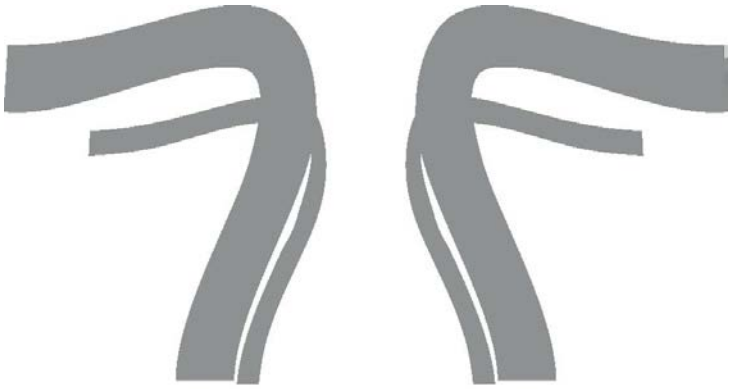


10th International Seminar on Speech Production, 5 – 8 May, 2014

Abstract booklet of Satellite Workshop on "Interpersonal coordination and phonetic convergence"

Date: May 4, 2014, 10am-6pm
Location: Neuer Senatssaal (Main building), University of Cologne

Organizers:

Tine Mooshammer, Institut für deutsche Sprache und Linguistik, Humboldt- Universität zu Berlin and Haskins Laboratories

Mark Tiede, Haskins Laboratories

Introduction

The investigation of coordinated behavior arising from interpersonal interaction is an emerging area of research recently attracting increased attention. Thanks to funding provided through the Humboldt University of Berlin and the National Institutes of Health we are pleased to provide this opportunity to consider these topics as presented by our distinguished invited speaker and contributing authors. We also wish to thank Oxana Rasskazova and Birgit Schenk for their assistance, and the organizers of the ISSP for their support of this satellite program.

Tine Mooshammer and *Mark Tiede*

Workshop Program:

Workshop on Interpersonal Coordination and Phonetic Convergence
Cologne, May 4, 2014

10:00-10:15 **Welcome**

10:15-11:15 **Invited Speaker** *Ivana Konvalinka*: Synchronized and complementary mechanisms underlying interpersonal coordination across behavioural, neural, and physiological domains

Levels of Processing

11:15-11:35 *Vijay Solanki, Jane Stuart-Smith, Rachel Smith, Pascal Belin, Alessandro Vinciarelli*: An Investigation into Phonetic Accommodation and its Neural Correlates in a Collaborative Conversational Task

11:35-11:55 *Francisco Torreira, Sara Bögels, Stephen Levinson*: Breathing and speech planning in turn-taking

11:55-12:10 **Coffee Break**

12:10-12:30 *Alessia Tosi*: Deception impedes lexical entrainment in verbal interaction

12:30-12:50 *Melanie Weirich, Susanne Fuchs, Adrian Simpson*: Listening tongues: An EMA and EPG study on tongue movements during the perception of speech

12:50-13:50 **Lunch Break**

Methodological Issues

13:50-14:10 *Rob Fuhrmann, Eric Vatikiotis-Bateson*: Coordination across task and event domains

14:10-14:30 *Molly Babel, Michael McAucliffe, Grant McGuire*: Spectral similarity and listener judgments of phonetic accommodation

Accommodation between speakers with different language backgrounds

14:30-14:50 *Nicholas Henriksen*: Convergence in bilingual rhythm

14:50-15:10 *Shiri Lev-Ari, Sharon Peperkamp*: Do people converge to the linguistic patterns of non-reliable speakers? Perceptual learning from non-native speakers

15:10-15:25 **Coffee Break**

Manipulation accommodation

15:25-15:45 *Benjamin Schultz, Irena O'Brien, Natalie Phillips, David McFarland, Deborah Titone, Caroline Palmer*: Interlocutors' speech rates converge: The effects of fast and slow confederate speech rates

15:45-16:05 *Natalie Lewandowski, Antje Schweitzer*: Degrees of control over influencing factors in phonetic convergence

16:05-16:25 *Maëve Garnier, Gabrielle Richard, Lucie Ménard*: Modulation of visible and non-visible articulatory movements in perturbed face-to-face communication

16:25-16:40 **Break**

16:40-17:40 **Special Session**: dual-EMA systems

17:40-18:00 **Closing Discussion**

Synchronized and complementary mechanisms underlying interpersonal coordination across behavioural, neural, and physiological domains

Ivana Konvalinka

Section for Cognitive Systems, DTU Compute, Technical University of Denmark
Department of Cognitive Science, Central European University, Hungary

A large body of research has investigated the mechanisms that enable people to coordinate their actions in order to achieve a joint goal. Successful interpersonal coordination has been shown to be facilitated via continuous feedback from two or more people, and their mutual adaptation to each other's actions. However, this adaptation need not always be symmetric. In this talk, I will present several studies investigating mechanisms of interpersonal coordination in the behavioural, neural, and physiological domain, showing that people rely on both symmetrical and asymmetrical interpersonal mechanisms when working to achieve coordination. Given similar task constraints, participants synchronize best when adopting mutual adaptation strategies. However, when there is an asymmetry in task difficulty between partners, they implicitly negotiate complementary leader-follower dynamics in order to effectively achieve the common goal, such that the person with the more difficult task takes on the role of the leader. I will discuss how these leader-follower roles can be predicted from dual EEG recordings of brain activity, and what these patterns tell us about the neural mechanisms underlying leadership. Finally, I will show how coordination in the physiological domain (respiration & heart rate) follows similar principles as coordination in the behavioural domain.

An Investigation into Phonetic Accommodation and its Neural Correlates in a Collaborative Conversational Task

Solanki, V., Stuart-Smith, J., Smith, R., Belin, P., Vinciarelli, A.
University of Glasgow (UK)

v.solanki.1@research.gla.ac.uk

Abstract

Speakers are known to make subtle yet communicatively important adjustments to their speech during dialogue in relation to their interlocutor, a process known as speech accommodation. At a conceptual level, it has been proposed that these adjustments might be important in the alignment of mental situational models when communicating with another person (Pickering & Garrod, 2013) and in social cohesion between speech partners (Coupland & Giles, 1988; Giles, Taylor, & Bourhis, 1973). Essentially, these models propose a fundamental need to reduce uncertainty in the signal through capitalising on the regularities inherent in the most similar aspects of the speech signals of both conversational partners. However, the experimental techniques used to verify these theories at a phonetic level have often overlooked the interactive and interdependent nature of conversation. They have instead favoured of an investigative approach which gains experimental control by delimiting the amount of possible variation in the signal through the use of pre-recorded speech. In order to provide strong evidence to support these hypotheses, studies looking at truly interactive speech phenomena are needed and thus far, these are somewhat lacking. The presented project utilizes a Dual-EEG methodology, where EEG signals of two people are simultaneously recorded whilst performing a structured collaborative task designed to elicit specific phonetic features in order to investigate the relationships between phonetic variation and brain activity.

Within the field of phonetics, there is a large body of research investigating the types of phonetic variables which contribute to the shifting of speech in relation to their conversational partners (Babel, 2012; Pardo, 2012; Tobin, 2013). These works generally tend to focus on the individual features of the phonetic repertoire that each interlocutor produces during conversation, e.g. VOT (Tobin, 2013), F0 (Babel, 2012). However, the idea that human beings interpret the acoustic features of another's speech based on just a small number of key features of the acoustic signal rather than making use of all available information contained within the speech signal, seems unlikely. Additionally, the studies mentioned above find that a large number of social and individual factors are strongly correlated with the variation which occurs in the acoustic patterning of conversational speech adaptation. The project presented here, takes this into account and assesses phonetic features from fine-grained temporal variables up to more supra-segmental aspects of the speech signal whilst controlling for social and individual variance.

In the fields of psychology and brain imaging, there is a growing body of research which aims to assess the link between the variation in the acoustic signal of human speech and neural activity (Ghitza, Giraud, & Poeppel, 2013). If it is the case that our mental models of a given interaction align across the course of a conversation based on the acoustic properties of our speech patterning, then it could be proposed that a form of alignment between brain processes in the interlocutors should also be present. However, little is known about whether a link exists between the underlying phonetic variation in interactive communication and the neural processes which are hypothesised to drive alignment. Those studies which have attempted to investigate this matter, whether from a phonetic or neural perspective, have often failed to capture the full spectrum of interaction because the models are usually inferred from non-interactive experiments (responding to pre-recorded speech), rather than from interaction itself.

In order to elicit free conversational speech, the DiapixUK task is used (Baker & Hazan, 2011), two participants must verbally interact in order to solve a spot-the-difference task where each participant has a different version of the same pictorial scene. The effects which are being assessed are subtle and a strong behavioural effect is required for sufficient activity to be detected by the EEG. It has been demonstrated that women are both more likely to accommodate and will tend to show greater accommodative magnitude (Bulatov, 2009; Namy, Nygaard, & Sauerteig, 2002), accommodation is more likely to be present in conversational pairs from the same local area but with different local accents (Kim, Horton, & Bradlow, 2011) and accommodation is more likely to occur amongst pairs who are more similar to one another or who tend to like each other (Bailly, Lelong, et al., 2010). Taking these aspects into account, the participant group will consist of female undergraduate students from the City of Glasgow conurbation who will self-select their partners based on standardised photographs. These restrictions on the participant pool serve as proxies for the features outlined above in order to elicit as large a degree of accommodation as possible. The acoustic data is transcribed and analysed for changes in voice-onset-time (VOT), vowel formants (F1 and F2), fundamental frequency (F0) and speech rate. The signals from each recording domain are correlated between speakers such that for each speaker dyad there are correlations for acoustic only data, EEG only data and for data across modalities (ie. EEG-Acoustic). In this way, the interactive and interdependent nature of conversation can be captured in relation to the phonetic and neural processes which are involved.

