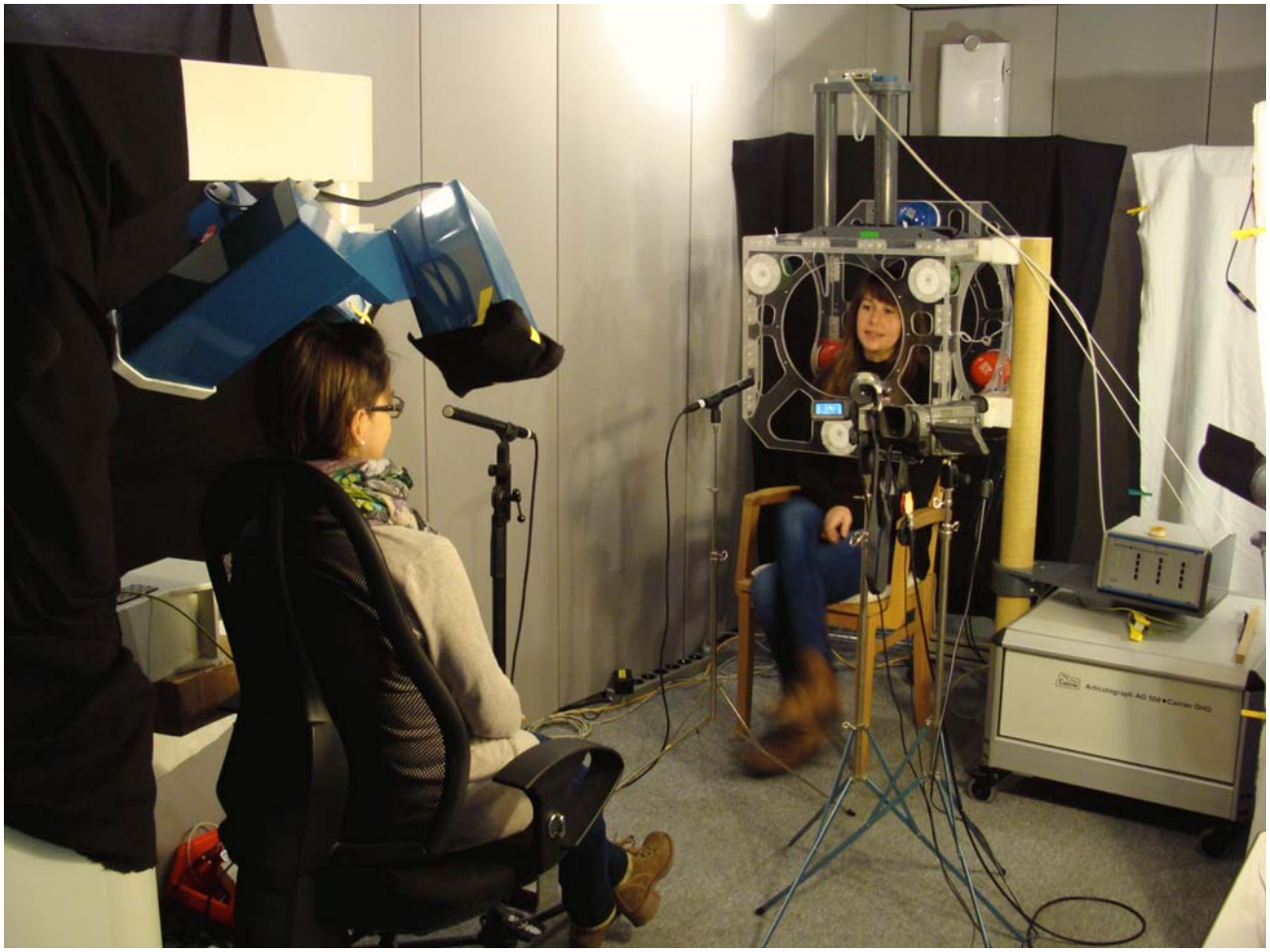
Used for pilot experiment at MARCS lab, Univ. W. Sydney:

