ISICS 2012

Book of Abstracts

Aix-en-Provence
3-5 September 2012

International Symposium on Imitation and Convergence in Speech

ISICS 2012: International Symposium on Imitation and Convergence in Speech
Aix-en-Provence, 3-5 September 2012

In the course of a conversational interaction, the behavior of each talker often tends to become more similar to that of the conversational partner. Such convergence effects have been shown to manifest themselves under many different forms, which include posture, body movements, facial expressions, and speech. Imitative speech behavior is a phenomenon that may be actively exploited by talkers to facilitate their conversational exchange. It occurs, by definition, within a social interaction, but has consequences for language that extend well beyond the temporal limits of that interaction. It has been suggested that imitation plays an important role in speech development and may also form one of the key mechanisms that underlie the emergence and evolution of human languages. The behavioral tendency shown by humans to imitate others may be connected at the brain level with the presence of mirror neurons, whose discovery has raised important issues about the role that these neurons may fulfill in many different domains, from sensorimotor integration to the understanding of others' behavior.

The focus of this international symposium is the fast-growing body of research on convergence phenomena between speakers in speech. The symposium also aims to assess current research on the brain and cognitive underpinnings of imitative behavior. Our main goal is to bring together researchers with a large variety of scientific backgrounds (linguistics, speech sciences, psycholinguistics, experimental sociolinguistics, neurosciences, cognitive sciences) with a view to improving our understanding of the role of imitation in the production, comprehension and acquisition of spoken language.

The symposium is organized by the Laboratoire Parole et Langage, CNRS and Aix-Marseille Université, Aix-en-Provence, France, with the financial support of Aix-Marseille Université, the CNRS, the Ville d'Aix-en-Provence, the Conseil Général des Bouches-du-Rhône, and the Région Provence-Alpes-Côte d'Azur.

ISICS 2012 is held under the auspices of the Brain and Language Research Institute at Aix-Marseille Université. It is an event supported by the International Speech Communication Association, ISCA.

ISICS 2012 is co-chaired by Noël Nguyen (LPL) and Marc Sato (GIPSA-Lab, Grenoble).

Local Organizing Committee

Noël Nguyen	LPL, CNRS & Univ. d'Aix-Marseille	Aix-en-Provence
Nadéra Bureau	LPL, CNRS & Univ. d'Aix-Marseille	Aix-en-Provence
Stéphanie Desous	LPL, CNRS & Univ. d'Aix-Marseille	Aix-en-Provence
Sophie Dufour	LPL, CNRS & Univ. d'Aix-Marseille	Aix-en-Provence
Amandine Michelas	LPL, CNRS & Univ. d'Aix-Marseille	Aix-en-Provence
Nadia Monségu	LPL, CNRS & Univ. d'Aix-Marseille	Aix-en-Provence

Scientific Committee

Patti Adank	University of Manchester	Manchester	UK
Martine Adda-Decker	LIMSI, CNRS	Paris	France
Gérard Bailly	GIPSA-Lab, CNRS	Grenoble	France
Roxane Bertrand	LPL, CNRS & Univ. d'Aix-Marseille	Aix-en-Provence	France
Mariapaola D'Imperio	LPL, CNRS & Univ. d'Aix-Marseille	Aix-en-Provence	France
Sophie Dufour	LPL, CNRS & Univ. d'Aix-Marseille	Aix-en-Provence	France
Carol Fowler	Haskins Laboratories, Univ. Connecticut	New Haven, CT	USA
Jonathan Harrington	University of Munich	Munich	Allemagne
Jennifer Hay	University of Canterbury	Christchurch	New-Zealand
Julia Hirschberg	Columbia University	New York	USA
Holger Mitterer	Max Plank Institute	Nijmegen	Pays-Bas
Lorenza Mondada	ICAR, CNRS & Université de Lyon II	Lyon	France
Kuniko Nielsen	Oakland University	Rochester, MI	USA
Noël Nguyen	LPL, CNRS & Univ. d'Aix-Marseille	Aix-en-Provence	France
Martin Pickering	University of Edinburgh	Edinburgh	UK
Marc Sato	GIPSA-Lab, CNRS	Grenoble	France
Jean-Luc Schwartz	GIPSA-Lab, CNRS	Grenoble	France
Véronique Traverso	ICAR, CNRS & Université de Lyon II	Lyon	France
Sophie Wauquier	SFL, CNRS & Université Paris 8	Paris	France