It is hypothesised that over the course of the interaction, signals from all recording domains will trend towards one another. If the signals tend toward one another irrespective of recording domain, it will provide evidence suggesting a strongly coupled link between the acoustic speech signal and ongoing brain processes. However, even if correlations can only be found within the same recording domains, this will still go some way to illuminating the link between phonetic adaptation and neural processing of interdependent conversational speech. If these predictions are supported by the data then this will provide support for theories of mental alignment in order to aid communication and social cohesion, along with the demonstration of a possible root for this alignment in fine-grained phonetic features of speech. Importantly, it will derive its evidence from a dynamic and ecologically valid conversational setting.

Keywords
Dual EEG, Brain-to-Brain Coupling, Social Interaction, Accommodation, Phonetic Entrainment

References

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

Bailly, G., Lelong, A., et al. (2010). Speech dominoes and phonetic convergence. *Proceedings of Interspeech 2010*, 1153–1156.

Baker, R., & Hazan, V. (2011). Diapixuk: task materials for the elicitation of multiple spontaneous speech dialogs. *Behavior research methods*, *43*(3), 761–770.

Bulatov, D. (2009). The effect of fundamental frequency on phonetic convergence. *Berkeley Phonology Lab Annual Report*, *2009*, 404–434.

Coupland, N., & Giles, H. (1988). *Communicative accommodation: Recent developments*. Pergamon Press.

Ghitza, O., Giraud, A.-L., & Poeppel, D. (2013). Neuronal oscillations and speech perception: critical-band temporal envelopes are the essence. *Frontiers in Human Neuroscience*, *6*(340).

Giles, H., Taylor, D. M., & Bourhis, R. (1973). Towards a theory of interpersonal accommodation through language: Some canadian data. *Language in society*, *2*(2), 177–192.

Kim, M., Horton, W. S., & Bradlow, A. R. (2011). Phonetic convergence in spontaneous conversations as a function of interlocutor language distance. *Laboratory Phonology*, *2*(1), 125–156.

Namy, L. L., Nygaard, L. C., & Sauerteig, D. (2002). Gender differences in vocal accommodation: The role of perception. *Journal of Language and Social Psychology*, *21*(4), 422–432.

Pardo, J. S. (2012). Reflections on phonetic convergence: Speech perception does not mirror speech production. *Language and Linguistics Compass*, *6*(12), 753–767.

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioural and Brain Sciences*.

Tobin, S. (2013). Phonetic accommodation in spanish-english and korean-english bilinguals. In *Proceedings of meetings on acoustics* (Vol. 19)

Breathing and speech planning in turn-taking

Francisco Torreira, Sara Bögels & Stephen C. Levinson
Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

In conversation, turn transitions between speakers often occur smoothly, usually within a time window of a few hundred milliseconds (Stivers et al., 2009). Since planning and producing a simple word takes around 600 ms (Levelt et al., 1999), it has been argued that conversational participants must start planning their utterance in overlap with their interlocutor's turn to achieve smooth turn transitions (Levinson 2013). In order to avoid overlaps and long gaps, speakers must also time their utterances accurately with respect to the end of their interlocutor's turn. However, direct evidence of early speech planning and accurate turn-end identification in conversation still remains scarce. In this talk, we present preliminary data from a project aimed at investigating whether the breathing behavior of conversational participants can provide such evidence. While previous studies on read speech have identified a relationship between breathing behavior and utterance duration (Whalen & Kinsella-Shaw, 1996; Fuchs et al., 2013), it remains to be investigated whether breathing behavior can be informative about speech planning in conversational speech as well.

Six dyadic unscripted conversations between Dutch male friends were recorded with head-mounted microphones and an Inductotrace^{TM} system of inductive plethysmography for around 40 minutes each. Each participant wore an Inductotrace band attached around his chest at the level of the axilla and a head mounted-microphone coupled to an amplifier. The speech and breathing signals were recorded simultaneously via an A/D converter connected to a computer.

We extracted and segmented 144 question and answer sequences from our data, and annotated all answerer's inbreaths that occurred between the start of the question and the start of the answer. We then examined a) if a relationship exists between the answerer's breathing behavior and the length of the answer, as has been found for read speech, and b) the timing of the answerer's inbreath relative to the end of the question (i.e. the moment when an answer is expected).

We found that 37% of the answers were not preceded by an inbreath, and, interestingly, that the presence vs. absence of an inbreath before the answer was related to the length of the answer. Figure 1 shows boxplots of answer duration for answers preceded and not preceded by an inbreath. It can be seen in this figure that answers preceded by an inbreath tended to be significantly longer than answers not preceded by an inbreath. This difference was statistically significant in a regression model with answer length as response and the presence of an inbreath as a predictor (β = 1087, t = 3.88, $p < .001$). It indicates that, as in controlled read speech, speakers' breathing behavior can be informative about the scope of speech planning in conversational speech as well.

Regarding the timing of the answerer's inbreaths relative to the questioner's turn end, we observed that the most frequent timing (i.e. the mode of the distribution) was located at the end of the question. This is illustrated in Figure 2a, which shows a density plot of this measure. It can also be seen in this figure that the distribution of inbreath timings is skewed to the left, with more inbreaths starting in overlap with the question (with negative values) than in the gap following the question (with positive values). Subsequent inspection of early vs. late inbreaths revealed that inbreaths occurring before long answers, for which inbreaths are presumably required, tended to cluster closely around the end of the question, whereas inbreaths preceding short answers displayed a wider spread and earlier timings in general. This is illustrated in Figure 2b, which shows the timing of inbreaths for answers shorter and longer than 2.5 s. These observations suggest that many of the early inbreaths could be vital or partly-vital inbreaths not primarily intended for speech, and that it is the late inbreaths that could be more specifically designed for speech. In our question and answer sequences, therefore, the timing of speech inbreaths before the longer answer appears to be sensitive to where precisely the question ends.

Additionally, given that preparing an inbreath requires at least a few hundred ms (i.e. activation of internal intercostals alone requires 140 to 320 ms, Draper et al. 1960) and that speech inbreaths are contingent on the planned length of the answer, the fact that inbreaths mostly occur close to or before the end of the question provides evidence that participants in a conversation often start planning their speech in overlap with their interlocutors' turn.

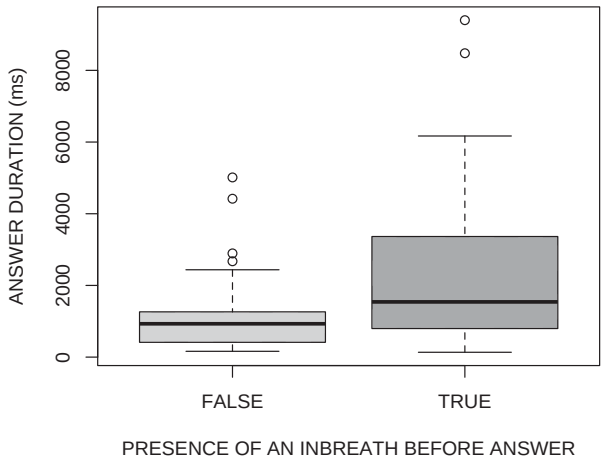


Figure 1 Answer duration as a function of the presence or absence of an inbreath before the answer.

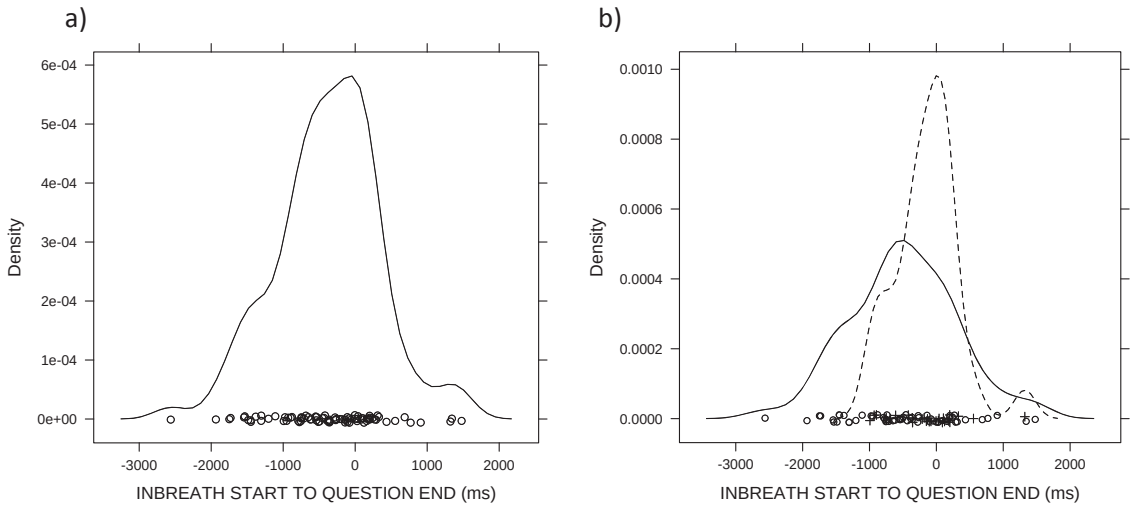


Figure 2 Answerer's inbreath start relative to question end (ms), divided in two groups of answer duration in the right panel (dashed: > 2500 ms; solid: < 2500 ms).

References

Draper, M. H., Ladefoged, P., and Whitteridge, D. (1960) Expiratory pressures and airflow during speech. *British Medical Journal*, 1(5189): 1837–1842.

Fuchs, S., Petrone, C., Krivokapic, J., and Hoole, P. (2013). Acoustic and respiratory evidence for utterance planning in German. *Journal of Phonetics*, 41(1):29–47.

Levelt, W., Roelofs, A., and Meyer, A. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22(1):1–37.