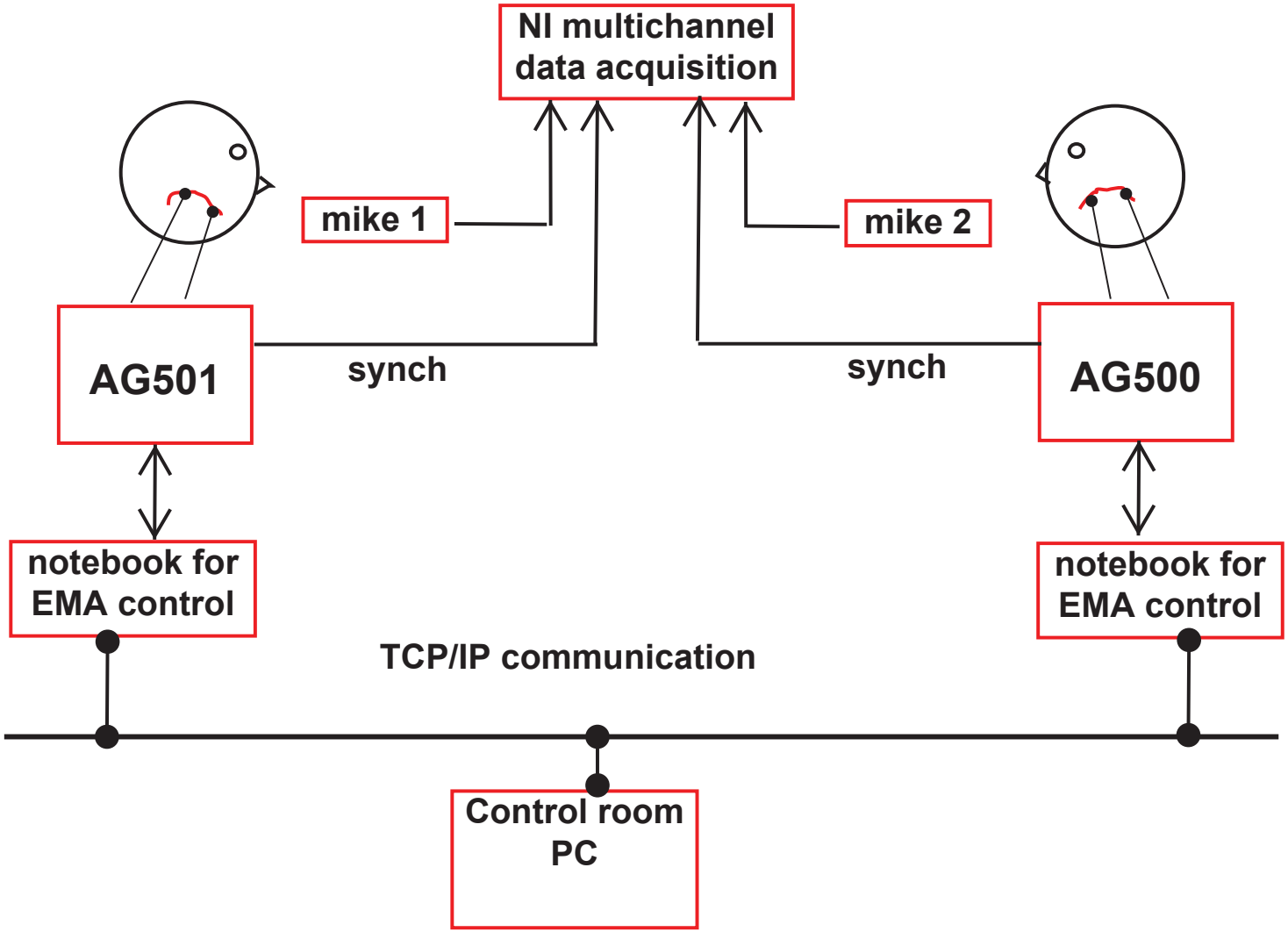
Tiede, M., Bundgaard-Nielsen, R., Kroos, C., Gibert, G., Attina, V., Kasisopa, B., Vatikiotis-Bateson, E. & Best, C. (2010). *Speech articulator movements recorded from facing talkers using two electromagnetic articulometer systems simultaneously*. Proceedings of Meetings on Acoustics, Vol. 11, 060007

Further experiments at Haskins Labs using this setup by M. Tiede and C. Mooshammer

Our setup: 1 x Carstens AG501 with 1x Carstens AG500

No mutual interference at inter-speaker distances > 2m




NI multichannel
data acquisition
mike 1
mike 2
AG501
AG500
synch
synch
notebook for
EMA control
notebook for
EMA control
TCP/IP communication
Control room
PC

Further features of recording setup

- No noisy equipment in recording booth
- HDV video: merged view of 2 cameras plus control PC monitor
- Additional PCs for data visualization during session

Pilot experiment based on Tiede et al. experiment at MARCS

EMA sensors

articulators: tongue tip, tongue body, upper lip, lower lip

head: upper incisors, bridge of nose, behind left and right ear

After recording, head sensors used

(1) to factor out head movement from articulator sensors

(2) to calculate the rigid-body parameters of head movement (3 translations, 3 rotations)

Subjects spoke alternating word sequences for 30s, e.g.

S1: “Topf Kopf Topf Kopf ......”

S2: “Kopf Topf Kopf Topf ......”

sometimes “disturbed” by instructions from investigator to speed up or slow down.

cf. Goldstein, L., Pouplier, M., Chen, L., Saltzman, E. & Byrd, D. (2007). *Dynamic action units slip in speech production errors*. Cognition 103, 386-412.

[some non-alternating sequences also recorded, e.g.

S1: “Topf Topf Topf Topf ......”

S2: “Kopf Kopf Kopf Kopf ......”

]

Illustrations of one approach to analyzing entrainment patterns between and within speakers

based on:

Barbosa, A., Déchaine, R.-M., Vatikiotis-Bateson, E. & Yehia, H. (2012). *Quantifying time-varying coordination of multimodal speech signals using correlation map analysis*. JASA 131 (3), 2162-2172.

probably many other possibilities, e.g.