ISICS 2012: International Symposium on Imitation and Convergence in Speech

Aix-en-Provence, 3-5 September 2012

MONDAY 3 SEPTEMBER

08:30-09:00 Registration

09:00-09:30 Opening Session

Keynote Session 1 - Chair: Marc Sato

09:30-10:30 *The motor contribution to speech perception: neurophysiological and modeling approaches*
Alessandro D'Ausilio

10:30-11:00 Coffee break

Oral Session 1 - Chair: Patti Adank

11:00-11:30 *Phonetic convergence in shadowed speech: phonological neighborhoods, individual differences, and the relation between acoustic and perceptual measures*
Jennifer Pardo, Kelly Jordan, Rolliene Mallari, Caitlin Scanlon, Eva Lewandowski

11:30-12:00 *Original objective and subjective characterization of phonetic convergence*
Amélie Lelong, Gérard Bailly

12:00-12:30 *A multicomponent approach to phonetic convergence*
Natalie Lewandowski

12:30-13:00 *Assessing phonetic compliance*
Véronique Delvaux, Kathy Huet, Myriam Piccaluga, Bernard Harmegnies

13:00-14:30 Lunch break

Poster Session 1

14:30-16:00 *Can a spoken dialog system be used as a tool to study convergence?*
Jose Lopes, Andrew Fandrianto, Maxine Eskenazi, Isabel Trancoso

Individual differences in phonetic convergence
Alan Yu, Carissa Abrego-Collier, Morgan Sonderegger

Quantification of speech convergence through non-linear methods for the analysis of time-series
Leonardo Lancia, Susanne Fuchs, Amélie Rochet-Capellan

The role of visual cues in imitating a new sound
Nancy Ward, Megha Sundara

Sensory-motor maps of speech: functional coupling and sensory-motor predictions during vowel perception and production
Krystyna Grabski, Marc Sato

Bio-acoustic fingerprints of individual differences in bilingual speech imitation ability: neuroimaging and spectral analysis
Susanne Reiterer, Xiaochen Hu, Nandini Singh

The sound of your lips: haptic information speeds up the neural processing of auditory speech
Avril Treille, Camille Cordeboeuf, Coriandre Vilain, Marc Sato

Synchrony and convergence of pause lengths in spontaneous conversation
Kristina Lundholm Fors

Convergence of laughter in conversational speech: effects of quantity, temporal alignment and imitation
Jürgen Trouvain, Khiet Truong

Speech imitation between speakers influences the realization of initial rises in French intonation
Amandine Michelas, Noël Nguyen

The bilingual advantage in phonological learning
Laura Spinu, Yulia Kondratenko

Oral Session 2 - Chair: Gérard Bailly

16:00-16:30 *Functional causality of the dorsal stream in sensorimotor integration of speech repetition*
Takenobu Murakami, Yoshikazu Ugawa, Ulf Ziemann

16:30-17:00 *Perceptually induced speech motor representations*
Chris Neufeld, Radu Craioveanu, Frank Rudzicz, Willy Wong, Pascal Van Lieshout

17:00-17:30 *Plasticity of sensory-motor goals in speech production: behavioral evidence from phonetic convergence and speech imitation*
Marc Sato, Krystyna Grabski, Maëva Garnier, Lionel Granjon, Jean-Luc Schwartz, Noël Nguyen

17:30-18:00 *Role of motor representations in perception and imitation of singing*
Yohana Lévêque, Antoine Giovanni, Daniele Schön

19:00 Reception at the Aix-en-Provence City Hall

TUESDAY 4 SEPTEMBER

Keynote Session 2 - Chair: Mariapaola D'Imperio

09:00-10:00 *Prosodic matching as a sequential resource in naturally occurring interaction*
Beatrice Szczepek Reed