Levinson, S. C. (2013). Action formation and ascription. In Stivers, T. and Sidnell, J., editors, *Handbook of Conversation Analysis*, pages 103–130. Wiley-Blackwell.

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., Hoymann, G., Rossano, F., de Ruiter, J. P., Yoon, K.-E., and Levinson, S. C. (2009). Universals and cultural variation in turn-taking in conversation. *PNAS*, 106(26):10587–10592.

Whalen, D. H. and Kinsella-Shaw, J. M. (1997). Exploring the relationship of inspiration duration to utterance duration. *Phonetica*, 54:138–152.

Deception may impede lexical entrainment in verbal interaction

Alessia Tosi, Holly Branigan, Adam Moore, & Martin Pickering (University of Edinburgh)
a.tosi@sms.ed.ac.uk

People interacting in dialogue tend to converge on the use of referring expressions (*lexical entrainment)*. This interactive adaptation has been claimed to be instrumental to interlocutors' mutual understanding and successfulness of communication [1].

Most research on coordination in dialogue has assumed clear shared goals for the interlocutors. This cooperation bias in the study of interaction reflects the real-life expectation that in communication people act cooperatively and transparently, and that they tell the truth. But of course this is not always the case. We report a study that investigates lexical entrainment in the context of deceptive interactions; our findings suggest that deception may impede the deceiver's tendency to lexical entrain.

There are broadly two types of approaches as of how linguistic coordination arises in dialogue. Unmediated accounts (see [1]) argue that interpersonal coordination is grounded in the functional architecture of the individual's language system and is automatically driven by priming processes (i.e., tendency to pick up and reuse expressions used before by oneself or the interlocutor), unaffected by extra-linguistic factors such as attributional cues about the interlocutor. Mediated accounts [4] assume that online language processing in dialogue is guided by explicit considerations of communicative design. According to this approach, the goal to understand and be understood gives rise to behavioural coordination (e.g., lexical entrainment).

Overall, findings from dialogue research have shown that under highly cooperative and goal-oriented circumstances, linguistic coordination seems to arise without particular effort from the interlocutors [1]. Yet, findings have also shown that alignment, and lexical entrainment in particular, can be mediated by the interlocutors' beliefs about their partner [5]. More generally, behavioural coordination has been proved to be a widespread phenomenon that have positive affective repercussions on interactions, fostering sense of affiliation, engagement and sympathy [6]. It remains unclear, however, what the directionality is between behavioural coordination one one side, and effectiveness of the interaction and sense of affiliation on the other.

Research on deception has shown that this involves sophisticated socio-cognitive representations and skills involved in the understanding and manipulation of another's beliefs, as well as self- and other-monitoring [2]. It also requires extra attentional and control resources for the inhibition of a truth-bias and the production of a conflicting false response [3]. As such, deceptive interactions might display different patterns of entrainment than the ‘cooperative’ interactions normally studied. Hence, given the evidence that one partner's lexical choice shapes and adapt to another's during highly collaborative task-oriented dialogue, how is this trend affected if one of the partners is deceiving the other?

One possibility is that as alignment facilitates the effectiveness of the interaction, it could also facilitate deception through a tailored adaptation of the deceptive message to earn the partner's trust. As a result deceivers may show a higher degree of lexical entrainment than truth-tellers. Alternatively, deceivers may show a lower degree of lexical entrainment; deceivers may perceive themselves as distant from their dialogue partner, a form of disaffiliation that results in decreased alignment tendency. A third possibility goes back to the cognitive demanding nature of deceiving; the cognitive load induced by deception may hampers deceivers' attentional resources, hence their processing of

linguistic input, resulting in a lower degree of alignment in deceivers than truth-tellers.

We used a picture-naming paradigm, in which participants (N=46) were primed by a confederate with highly acceptable-yet-disfavored names for target pictured objects (e.g., “tumbler” for glass). Throughout 7 blocks of trials, participant and confederate took turns to name objects for each other and choose an object to match their partner's description. Every few trials, they did a memory test in which they tried to correctly identify objects they had so far seen but not selected during their matching turns. To perform well on the memory test, the correct objects had to be previously named by their partner. Participants in the **DECEPTIVE** condition were instructed to jeopardize their partner's performance on the memory test (and promised financial reward for it) by being cued to sometimes name the wrong object. We measured lexical entrainment as participants' tendency to re-use the disfavored names introduced by their partner for target pictures, under the assumption that participants were unlikely to use the disfavoured names unless these were previously used by their dialogue partner.

Mixed-effects analyses showed that participants deceiving their partner were less likely to lexically align (52.5%) than participants telling the truth (61.6%) (Beta = -.573, SE = .270, p = .033). Analysis of the time course of the alignment tendency in the **DECEPTIVE** and **TRUTHFUL** conditions revealed that deceptive participants tended to align less and less over the course of the interaction, whilst the alignment tendency of truthful participants remained stable over time (Beta = -1.58, SE = .067, p =.017). Deceivers were also less likely to perform correctly on the memory tests (17% vs. 42% correct responses) ($p < .0001$), a task that did not bear any direct connection to their deceptive act. These results do not only confirm that being deceptive is overall cognitive more taxing than being truthful, but they also suggest that it can affect agents' performance on concurrent secondary tasks, with deceivers being overall poorer performers.

Taken together, our results suggest that being deceptive disrupt the interactive process of lexical entrainment. Further investigation is needed in order to assess whether this is the result of the deceivers' sense of estrangement from their dialogue partner, which also led them to linguistically distance themselves, or instead, it is due to the cognitively and emotionally taxed nature of deceptive behaviour which hampers deceivers' attentional resources and thus prevents deceivers to engage in linguistic coordination dynamics. Deceivers' poorer performance on the memory tests seems to corroborate this second explanation.

[1] Pickering, M.J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behav. Brain Sci*. 27, 169–225.

[2] Duran, N. D. & Dale, R. (2012). Increased vigilance in monitoring others' mental states during deception. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (pp. 1518-1523). Austin: TX: Cognitive Science Society.

[3] Walczyk, J.J., Roper, K.S., Seeman, E., & Humphrey, A.M. (2003). Cognitive mechanisms underlying lying to questions: response time as a cue to deception. *Applied Cognitive Psychology,* 17, 755-774.

[4] Brennan, S.E., & Clark, H.H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 22(6), 1482–1493.

[5] Branigan, H. P., Pickering, M. J., Pearson, J., McLean, J. F., & Brown, A. (2011). The role of beliefs in lexical alignment: Evidence from dialogs with humans and computers. *Cognition, 121,* 41–57.

[6] Lakin, J.L., Chartrand, T.L., & Arkin, R.M. (2008). I am too just like you: Nonconscious mimicry as an automatic behavioral response to social exclusion. *Psychological Science, 19,* 816–822.

Listening tongues:
An EMA and EPG study on tongue movements during the perception of speech

Melanie Weirich[1]*, Susanne Fuchs*[2] *& Adrian P. Simpson*[1]
[1]*Institut für Germanistische Sprachwissenschaft, Friedrich-Schiller-Universität Jena*
Zentrum für Allgemeine Sprachwissenschaft, Berlin
melanie.weirich@uni-jena.de

Recent investigations in inter-personal coordination have found that humans can adapt their behavior to each other (hereafter convergence), for instance, over the course of a dialogue. One fundamental question arising from the pure observation of this phenomenon is how and why is this possible? The answers are manifold and mostly depend on the theoretical framework of a) how speech production, perception, and processing are linked, b) how flexible units in the lexicon are and what is stored (Goldinger, 1998), and c) what contribution the social relation between interlocutors plays (Giles & Coupland, 1991). In our studies we will focus on the first account. Yuen et al. (2010) showed on the basis of electropalatographic (EPG) data that articulatory movements are active during speech perception. The authors suggested that the articulatory information "is activated automatically and involuntarily in speech perception" (p. 595). Such a proposal is in agreement with Pulvermüller & Fadiga (2010) summarizing the link between action and perception at the neural level. It is also in agreement with work by Watkins & Paus (2004), Watkins et al. (2003) and Fadiga et al. (2002) who found increased excitability of the speech motor system during speech perception, even if no output is produced. Hickock (2010) argues strongly against the mandatory activation of motor areas during speech perception, especially because there is clinical evidence that in some cases speech perception can be completely intact although speech production may be disordered.

An EMA (Electromagnetic Articulograph) study was conducted analyzing listeners' tongue movements during the perception of speech. 5 speakers (2 male and 3 female) were recorded while listening to questions by the experimenter (e.g. "Did you look at Gabi or Gabbi?") which they were subsequently required to reply with a pre-defined answer ("I looked at GABI."). This question-answer-paradigm was used to trigger contrastive accent realizations (which are not the focus of this study). The speech material is part of a bigger corpus and contains the name /GVbi/ with the vowel being /i: ɪ e: ɛ a: a o: ɔ u:/ or /ʊ/. Each target word was repeated 10 times and while subjects were recorded during listening and speaking, here, we will concentrate on the tongue movement during *listening*.

Results reveal tongue movements during listening in alignment with the accented syllables (as shown in Figure 1) in 105/500 (10 vowels * 10 repetitions * 5 speaker) trials. There seems to be no strong influence of vowel, but most instances were found for the Giebi/Gibbi question (28/105). While the absolute amount is subject-specific (it ranges from 3 to 42/100), overall, the majority of instances of alignment of perceived accent and articulatory movements was found at the *beginning* of the experiment (over 50% during repetitions 1-3, cf. Figure 2). These findings point to a strong connection between speech production and perception, and the variation due to repetition number does not point to a reaction in terms of convergence. By contrast, it seems that the processing load plays a role, in the way that more tongue movement is found during a *new* incoming signal (with a higher load) than during a known repetition. Here, the possible role of learning comes to the fore.