Lancia, L., Fuchs, S. & Tiede, M. (2013). *Application of concepts from Cross-Recurrence Analysis in speech production: An overview and a comparison to other nonlinear methods.* JSLHR (in press)

Basic idea:

- Move a 2-second window in steps of 40ms across each trial

  (window length and step size can be chosen to match the time scale of the movement patterns of interest)
- At each time step compute the correlation between signal 1 and lagged versions of signal 2

  signal 1: e.g. speaker 1, tongue body vertical position (TB_Z)

  signal 2: e.g. speaker 2, tongue tip vertical position (TT_Z)

  lags used here: -2 to +2 seconds in steps of 4ms

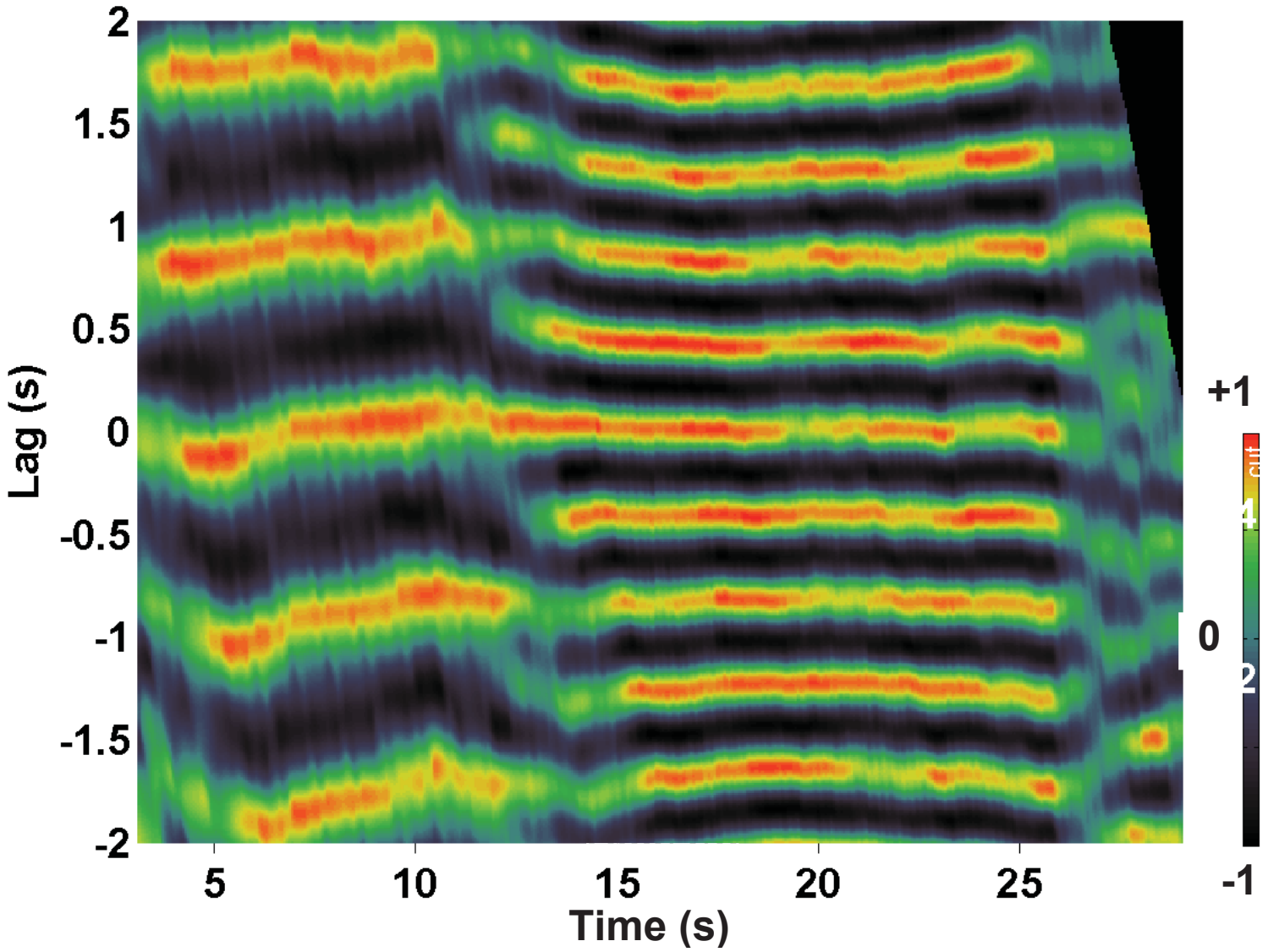
===> sonagram-style display of cross-correlations over time

Example 1

S1:
“Kopf Kopf ...”
S2:
“Topf Topf ...”