10:00-10:30 Coffee break

Oral Session 3 - Chair: Marie Postma

10:30-11:00 *Prosodic structuring imitation in French L1 context - A first step towards correcting phonetic-prosodic features in L2 French*
Olivier Nocaudie, Corine Astésano

11:00-11:30 *Perceptual learning and convergence in sound change*
Bridget Smith

11:30-12:00 *Unmerging mergers-in-progress through spontaneous phonetic imitation*
Molly Babel, Michael McAuliffe, Graham Haber

12:00-12:30 *Effects of direct dialect imitation on tonal alignment in two Southern varieties of Italian*
Mariapaola D'Imperio, Rossana Cavone, Caterina Petrone

12:30-14:00 Lunch break

Poster Session 2

14:00-15:30 *Effects of imitative training techniques on L2 production and perception*
Ewa Wanat, Rachel Smith, Tamara Rathcke

Discursive convergence in conversation
Mathilde Guardiola, Roxane Bertrand

Accommodation of backchannels in spontaneous speech
Antje Schweitzer, Natalie Lewandowski

Effects of dialect/language interference and memory on direct imitation of German question intonation
Caterina Petrone, Leonardo Lancia, Cristel Portes

Convergence, complementarity, co-ordination: partner-specific effects in the emergence of procedural conventions
Gregory Mills

Imitation, convergence, and phonological learning: a study of bilingual and monolingual patterns in the reproduction of word-final stop realizations in novel accents of English
Laura Spinu, Jiwon Hwang

Accommodation and sociolinguistic meaning: phonetic after-effects of being and interacting with a (dis)engaged interviewer
Kodi Weatherholtz, Abby Walker, Kathryn Campbell-Kibler

Interpersonal long-term phonetic accommodation-patterns in close acquaintances
Yshai Kalmanovitch

Developing sound categories in adult language learners' imitated and read speech
Terhi Peltola, Pertti Palo

Convergence in talk-in-interaction across languages: multilingual format tying in peer talk-in-interaction
Gudrun Ziegler, Natalia Durus, Neiloufar Family

Does the interpreter converge with the speaker? Analysing prosodic characteristics of simultaneously interpreted texts
George Christodoulides, Anne Catherine Simon

Prosodic convergence and divergence: the building of coherence and shared meaning in conversational dialogues
Li-Chiung Yang, Shu-Chuan Tseng

Oral Session 4 - Chair: Jennifer Pardo

15:30-16:00 *Auditory perception bias affects F0 imitation*
Marie Postma, Eric Postma

16:00-16:30 *Effects of seeing and hearing vowels on neonatal facial imitation*
Marion Coulon, Cherhazad Hemimou, Arlette Streri

16:30-17:00 *Articulations or preconceptions? An investigation of visual speech alignment findings*
Kauyumari Sanchez

17:00-17:30 *Breathing changes during listening and subsequent speech according to the speaker and the loudness level*
Amélie Rochet-Capellan, Susanne Fuchs, Leonardo Lancia, Pascal Perrier

Keynote Session 3 - Chair: Noël Nguyen

17:45-18:45 *Mechanisms of speech adaptation*
Maëva Garnier

20:00 Banquet

WEDNESDAY 5 SEPTEMBER

Keynote Session 4 - Chair: Jürgen Trouvain

09:00-10:00 *Mechanisms for interactive alignment during conversation*
Simon Garrod

10:00-10:30 Coffee break

Oral Session 5

10:30-11:00 *Vocal imitation positively affects language attitudes*
Patti Adank, Andrew Stewart, Louise Connell

11:00-11:30 *Other-repetition: displaying others' lexical choices as "commentable"*
Mathilde Guardiola, Roxane Bertrand, Sylvie Bruxelles, Carole Étienne, Emilie Jouin-Chardon, Florence Oloff, Béatrice Priego-Valverde, Véronique Traverso