Figure 1: Oscillogram, spectrogram and vertical (thin lines) and horizontal (thick lines) movement of tongue mid (yellow) and tongue back (green) for one subject during *listening* to the question "Sahst du Giebi oder Gibbi an".

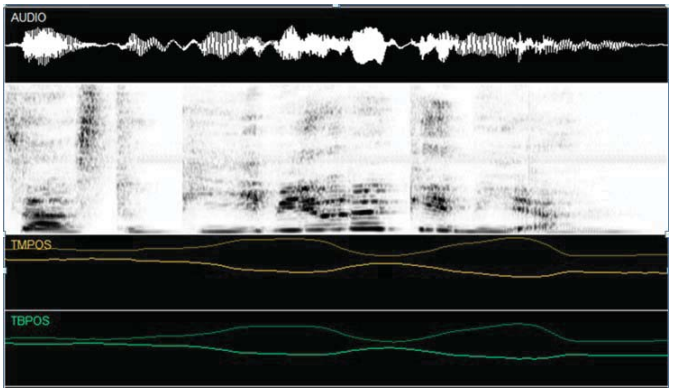
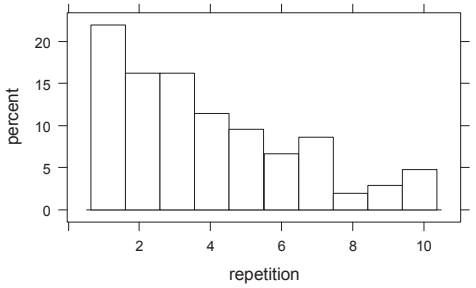


Figure 2: Distribution of occasions with aligned tongue movements during listening of all speakers (in percent) for repetitions 1 -10



However, a critical point of this EMA study is that listeners could see their answers and thus the target word while they were listening to the questions. A potential consequence is that participants compared what they heard with what they should read and this might have influenced the tongue movements. Therefore, a second experiment with EPG is currently being conducted. This second study addresses this drawback by changing the methodology: Subjects are asked questions with two words under contrastive focus as in the first study, however, this time they cannot read their answers. Instead, after listening to the question they are shown a number (1 or 2), which signifies what their answer should be (target word 1 or 2). In this way, the cognitive load is increased and participants have to concentrate on memorizing the two words and their order. If we still find tongue movement during listening, this cannot result from silent reading and comparing words but is rather only due to the perception of speech. The speech material comprises high vowels, alveolar stops and sibilants (i.e. sounds with tongue palate contact at the hard palate) and includes nonsense words and high frequency words. We predict we will find more tongue movement in the nonsense words (where the cognitive load is higher) if learning plays a role. Also, as in the EMA study, tongue movements should be more likely in the first repetitions than at the end.

References:

Fadiga, L., Craighero, L. Buccino, G. & Rizzolatti, G. (2002) Speech listening speci®cally modulates the excitability of tongue muscles: a TMS study. *European Journal of Neuroscience* 15, 399-402.

Giles, H. & Coupland, N. (1991). *Language: Contexts and Consequences.* Buckingham: Open University Press.

Goldinger, S. D.(1998).Echoes of echoes? An episodic theory of lexical access. *Psychol.Rev.* 105, 251-279.

Hickok, G. (2010) The role of mirror neurons in speech perception and action word semantics. *Language and Cognitive Processes* 25 (6), 749-776.

Pulvermüller, F. & Fadiga, L. (2010) Active perception: sensorimotor circuits as a cortical basis for language. *Nature Reviews Neuroscience* 11. 351-360.

Watkins, K. E. & Paus, T. (2004), Modulation of Motor Excitability during Speech Perception: The Role of Broca's Area. *Journal of Cognitive Neuroscience* 16 (6), 978-987.

Watkins, K. E., Strafella, A.P. & Paus, T. (2003) Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia* 41, 989-994.

Yuena, I., Davis, M. H., Brysbaerta, M. & Rastlea, K. (2010) Activation of articulatory information in speech perception. *PNAS* 107 (2), 592-597.

Coordination across task and event domains

Robert Fuhrman and Eric Vatikiotis-Bateson[1]
Linguistics, University of British Columbia, Vancouver, Canada

We report on two studies that use *correlation map analysis* (CMA) [1] assess the time-varying coordination associated with two types of communicative sound production: one musical, the other speech-based. One study focuses on the coordination *between* two Plains Cree Indians (Alberta, Canada) as they sing and drum (Fig. 1). Preliminary results show that coordination in the motor domain of producing sound differs from the auditory-acoustic domain of the produced sounds. The speech study examines coordination *within* talkers as they vary their vocal effort (e.g., loudness) in different speaking contexts (Fig. 2). Pilot results show that within-talker coordination increases with greater vocal effort, but there is a corresponding loss of coordination with the environment (based on postural force measures).

Cree drumming and singing

A father-son duo was video-recorded while singing Cree songs and playing the tom-toms (Fig. 1). 2D motion was extracted for the torso, head, and drum for each performer using *optical flow analysis* [2]. Fig. 1 shows sample *regions of interest* (ROI), within which image velocities are summed to create a time-varying signal for each ROI [3 for procedure]).

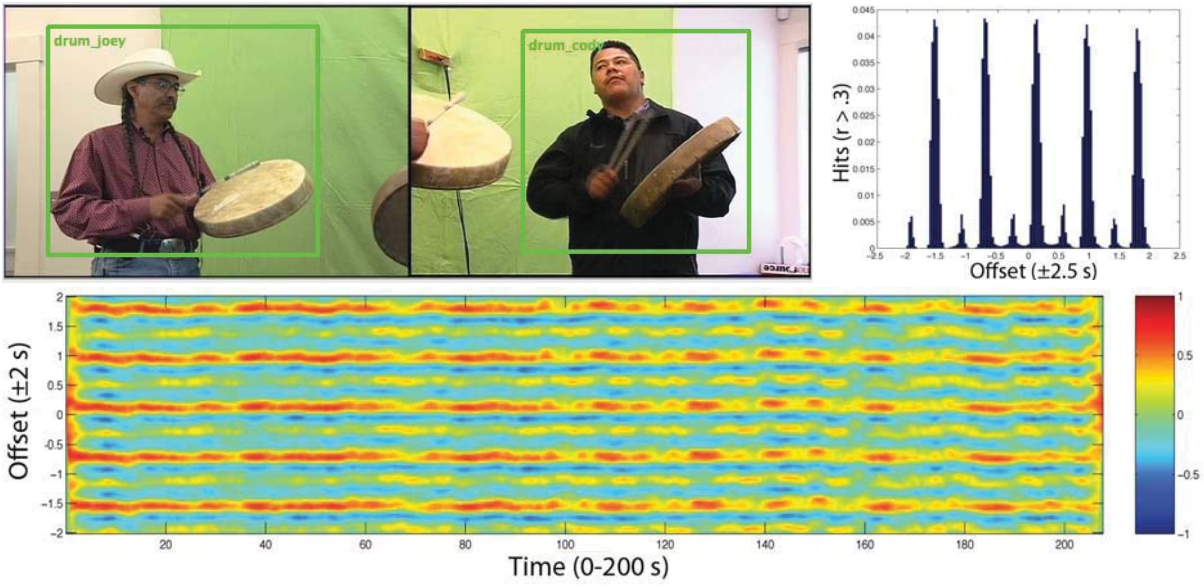


Fig. 1. *2D motion (velocity) of two drummers summed within ROI (box). Instantaneous correlation is plotted as a function of temporal offset. The histogram shows positive correlations (r > .3), the small peaks index the father's bimodal downstroke. The entire correlation map is plotted in the lower panel.*

The drummers use different stroke styles: the father pauses during his down-stroke while the son produces a unimodal downstroke. Despite this difference, the drummers' motions are continuously correlated through time (Fig. 1, bottom) at a mean temporal offset of about 130 ms. However, the acoustic onset of the drum beats is offset only 3-4 ms. Understanding the timing discrepancy between motor and acoustic events in this simple-to-observe setting may help us understand the more complex control and adaptation of speech timing.

[1] Corresponding author – email: evb@mail.ubc.ca

Vocal effort

An earlier study using singers reciting and singing song lyrics, in no way adapted to the speaking task, provided little support for the hypothesis that coordination of the body, head, and voice increases (and simplifies in terms of signal characteristics) as vocal effort increases[4]. Returning to the original pilot data in which the talker adapted her speech material to fit the task constraint – e.g., "you are shouting across a parking lot" – both the main hypothesis and its corollary (reduced coordination with the environment) were supported. As shown in Fig. 2, the proportion of positive correlations (r > .5) increases with vocal effort (from *soft* to *yell*) for all pairwise correlations except those involving *center of pressure* (COP) measured via force plates under each foot and a body, head, or acoustic (F0, RMS amplitude) measure. Two of the 4 correlations involving COP (COP x F0, COP x Body) reduce for at least the yell condition. Interestingly, despite the gradual decrease in correlation between COP and FO, there is no change in the temporal offset of the moments of high correlation (r > .5). At the meeting, we will present results for six additional talkers producing task-relevant speech.