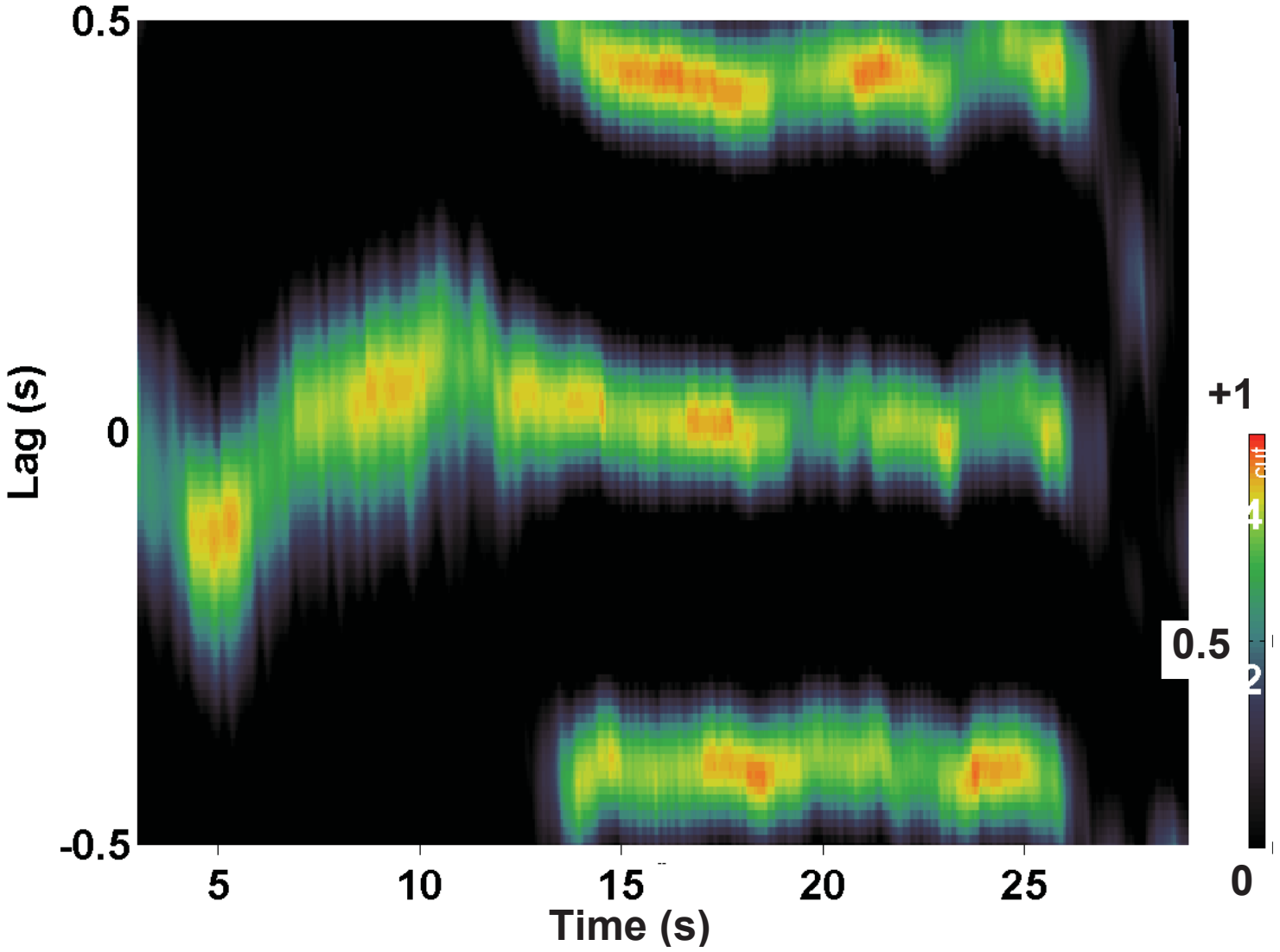
Normal to fast tempo

S1 TB_Z
vs.
S2 TT_Z



Zoom in around zero lag.