11:30-12:00 *The temporal dynamics of alignment in multimodal interaction*
Bert Oben, Geert Brône, Kurt Feyaerts

12:00-12:30 *The influence of gender stereotype threat on speech accommodation in same & mixed-gender negotiations*
Lauren Aguilar

12:30-13:00 Closing Session

The motor contribution to speech perception: neurophysiological and modeling approaches

Alessandro D'Ausilio[1]*, Leonardo Badino*[1]*, & Luciano Fadiga*[1,2]

[1]Italian Institute of Technology (IIT). Robotics, Brain and Cognitive Sciences Department (RBCS). Mirror Neurons and Interaction Lab.
[2]University of Ferrara. DSBTA - Section of Human Physiology.
alessandro.dausilio@iit.it,leonardo.badino@iit.it,luciano.fadiga@iit.it

Classical models of language consider an antero-posterior distinction between perceptive and productive functions. In the last 15 years, this dichotomy has been weakened because of empirical evidence suggesting a more integrated view. Passive listening to phonemes and syllables activate motor and premotor areas. These activations were somatotopically organized according to the effector recruited in the production of these phonemes. However, a feature of action-perception-theories is that motor areas are considered necessary for perception. In fact, it has been argued that in absence of a stringent determination of a causal role played by motor areas in speech perception, no final conclusion can be drawn in support of motor theories of speech perception. The mere activation of motor areas during listening to speech might be caused by a corollary cortico-cortical connection that has nothing to do with the process of comprehension itself (Fadiga et al., 2002). A possible solution might come from the selective alteration of neural activity in speech motor centers and the evaluation of effects on perception. Therefore, we designed a series of TMS experiments to tackle the causal contribution of motor areas to speech perception (D'Ausilio et al., 2009). We demonstrated that activity in the motor system is causally related to the discrimination of speech sounds and might be more critical under adverse listening conditions or when coping with inter-speaker variability. Interestingly, this functional association is somatotopically organized according to an effector-sound motor map. Listening to reproducible speech sounds might activate the same motor gestures necessary for production and thus help sensory classification and decision. This process of matching the actions of others onto our own sensorimotor repertoire is thought to be important for action recognition in general, providing a non-mediated "motor perception" based on a bidirectional flow of information along the mirror parieto-frontal circuits. Computational models that use state-of-the-art machine learning techniques for hand actions execution/observation have shown that when sensorimotor data are available during learning, visually presented actions are subsequently identified significantly better than when systems learn actions on the basis of visual information only. Since speech is a particular type of action (with acoustic targets) it is expected to activate a mirror neurons mechanism. Indeed, automatic phonetic classification under adverse listening conditions, is significantly improved when motor data are used during training of classifiers as opposed to learning from purely auditory data (Castellini et al., 2011).

References

Fadiga, L., Craighero, L., Buccino, G., & Rizzolatti, G. 2002. Speech listening specifically modulates the excitability of tongue muscles: a TMS study. *European Journal of Neuroscience* 15, 399–402.

D'Ausilio. A., Pulvermüller, F., Salmas, P., Bufalari, I., Begliomini, C., & Fadiga L. 2009. The motor somatotopy of speech perception. *Current Biology* 19, 381–385.

Castellini, C., Badino, L., Metta, G., Sandini, G., Tavella, M., Grimaldi, M., & Fadiga, L. 2011. The use of phonetic motor invariants can improve automatic phoneme discrimination. *PLoS One* 6, e24055.