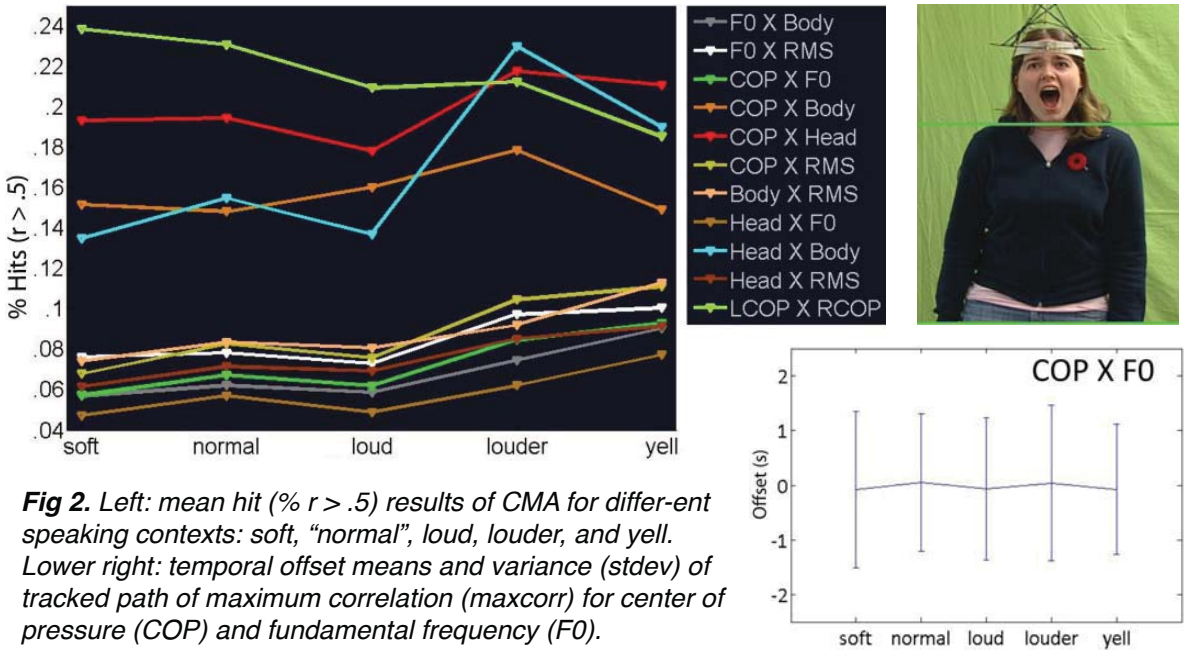


Fig 2. *Left: mean hit (% r > .5) results of CMA for differ-ent speaking contexts: soft, "normal", loud, louder, and yell. Lower right: temporal offset means and variance (stdev) of tracked path of maximum correlation (maxcorr) for center of pressure (COP) and fundamental frequency (F0).*

Bibliography

1 Barbosa, A. V., Dechaine, R.-M., Vatikiotis-Bateson, E. & Yehia, H. C. Quantifying time-varying coordination of multimodal speech signals using correlation map analysis. *The Journal of the Acoustical Society of America* **131**, 2162-2172 (2012).

2 Horn, B. K. P. & Schunk, B. G. Determining optical flow. *Artificial Intelligence* **17**, 185-203 (1981).

3 Barbosa, A. V., Yehia, H. C. & Vatikiotis-Bateson, E. in *Auditory and Visual Speech Processing -- AVSP08* (eds Roland Goecke, Patrick Lucey, & Simon Lucey) 173-177 (Causal Productions, 2008).

4 Vatikiotis-Bateson, E. *et al.* in *Proceedings of the 7th Triennial Conference of European Society for the Cognitive Sciences of Music (ESCOM 2009)* (ed Tuomas Eerola Jukka Louhivuori, Suvi Saarikallio, Tommi Himberg, Päivi-Sisko Eerola) 604-609 (2009).

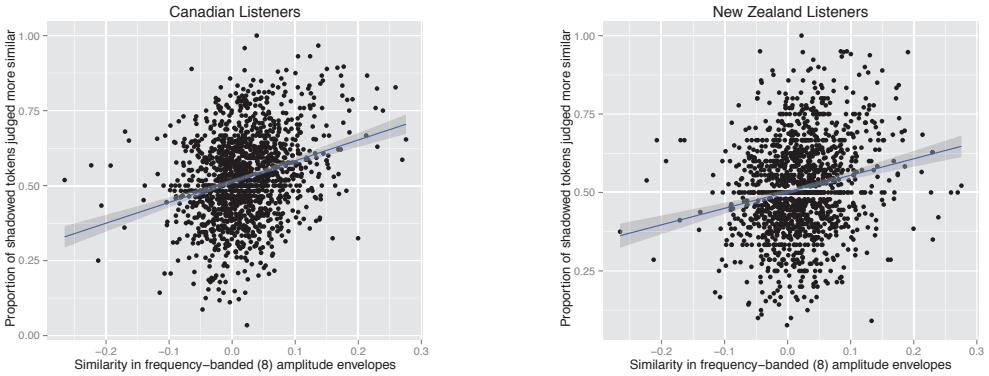
Spectral similarity and listener judgments of phonetic accommodation

A fundamental problem in studies of phonetic accommodation is knowing what to measure to properly assess the degree of convergence between interlocutors. Hand-selecting acoustic parameters (e.g., F1, F2, duration) may miss the relevant parameters. On the other hand, if accommodation is a strategy used in real-world communication or if it is has implications for sound change, the role of listener judgments of accommodation is critical. Thus, many researchers rely on listener judgments of similarity, typically implementing an AXB listening task. Listener-based experiments, however, are costly and time consuming. Given these issues, there is a strong desire in the field to have a reliable holistic acoustic measure of convergence. To date there has been some use of more general acoustic measures of similarity. For example, Delvaux and Soquet (2007) based their analysis on mel frequency cepstral coefficients (MFCCs), finding evidence for convergence. Lewandowski (2012) used a measure of spectral similarity that averaged cross-correlation values from amplitude envelopes in four logarithmically spaced frequency bands in a range of 80-7800Hz. This method also proved fruitful in her analysis, showing that participants with increased "phonetic talent" accommodated their interlocutor more.

No study, however, has compared a holistic measure of spectral similarity to listener judgments, nor has a study compared different measures of spectral similarity. In this paper we compare three types of measures of spectral similarity and their relationship to listener judgments across two data sets of single word productions from auditory naming tasks. To assess phonetic distance we used a dynamic time warping algorithm using (1) spectrograms and (2) 12 MFCCs derived from the spectra of those spectrograms. Both of these are asymmetric distance functions, meaning that we compare shadower-to-model distance and model-to-shadower distance. We also used (3) the amplitude envelope similarity measure from Lewandowski using 4 and 8 frequency bands. We intend to also test 16 frequency bands and ERB bands.

Our listener-based data sets are from Babel, McGuire, Walters, and Nicholls (in press) and Babel, McAuliffe, and Haber (2013). Babel et al. (in press) reports on single word accommodation in North American English (NAE) speaking shadowers and NAE-speaking AXB listeners. Babel et al. (2013) was a study of New Zealand shadowers who were presented with an Australian model – the AXB listeners in this study spoke Canadian English. We have also recruited AXB listeners (n=93) in New Zealand for these data; this offers the crucial comparison of whether the spectral measures better align with the NZ listeners or the Canadian listeners. Given that the spectral measures are linguistically naive, we predict that the spectral similarity measures will correlate more strongly with the responses from the Canadian listeners who are also linguistically naive with respect to what constitutes similarity in Antipodean dialects.

In the interest of space, we focus on the results comparing spectral measures to listener judgments for the NZ data from Babel et al. (2013), and highlight those spectral measures which also correlated with listeners' judgments with the NAE-speaking data set from Babel et al. (in press). First, we established that Canadian listeners' judgments were well correlated with NZ listeners' judgments [$t(1430) = 19.3$, $p < 0.001$, $r = 0.45$]. The 4-band and 8-band amplitude envelope similarity measures correlated with listener judgments from both populations with the 8-band analysis correlating more strongly: Canadian listeners and 8-bands [$t(1430) = 11.09$, $p < 0.001$, $r=0.28$], New Zealand listeners and 8-bands



(a) Canadian listeners' judgments of similarity. (b) New Zealand listeners' judgments of similarity.

[$t(1430) = 7.6$, $p < 0.001$, $r = 0.20$]. These results are shown in the figures above. A r-to-z Fisher transformation showed the correlation was stronger for the Canadian listeners ($p < 0.05$), perhaps due to linguistic naivety on the part of both the listeners and the spectral measures.

All spectral measures of similarity which correlated with listeners' judgments indicated convergence in their own right as well, while spectral measures which did not correlate with listener judgments did not (e.g., the MFCC analyses).

In short, not all of the acoustic measures of spectral similarity correlate with listeners' judgments of convergence, and those that do, do not correlate very strongly. This throws a wrench in the hopes of finding a singular spectral measure of similarity to replace the role of listeners in assessing convergence. Understanding what listeners attend to in experimental settings is of crucial importance not just for studies of phonetic accommodation, but also more generally for studies of speech. As experimenters move away from simple cue trading experiments to larger scale, more naturalistic studies, the production-to-perception mapping becomes more complex. Thus, understanding the usefulness of holistic measures of the speech signal is vital.

References

Babel, M., McAuliffe, M., & Haber, G. (2013). Can mergers-in-progress be unmerged in speech accommodation?, Frontiers in Psychology, 4, 653.

Babel, M., McGuire, G., Walters, S., & Nichols, A. (in press/2014). Novelty and Social Preference in Phonetic Accommodation. Journal of Laboratory Phonology.

Delvaux, V., & Soquet, A. (2007). The influence of ambient speech on adult speech productions through unintentional imitation. Phonetica, 64(2-3), 145-173.

Lewandowski, Natalie (2012): Talent in nonnative phonetic convergence. Dissertation. Universitt Stuttgart. http://elib.uni-stuttgart.de/opus/volltexte/2012/7402/

Convergence in bilingual rhythm

Current models of second language speech learning (Best & Tyler, 2007; Flege, 1995) account for cross-language variability based on vowel and consonant production, but to date we lack information on their applicability to prosodic units such as speech rhythm. By 'rhythm', we refer to how weak and strong units of prosodic structure are timed in speech. In this presentation we examine rhythm data for native speakers of Spanish (syllable-timed language) who are highly proficient learners of English (stress-timed language). Specifically, we ask if individual speakers converge, or become more similar, in the rhythm values they produce in their L1 and L2. There is evidence that bilingual speakers are influenced by their L1 in L2 rhythm production (White & Mattys, 2007), but we have little information on whether there are comparable L2 → L1 effects.