Note shift in lag giving maximum correlation

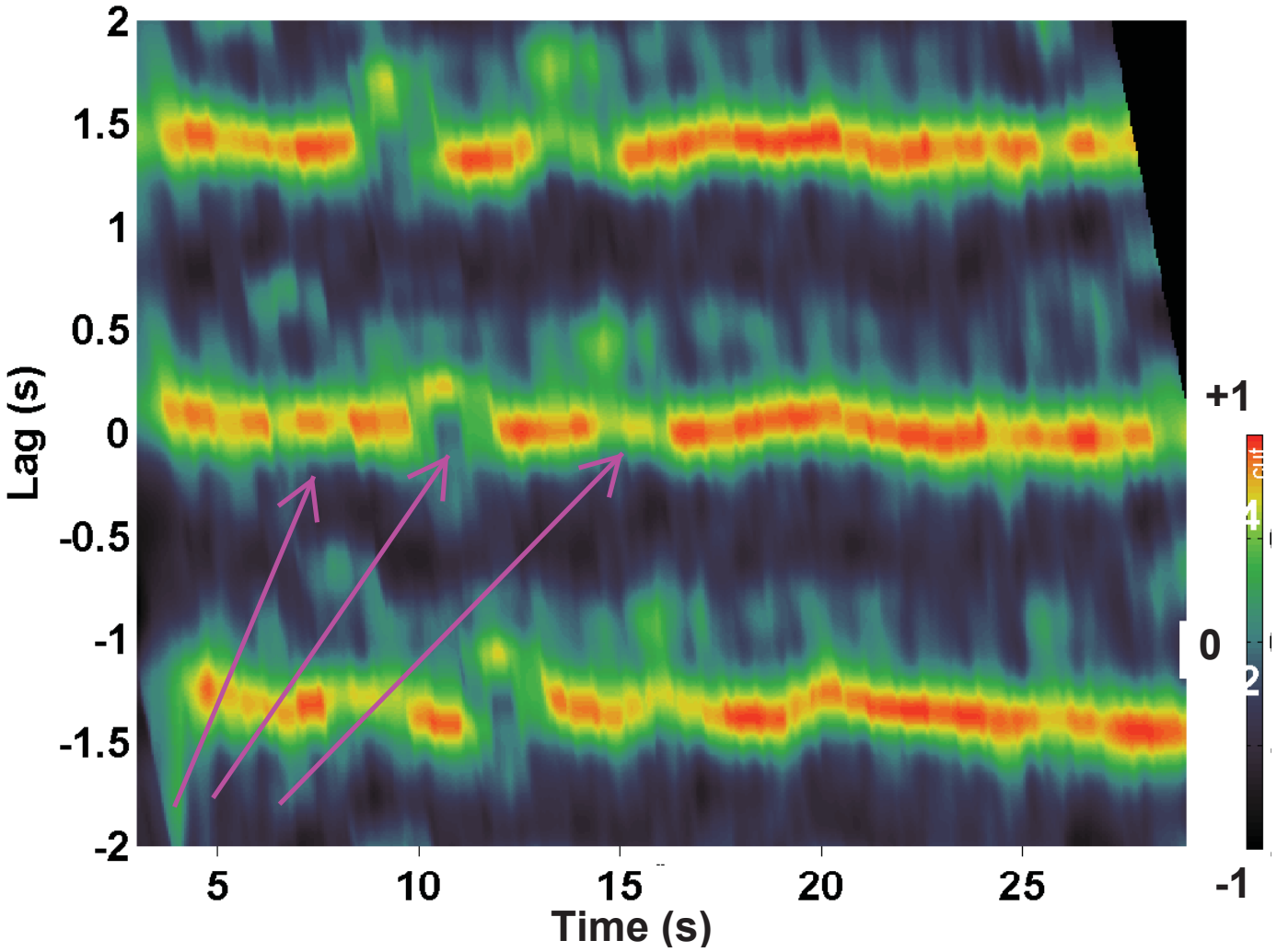


Example 2

S1:
“Köpfe Töpfe ...”
S2:
“Töpfe Köpfe ....”