Fifteen speakers participated in a sentence reading task. Five NSs of Peninsular Spanish read aloud sentences in Spanish, five native speakers of American English read aloud sentences in English, and five Spanish-speaking L2 learners of English read aloud sentences in Spanish and English. The bilingual speakers were native speakers of Peninsular Spanish who are highly proficient L2 learners of English. The test sentences were taken from Arvaniti (2012), the most up-to-date study on the rhythm metrics in Spanish and English. Acoustic analysis was carried out in Praat (Boersma & Weenink, 2013). For each sentence, six rhythm metrics were calculated: %V, SD-V, Varco-V, nPVI-v, SD-C, and Varco-C. These metrics are commonly used in the experimental literature to quantify rhythm differences across languages (Arvaniti, 2009; 2012; Dellwo, 2006; Grabe & Low, 2002; Prieto et al., 2011; Ramus, 2002).

We highlight three noteworthy results. First, when comparing monolingual Spanish and English speaker groups, differences on all rhythm metrics were statistically significant, as expected (see Figures 1-6). Second, for all measures except SD-C, bilinguals show intermediate values between both monolingual control groups. In other words, their rhythm metrics showed statistical difference when compared to both monolingual groups. The implication is that bilingual L1 and L2 rhythms have converged into a shared phonological space. Third, for all measures except Varco-C, bilinguals show statistically separate values in their L1 and L2 rhythms. Follow-up individual analyses confirm this trend for all bilingual speakers.

In summary, Spanish-English bilinguals show converged, but separate rhythm values in their L1 and L2. The finding that bilingual speakers maintain a contrast in rhythm properties in English and Spanish suggests cross-language dissimilation (Flege, 1995). However, we also show that speakers display intermediate values between those of English and Spanish monolinguals in both languages, suggesting phonetic convergence in the bilingual grammar. To the best of our knowledge, this is the first study to indicate that bilingual speakers show bidirectional influences (i.e., mutual convergence) in their L1 and their L2 rhythms. Finally, we propose a working model of second language rhythm based on the notion of the 'phonetic syllable' (Dupoux, 2001) as a category within L2 speech learning.

FIGURES: ENGLISH (monolingual English speakers); BIL_ENGLISH (Bilingual speakers in English); BIL_SPANISH (Bilingual speakers in Spanish); SPANISH (monolingual Spanish speakers

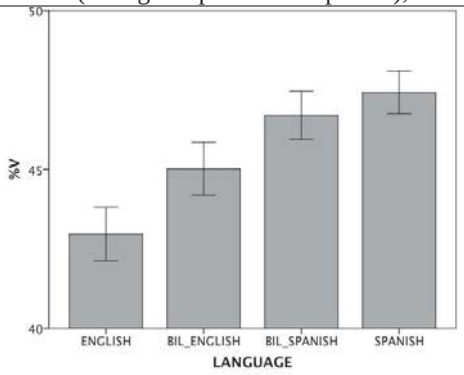


Fig. 1: %V across four language groups

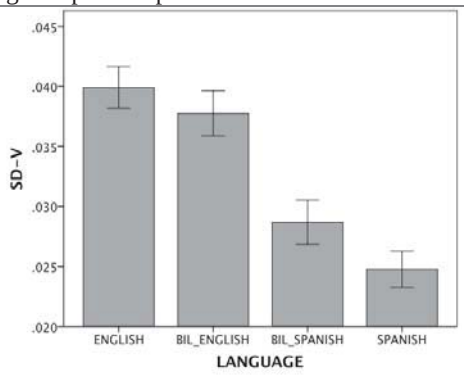


Fig. 2: SD-V across four language groups

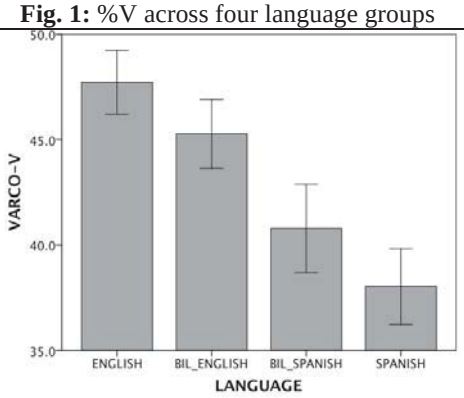


Fig 3: *Varco-V across four language groups*

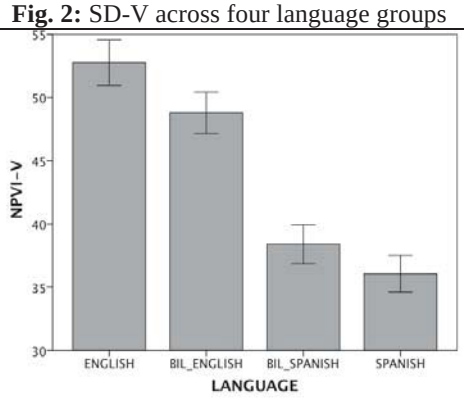


Fig. 4: nPVI-V across four language groups

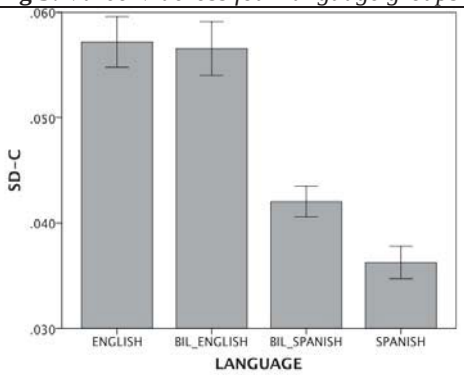


Fig. 5: SD-C across four language groups

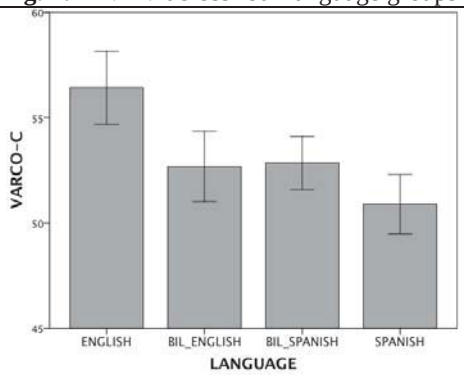


Fig. 6: Varco-C across four language groups

SELECTED REFERENCES: ARVANITI, A. (2009). Rhythm, timing and the timing of rhythm. *Phonetica* (66), 46-63. ARVANITI, A. (2012). The usefulness of metrics in the quantification of speech rhythm. *Journal of Phonetics* 40: 351-373. DELLWO, V. (2006). Rhythm and Speech Rate: A Variation Coefficient for deltaC. *Language and language-processing*, 231-241. GRABE, E. & LOW, E.L. (2002). Durational Variability in Speech and the Rhythm Class Hypothesis. *Papers in Laboratory Phonology* (7). RAMUS, F. (2002). Acoustic correlates of linguistic rhythm: Perspectives. *Proceedings of Speech Prosody*, Aix-en-Provence (11), 115-120. WHITE, L., & MATTYS, S.L. (2007). Calibrating rhythm: first language and second language studies. *Journal of Phonetics (*35), 501-522.

Do people converge to the linguistic patterns of non-reliable speakers?

Perceptual learning from non-native speakers

Shiri Lev-Ari[1,2] and Sharon Peperkamp[1]

[1]Laboratoire de Sciences Cognitives et Psycholinguistique (ENS, EHESS, CNRS)

[2]Max Planck Institute for Psycholinguistics

People's linguistic representations are shaped by the input from the environment. Exposure to speech with certain characteristics can thus influence individuals' later perception both when processing the speech of the same speaker (Norris, McQueen & Cutler, 2003) and when processing the speech of another speaker (Kraljic & Samuel, 2007). Not all speakers and situations, however, are equally reliable and representative, and individuals seem to be sensitive to this variation when learning from the environment. Thus, listeners do not change their phonological representations in accordance with input provided by a speaker holding a pen in her mouth even though they change their representations after exposure to the exact same tokens when the speaker does not seem to have any obstruents in her mouth (Kraljic, Brennan & Samuel, 2008). The question we investigate in this study is whether listeners similarly adjust their representation in accordance with input provided by reliable, but not by unreliable speakers, and in particular, whether listeners' representations are therefore influenced by the speech of native, but not by the speech of non-native speakers.

One-hundred-fifty-nine native French speakers selected pictures according to recorded instructions in French provided by a native French speaker or a native Dutch speaker. Half of the participants in each speaker condition heard /b/s whose Voice Onset Times were imperceptibly modified, and unmodified /p/s. The other half heard imperceptibly modified /p/s and unmodified /b/s. Dutch and French are both short-lag languages, and the two speakers' unmodified VOTs did not differ from one another. Following the picture selection task, participants performed a phoneme categorization task with either the same

speaker or with a new native French speaker. Results show that only participants in both conditions showed adaptation when tested with the same speaker but only those who listened to the native speaker generalized their learning to a new speaker.

These results indicate that people are indeed sensitive to the reliability of the input and do not adapt their representations according to input that is perceived to be unreliable. The results also provide evidence for a disassociation between adaptation to the characteristic of specific speakers and adjustment of linguistic representations in general based on these learned characteristics, as participants were able to adapt to specific speakers without generalizing the learning.

This study also has implications for theories of language change, as theories of substratum language influence propose that a large proportion of non-native speakers in a community can bring about linguistic changes. The results of this study constrain this possibility.

References

Kraljic, T. & Samuel, A.G. (2007). Perceptual adjustments to multiple speakers. *Journal of Memory and Language, 56, 1-15*.

Kraljic, T., Brennan, S.E., & Samuel, A.G. (2008). Accommodating variation: Dialects, idiolects, and speech processing. *Cognition, 107, 1*, 51-81.

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology, 47, 2*, 204-238.

Interlocutors' speech rates converge: The effects of fast and slow confederate speech rates

Benjamin G. Schultz[1], Irena O'Brien[1,2], Natalie Phillips[2,3], David McFarland[2,4], Deborah Titone[1,2], & Caroline Palmer[1,2]

[1]*McGill University,* [2]*Centre for Research on Brain, Language, and Music,* [3]*Concordia University,* [4]*University of Montreal*

INTRODUCTION

When interlocutors engage in speech with one another, they often change their behavior to conform to the behaviors of the partner or group (Giles, 1977). This is called *accommodation*. It has been shown that the speech of conversational partners can converge in accent (Giles, 1977), pitch (Bosshardt, Sappok, Knipschild, & Holscher, 1997), and intensity (Huber, 2008). The convergence of speech rate within conversations, however, has received less attention (but see Jungers, Palmer, & Speer, 2002 for speech rate priming from audio recordings). The present study examined the convergence of speech rate in scripted dialogues read by a confederate and participants. The confederate spoke at either a fast (6 syllables per second) or slow (3 syllables per second) rate.

A beat tracking algorithm (Ellis, 2007) was applied to acoustic data to measure speech rate through the dependent variable *inter-beat interval* (IBI), defined as the temporal interval between successive stressed syllables. Syllable stress was measured as energy summed across frequency bands that are perceptually salient to humans.

METHOD

Participants

McGill University undergraduate students (N = 38) with North American English as their first language were recruited. Participants (P) had a mean age of 20.7 years (SD = 2.48, *range* = 18-30 years), and 35 participants were female (three male). The Confederate (C) was a female, McGill University undergraduate of 21 years who spoke North American English. None of the Participants were aware that the Confederate was a Confederate as indicated by verbal questioning of the Participant at the end of the experiment.

Stimuli

Two scripts were used for the dialogues: excerpts from *Death of a Salesman* (Arthur Miller, 1949) and *The Importance of Being Earnest* (Oscar Wilde, 1908). Half of the Ps received *Death of a Salesman* in the Fast condition, and *The Importance of Being Earnest* in the Slow condition, and the other half received the reverse. The C always produced the first utterance of the dialogue.

Procedure

The P was introduced to the C under the guise that the confederate was another participant. The P then read the two scripted dialogues with the C in the two confederate speech rate conditions (Fast and Slow). For each of the Speed conditions, the C was trained to a metronome rate prior to speaking with the P. The C entered the room with a caffeinated energy drink for the Fast condition to provide a cover story for the different speech rate.

RESULTS & DISCUSSION

Effects of confederate speech rate

To test that Ps' speech rates were faster in the Fast condition compared to the Slow condition, a repeated-measures ANOVA was conducted on mean IBI values (collapsed across utterances) with Role (P, C) and Speed (Fast, Slow) as within-subjects variables. There were significant main effects of Role, Speed, and a significant interaction (ps < .01). Pair-wise comparisons between Fast and Slow conditions (see Figure 1a) for the P and C indicated that IBIs were smaller (i.e., faster) in the Fast condition compared to the Slow condition for Ps and the C (ps < .01). This result suggests that the speech rate of the C influenced the speech rate of the Ps.

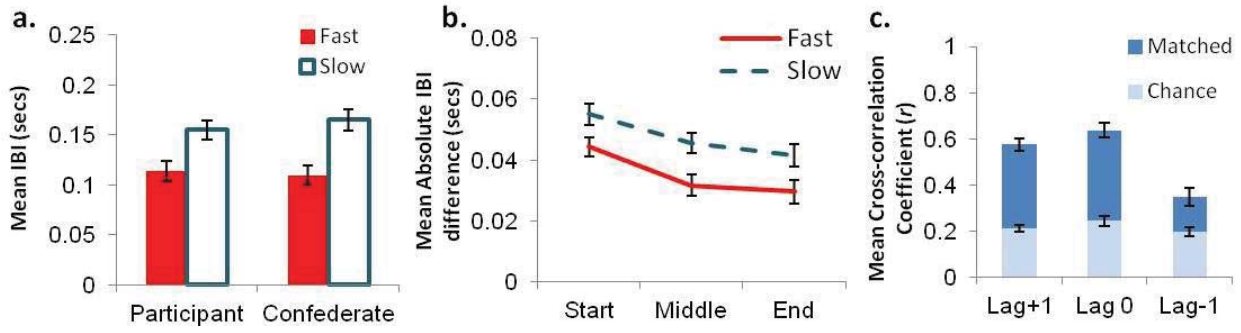


Figure 1. a) Mean inter-eat interval (IBI) across utterances for the Participant (P) and Confederate (C) in Fast and Slow conditions. b) Mean absolute difference between the IBI of the P and C for Fast and Slow conditions for the first (Start), Middle, and final (End) eight utterances. c) Mean cross-correlation coefficients between IBIs of the P and C at Lags +1, 0, and -1. Chance levels were calculated using jackknifing techniques. Error bars represent standard error of the mean.

Convergence of speech rate

To test that speech rates converged during the dialogue, the absolute difference between the IBI of the P and C for each utterance was calculated. The mean absolute IBI difference for the first (Start), Middle, and final (End) eight utterances were calculated to create the variable Section. There were significant main effects of Speed and Section (*ps* < .01) and no significant interaction (p = .88). As shown in Figure 1b, absolute IBI differences decreased over Sections for both Fast and Slow Speed conditions, suggesting that the speech rate of the P and C converged during the dialogue.

Cross-correlational analyses

To test the influence of the P and C on each other's speech rate, cross-correlational (XC) analyses were conducted on the IBIs in each dialogue. Lag 0 was the XC of C's n^{th} utterance with P's n^{th} utterance. Lag +1 was the XC of the C's n^{th} utterance with the P's n^{th}+1 utterance. Lag -1 was the XC of the C's n^{th} utterance with the P's n^{th}-1 utterance. Lags 0 and +1 represented the influence of the C's speech rate on P's, and Lag -1 represented the influence of the P's speech rate on the C's. As shown in Figure 1c, Lags 0 and +1 were significantly higher than Lag -1 (*ps* < .001). Chance levels were estimated using jackknifing techniques (Quenouille, 1949). Lags +1, 0, and -1 were all significantly greater than chance (*ps* < .01). These results suggest that the C had a greater influence on the speech rate of Ps than *vice-versa*, but that Ps still had a significant influence over the speech rate of the C.

CONCLUSIONS

This experiment tested whether speech rate, as measured by a beat tracking algorithm, converges during a scripted conversation with a confederate. Results showed that speech rate converged, thus supporting the *speech accommodation theory* (Giles, 1977). Furthermore, there was evidence for mutual adaptation of speech rate; there was a bidirectional influence of the confederate's and participants' speech rates. Results are interpreted through the dynamic attending theory (Jones & Boltz, 1989) that posits that partners within a joint task should synchronize and adapt to each other's rates.

REFERENCES

Bosshardt, H.-G., Sappok, C., Knipschild, M., & Holscher, C. (1997). Spontaneous imitation of fundamental frequency and speech rate by nonstutterers and stutterers. *Journal of Psycholinguistic Research, 26,* 425-448. Ellis, D. P. W. (2007). Beat tracking by dynamic programming. *Journal of New Music Research, 36,* 51-60. Giles, H. (Ed) (1977). *Language, ethnicity, and intergroup relations*. London: Academic Press. Huber, J. E. (2008). Effects of utterance length and vocal loudness on speech breathing in older adults. *Respiratory Physiology & Neurobiology, 164,* 323–330. Jones, M. R. & Boltz, M. (1989). Dynamic attending and responses to time. *Psychological Review, 96,* 459–491. Jungers, M.K., Palmer, C., & Speer, S.R. (2002). Time after time: The coordinating influence of tempo in music and speech. *Cognitive Processing, 1-2,* 21-35. Quenouille, M. (1949). Approximate tests of correlation in time series. *Journal of the Royal Statistical Society, Series B, 11,* 68-84. Miller, A. (1949). *Death of a Salesman*. New York: The Viking Press. Wilde, O. (1908). *Collected Works of Oscar Wilde*. MA: Methuen Press.

Degrees of control over influencing factors in phonetic convergence

Natalie Lewandowski & Antje Schweitzer
Institut für Maschinelle Sprachverarbeitung
Universität Stuttgart

Phonetic convergence, the increase in similarity between two speakers' pronunciations during a dialogical interaction (Pardo, 2006), has been mostly researched from two particular perspectives – Communication Accommodation Theory (CAT; Shepard et al., 2001) and the process-oriented view of the Interactive Alignment Account (Pickering & Garrod, 2004). While the first suggests a high degree of influence the speakers themselves can exert during the dialog, the latter assumes the mechanism of convergence to proceed largely automatically, with little room for influence on the part of the speaker herself.
CAT proposes social aspects of the communicative encounter to bear the most important influence on the direction of the adaptation. The desire to reduce or increase social distance for various reasons is said to steer the changes either into a convergent or a divergent mode. Differences in social roles during the interaction (as in interviewer/interviewee), social status discrepancies (employer/employee), as well as the need to stress group membership (e.g. through dialect or language choice) have been proposed to drive these changes. The opposite view, according to which adaptation is biologically-grounded, assumes that there is no possibility to interfere with the process once started.

In our view both accounts should in fact be merged into a hybrid model of convergence where automatic processes and those more susceptible to (sub)conscious influence are combined. Furthermore, not only social and situational aspects, but also the speakers' personalities and cognitive abilities should be included in the framework. We present data on speaker adaptation from a study in second language convergence (German native speakers engaged in English conversations with English native speakers; Lewandowski, 2012) and from a corpus of German spontaneous conversations (GECO) between previously unacquainted talkers (Schweitzer & Lewandowski, 2013).