No tempo change

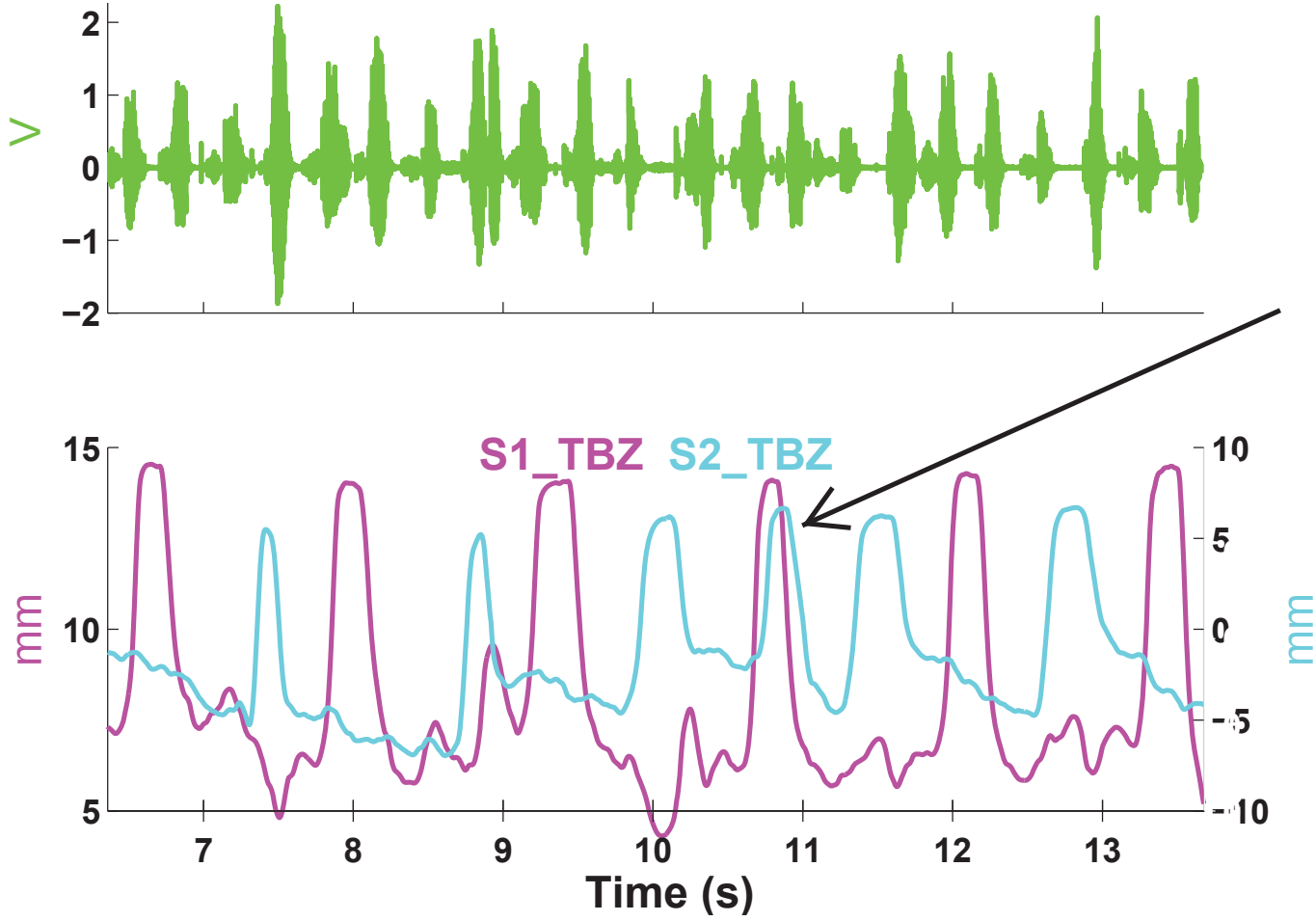
S1 TB_Z vs.
S2 TT_Z



Basic alignment of TB and TT around zero lag (i.e. S1 “Köpfe” with S2 “Töpfe”). But note instabilities in correlation!

Alternating phase of tongue-tip and tongue-body is unstable.

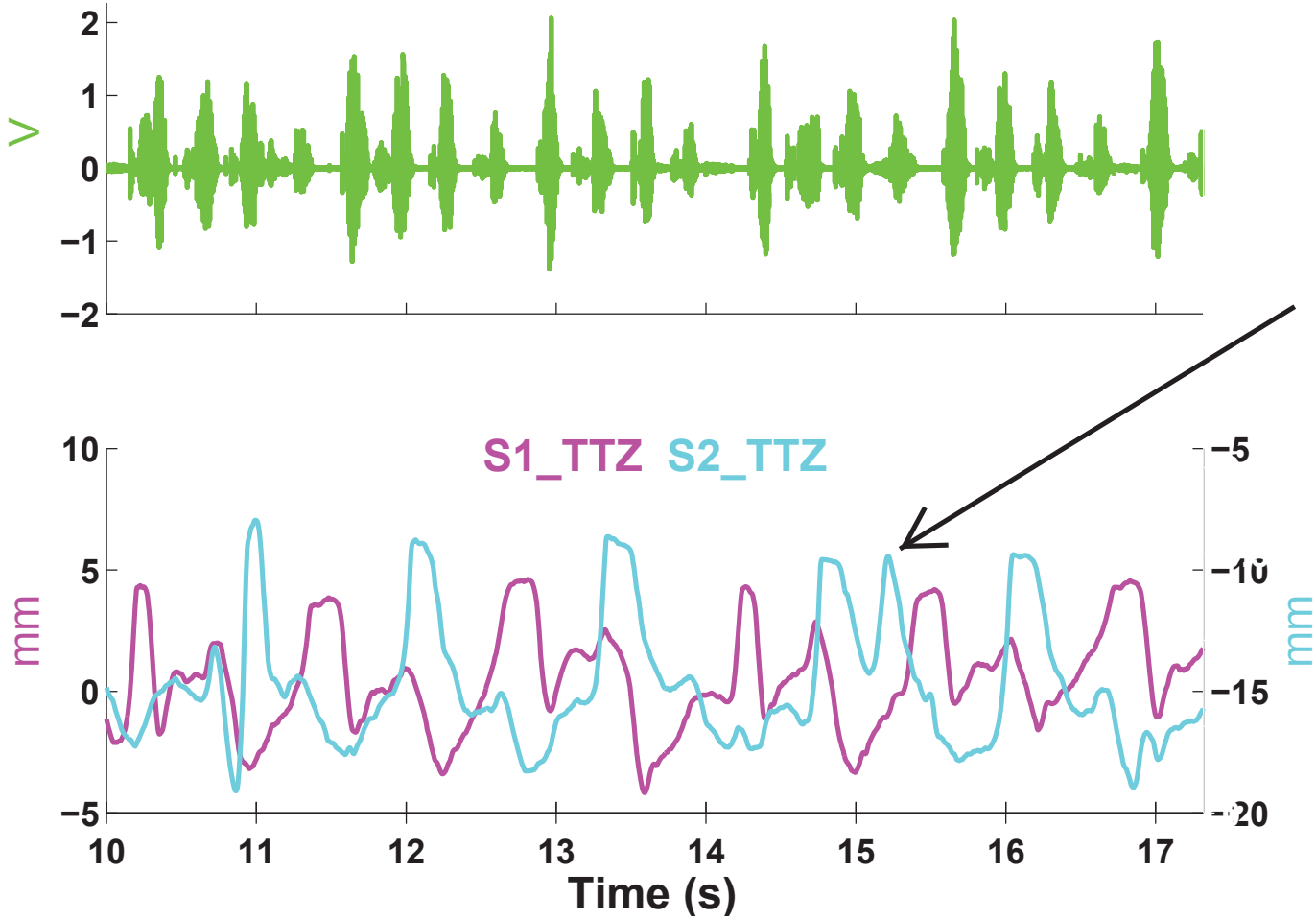
Within-speaker entrainment towards all articulators in phase



Intrusive gesture of TB for S2

Alternating phase of tongue-tip and tongue-body is unstable.

Within-speaker entrainment towards all articulators in phase



Intrusive gesture of TT for S2

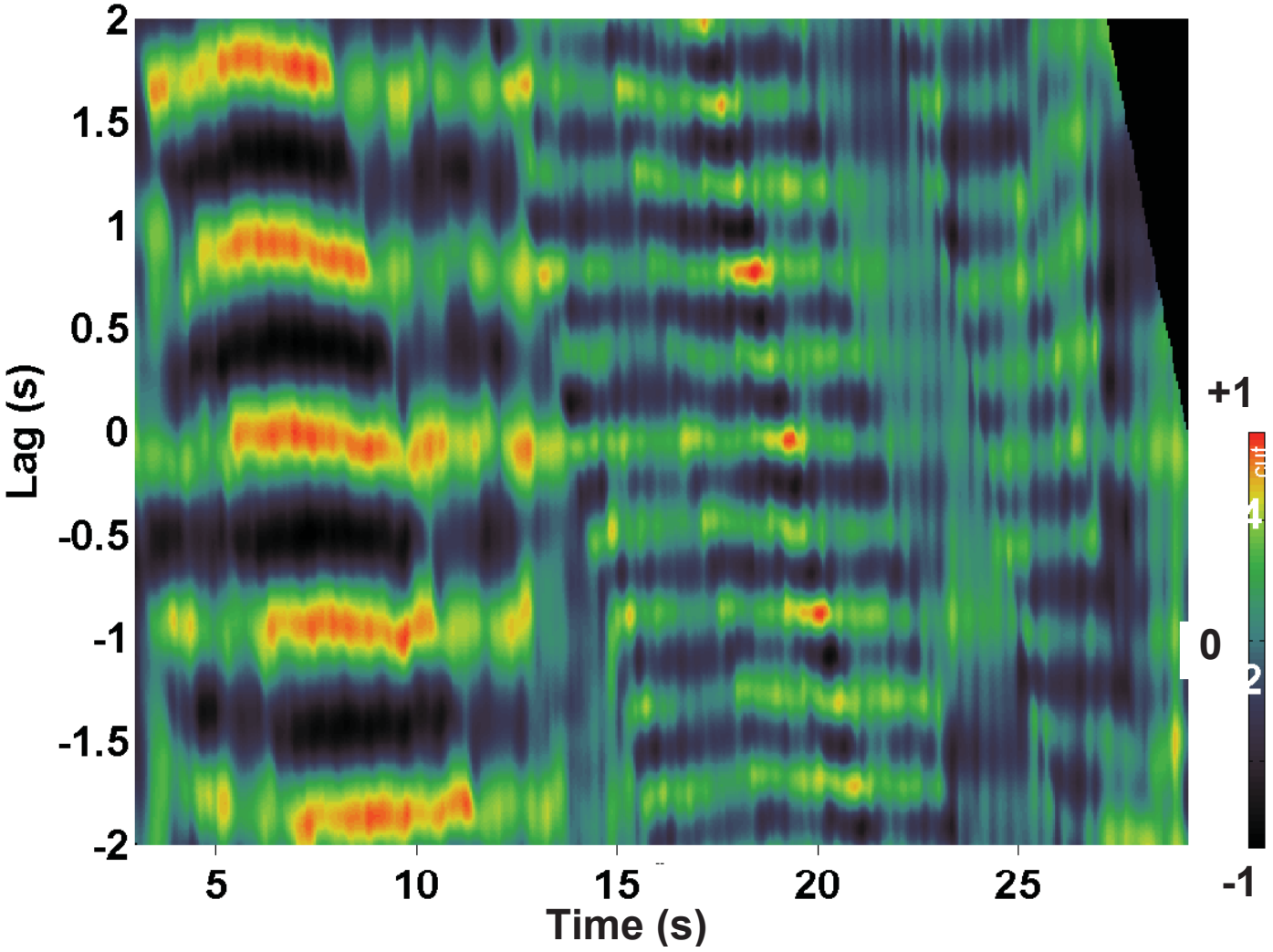
Within-speaker entrainment between head and articulators in a rhythmic speech task?

Tiede, M., Goldstein, L., Mooshammer, C., Nam, H., Saltzman, E. & Shattuck-Hufnagel, S. (2011) *Head movement is correlated with increased difficulty in an accelerating speech production task*. Proc. International Seminar on Speech Production, Montreal

S1:
“Kopf Kopf ....”
S2:
“Topf Topf ....”

Normal to fast tempo

S2 lip aperture
vs.
S2 head vertical pos.



In our data many cases of sporadic entrainment between articulatory movement and head movement for speaker 2, but no obvious link to task difficulty.

(except possible increase in magnitude of non-rhythmic head movements in the vicinity of articulatory break-down)

Between-speaker entrainment of head movement?

Since Speaker 1 showed very little head-movement (and overall much weaker correlations between lip and head than Speaker 2) no between-speaker head-movement entrainment was found.

Our speakers generally entrained their speech quickly at the start of the trial.

Head movement may still be useful to support entrainment when start conditions are less predictable.

Final comments

Examples of entrainment in this presentation based on rhythmic utterances (“speech-error” paradigm).