The analyses performed on the latter revealed a positive correlation of the speakers' mutual likeability ratings with convergence of their speaking rate. The more the participants liked each other, the more similar their speaking rates were. The likeability ratings were collected after every conversation and consisted of a range of questions concerning the speaking partner (e.g. how friendly, nice and sociable she seemed). This is in line with other studies in which mutual liking has an important bearing on adaptation phenomena (e.g. Abrego-Collier et al., 2011).

Convergence between an early and late point in the English-German L2 dialogs was measured with amplitude envelopes at word level and were correlated with the speakers' pronunciation talent ratings (obtained in an earlier study; Jilka, 2009). Phonetically gifted speakers proved to be significantly better at converging toward their speaking partners. The lack of accommodation on the part of the less talented speakers though, did not mean that they maintained their own pronunciation style throughout the conversation. Comparisons of their own productions from different points in the dialogs suggests that their pronunciation shifted as well, albeit not in the direction toward their English partners. Furthermore, the native English speakers, who were asked to

maintain their own pronunciation and not align with their German partners, also showed significant positive shifts in their speaking style instead (Lewandowski, 2012). The conscious attempt to stop convergence (or, in an ideal case, not to let it happen at all) seems to have failed. These two results speak for the existence of an automatic component, which, however, is not sufficient to explain all the variation arising within a conversational interaction.

Further tests on the dataset revealed that the personality factors of openness, agreeableness, and the ability to switch attention in a fast and accurate way, correlate positively with the amount of convergence measured in the course of the dialogs. All these results taken together allow us to suggest a hybrid model of convergence, with many layers of factors dynamically interacting to shape the amount of convergence within a conversation (Figure 1).

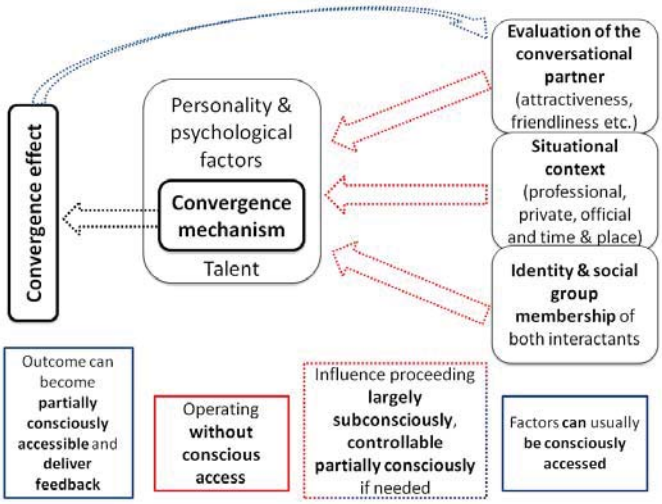


Figure 1. A hybrid model of convergence with the degree of control over different factors (adapted from Lewandowski, 2012).

A speaker's talent and personality are proposed to be located at the core of the mechanism, with no possibility of conscious change or influence. The factors on the right side, including the evaluation of the speaking partner and the situational context, are usually consciously accessible and can well impact the adaptation. So can the effects of the accommodation process, which the talker may become conscious of after producing adapted speech. The model thus assumes an automatically initiated convergence mechanism, which can be nevertheless altered by extrinsic and intrinsic influences in the dialog setting.

References

Jilka, M., 2009. Assessment of Phonetic Ability. In: Language Talent and Brain Activity, G. Dogil and S. Reiterer, Eds. Mouton De Gruyter, Berlin, 17 – 66.

Lewandowski, N., 2012. Talent in nonnative phonetic convergence. Ph.D. thesis, Universität Stuttgart.

Pardo, J. S., 2006. On phonetic convergence during conversational interaction. In: Journal of the Acoustical Society of America 119 (4), 2382–2393.

Pickering, J. M., Garrod, M., 2004. Toward a mechanistic psychology of dialogue. In: Behavioral and Brain Sciences 27 (2), 169–190.

Schweitzer, A., Lewandowski, N., 2013. Convergence of articulation rate in spontaneous speech. In: Proceedings of Interspeech 2013, Lyon, 525–529.

Shepard C. A., Giles H., Le Poire B. A. ,2001. Communication Accommodation Theory. In: The New Handbook of Language and Social Psychology, Eds. Robinson W. P., Giles H., 33–56.

Modulation of visible and non-visible articulatory movements in perturbed face-to-face communication

Maëva Garnier[1], Gabrielle Richard[2], Lucie Ménard[2]

[1]GIPSA-Lab, Département Parole et Cognition, UMR CNRS 5216 & Grenoble Universités, France
[2]Laboratoire de phonétique, Université du Québec à Montréal, Canada

Introduction

On one hand, it is now well known that seeing speech improves its perception, especially when speech is degraded by a noisy background [1]. On the other hand, some studies have shown that speakers adapt their speech production in noisy conditions. This adaptation, also called the « Lombard effect », mainly consists in talking louder and at higher pitch [2]. However, it is also accompanied by other speech modifications, such as increased amplitude and speed of lip articulation [3-4]. This raises the question of whether this hyper-articulation observed in Lombard speech can be considered as a communicative strategy to improve visual intelligibility.
This study aims at bringing elements of answers, by examining whether, in noise:
- speakers enhance significantly more their visible articulatory movements when their speech partner can see them compared to when the partner can only hear them.
- all the articulatory movements are enhanced similarly, or if the most visible ones (lips, jaw) are more enhanced than the others (tongue).

Material and Method

Six French Canadian speakers were recorded while speaking in a quiet environment and in a cocktail-party noise of 85 dB played over loudspeakers. Three conditions of interaction were examined: (S1) No Interaction: The speaker read sentences aloud. (S2) Audio Only (AO): The speaker gave instructions to the experimenter who was standing at a writing board placed 2m in front of him and the back to him. (S3) Audio Visual (AV): The experimenter was standing at the same place as before, this time facing the speaker. Seven target-words were selected: /pap/, /pip/, /pup/, /pɛp/, /map/, /tap/, /nap/. They were produced in the carrying sentence « le mot ___ me plaît » (*I like the word* ___) and repeated ten times in each condition. In the two interactive conditions (S2 and S3), the speaker chose freely the order of production of the 70 sentences, so that the experimenter could not predict the target-word. The audio signal was recorded synchronously with the 3D movements of the lips, the jaw and the tongue, using electromagnetic articulography (Carstens AG 500).

Results

The results confirmed that all the speech modifications from a quiet to a noisy situation (i.e. the increase of both voice intensity and fundamental frequency, the amplification of tongue and lip movements) are significantly greater when speakers interact with a speech partner (S2 and S3), compared to when they only read sentences aloud (S1).
The comparison of AO and AV interactive conditions did not show such a common tendency across all the speakers. Instead, the behaviours observed can be distinguished in three strategies:
- Three speakers showed the behaviour that we expected at first: they increased vocal intensity more or comparably in the AO condition, compared to the AV condition. On the

contrary, they enhanced their very visible articulatory movements (lip aperture, spreading, closure and protrusion, and jaw aperture) more in the AV condition, when they could be seen by their interlocutor, compared to the AO condition. However, they did not enhance their less visible tongue movements more in AV condition.
-Two speakers showed another behaviour: they did not demonstrate any significant difference in speech adaptation between the AO and AV conditions, for both acoustic and articulatory modifications.
-Finally one speaker showed an opposite behaviour to that we expected [5]: he enhanced both visible and non visible articulatory movements more in the AO condition than in the AV conditions, following the same tendency as the increase of his vocal effort.
For all vowels, the tongue was lower and more forward in the noisy condition compared to the quiet one. Tongue displacements were significantly amplified, with increased speed for the tip of the tongue and reduced speed at its root.

Conclusions

The results obtained from these speakers support the hypothesis that speakers can modulate their production of visible cues in adaptation to the perceptual modalities of interaction. However, they also show that such a specific enhancement of visible speech cues may be made by some speakers only, and not by all individuals.
Evidences were brought that some speakers enhance their visible articulatory movements more in adverse conditions when they can be seen by their interlocutor, and that this hyper-articulation does not necessarily correlate with vocal intensity or with non-visible tongue articulation.
For the other speakers who do not seem to use the visual channel to improve their intelligibility, increasing voice intensity appears to be the main strategy to compensate for the perturbation of intelligibility – which is greater in AO interaction than in AV interaction. For these speakers, the amplification of all articulatory movements, regardless of their visibility, may be just a consequence of an increased vocal effort.
Complementary analyses will be conducted: Indeed, the enhancement of visible articulatory movements from quiet to noise was greater in AO than in AV condition for some speakers, and greater in AV than in AO condition for others. However, when considering these movements in noise only – and not their enhancement from quiet to noise – they were almost always greater in the AO condition, compared to the AV condition. Therefore it seems necessary to take account of potential ceiling effects and articulatory limitations in the interpretation of the adaptations observed.

References

[1]Sumby, H. and Pollack, I. W. (1954). "Visual Contribution toSpeech Intelligibility in Noise." Journal of the Acoustic Society of America **26**: 212-215.

[2]Junqua, J. (1993). "The lombard reflex and its role on human listener and automatic speech recognizers." Journal of the Acoustic Society of America **93**(1): 510-524.

[3]Garnier, M., Henrich, N. and Dubois, D. (2010). "Influence of Sound Immersion and Communicative Interaction on the Lombard Effect." Journal of Speech, Language and Hearing Research **53**(3): 588-608.

[4]Garnier, M. (2008). "May speech modifications in noise contribute to enhance audio-visible cues to segment perception?" Proceedings of AVSP, Tangalooma, Australia.

[5]Garnier, M., Richard, G. and Ménard, L. (2012). "Do speakers make use of the visual channel to improve their intelligibility in adverse conditions? A pilot study." Proceedings of Interspeech, Portland, USA.