But Tiede & Mooshammer have recently applied the dual-EMA setup to a more traditional phonetic convergence paradigm.

Kinematic properties of velar consonants in test items more similar after a period of face-to-face interaction.

Tiede, M. & Mooshammer, C. (2013) *Evidence for an articulatory component of phonetic convergence from dual electromagnetic articulometer observation of interacting talkers*. Proceedings of Meetings on Acoustics, Vol. 19, 060138

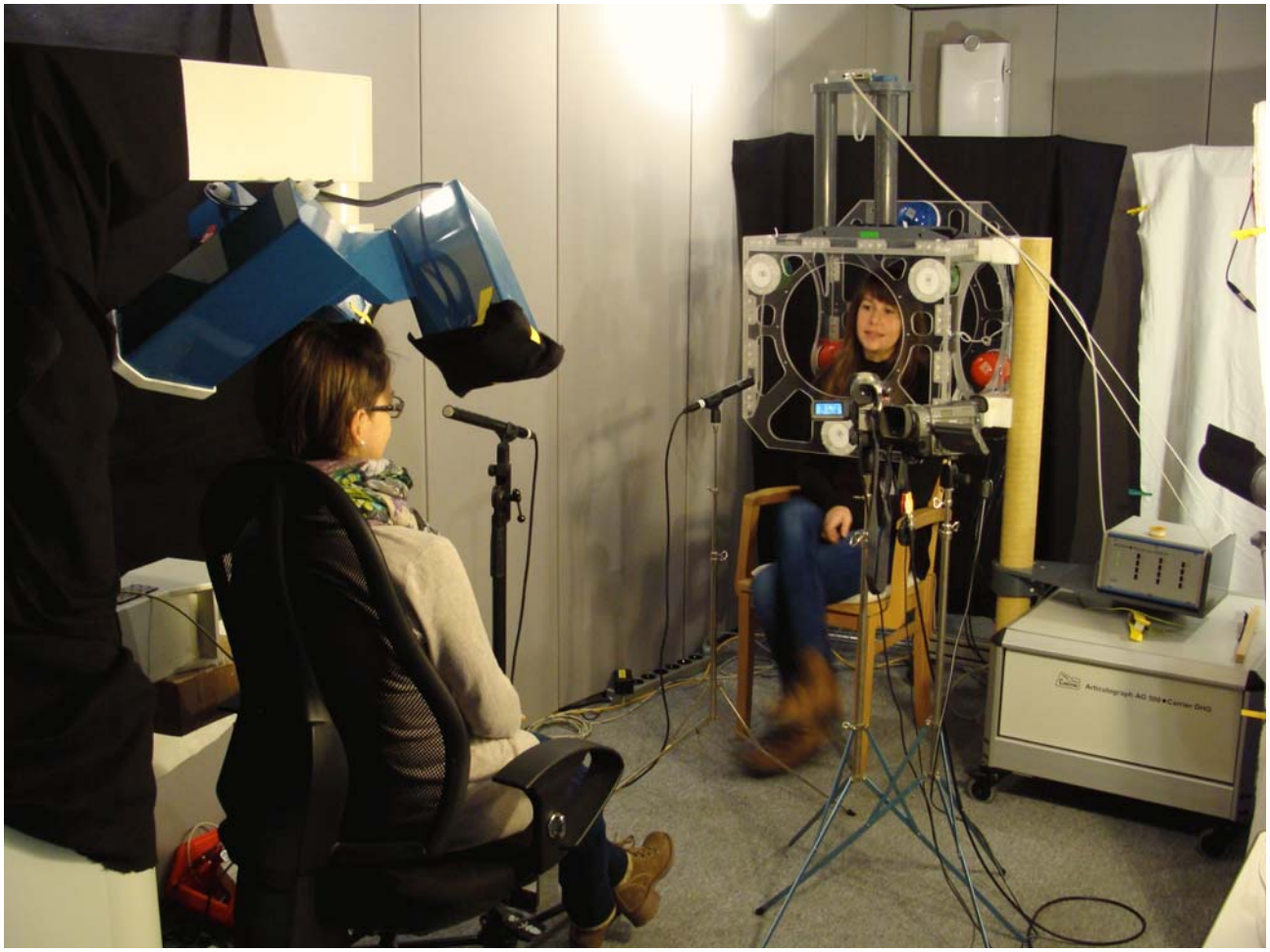
Outlook

Upgrade AG500 to AG501 (2x AG501 setup is possible because AG501 can be configured to different transmitter frequencies)

Advantages of AG501:

- data-processing much less time-consuming
- large head-movements captured with much less measurement error
- subject less enclosed (easier to combine with optical systems)
- ===> improved possibilities for more natural subject interaction

Other labs (e.g. USC) are exploring use of dual NDI Wave systems.



Thanks to Susanne Waltl for help with the recording

Thanks to Véro and Sandra for letting their tongues be twisted

May 4, 2014: Workshop on "Interpersonal coordination and phonetic convergence" organized by Tine Mooshammer & Mark Tiede (satellite meeting of 10th Int. Seminar on Speech Production, Cologne, May 5-8, 2014). **!!Abstract deadline Dec. 15, 2013!!**