Phonetic Convergence in Shadowed Speech: Phonological Neighborhoods, Individual Differences, and the Relation Between Acoustic and Perceptual Measures

Jennifer S. Pardo, Kelly Jordan, Rolliene Mallari, Caitlin Scanlon, & Eva Lewandowski

Department of Psychology, Montclair State University
pardoj@mail.montclair.edu

Phonetic convergence occurs both when individuals interact in conversation and when listeners rapidly repeat words presented over headphones (e.g., Goldinger, 1998; Pardo, 2006). Previous studies have found that characteristics of phonological neighbors also influence both the perception and production of words. Words from high-density neighborhoods are relatively harder to perceive, requiring greater phonetic resolution (Bradlow & Pisoni, 1999). Furthermore, Munson & Solomon (2004) found that words from high-density neighborhoods were produced with greater vowel expansion than words from low-density neighborhoods. The current study examined the influence of phonological neighborhood on phonetic convergence in a speech shadowing task. If high-density words are produced with more extreme vowel formants and their phonetic details are more highly resolved, then shadowers should converge to high-density words more than low-density words, especially with respect to vowel formants. A set of talkers produced target words that varied in frequency and frequency-weighted neighbor density independently. Another set of talkers produced baseline and shadowed tokens of the target words produced by the first set of talkers. Separate listeners judged the perceptual similarity of shadowed to model tokens in an AXB task designed to assess phonetic convergence. Finally, measures of inter-talker distances in vowel formants for baseline and shadowed speech were compared to the perceptual measures.

The results revealed large individual differences in the relationship between lexical and indexical influences on speech perception and production. Measures of vowel formants replicated the effects reported by Munson and Solomon—low frequency words and words from high-density neighborhoods were produced with more expanded vowels. However, there were no effects of phonological characteristics on perceived phonetic convergence or inter-talker distances in vowel formant frequency. Rather, individual shadowers showed unique patterns of convergence that were not systematically related to phonological characteristics. Moreover, acoustic measures of inter-talker similarity in vowel formants were unrelated to perceptual judgments of phonetic convergence. These findings have important methodological and theoretical implications for understanding the complexities of phonetic convergence. Studies of convergence should not rely solely on measures of single physical dimensions and should take into account individual differences. Lexical factors impact speech production and perception, but their effects appear to be independent of those that evoke phonetic convergence.

References

Bradlow, A. R., & Pisoni, D. B. 1999. Recognition of spoken words by native and non-native listeners: Talker-, listener-, and item-related factors. *Journal of the Acoustical Society of America* 106, 2074–2085.

Goldinger, S. D. 1998. Echoes of echoes? An episodic theory of lexical access. *Psychological Review* 105, 251–279.

Munson, B., & Solomon, N. P. 2004. The effect of phonological neighborhood density on vowel articulation. *Journal of Speech, Language, and Hearing Research* 47, 1048–1058.

Pardo, J. S. 2006. On phonetic convergence during conversational interaction. *Journal of the Acoustical Society of America* 119, 2382–2393.

Original objective and subjective characterization of phonetic convergence

Amélie Lelong and Gérard Bailly

GIPSA-lab, UMR 5216 CNRS/INPG/UJF/U. Stendhal
{amelie.lelong,gerard.bailly}@gipsa-lab.grenoble-inp.fr

Introduction

Individuals accommodate their communication behavior either by increasing similarity with their interlocutors (i.e. convergence) or on the contrary by increasing their differences (i.e. divergence). Speech accommodation has been observed at both linguistic and non linguistic levels. Several studies have been conducted on phonetic dimensions such as pitch, speech rate, loudness or dispersions of vocalic targets with various experimental paradigms ranging from close-shadows of prerecorded stimuli to more ecological face- to-face conversations.
Multiple objective and subjective characterizations of phonetic convergence have been proposed. This paper discusses limitations of current proposals, notably in terms of top-down strategies that may be used by labelers and listeners when characterizing/perceiving the stimuli. We put forward and evaluate here two novel techniques: objective characterization by speaker recognition techniques and subjective characterization by a novel paradigm named "speaker switching".
We will illustrate these techniques with stimuli collected during an original experimental paradigm called verbal dominoes (**?**), a speech game that can be played by several interlocutors and consisting in chaining rhyming words.

Objective characterization

Objective characterizations of convergence between two audio stimuli often involve the calculation of distances or correlations between time-aligned patterns. These characterizations are thus bounded to an a priori segmentation and labeling of relevant segments of interest, ranging from specific phonetic events (**?**) to whole words (**??**). Distribution of various phonetic cues – VOT, formant frequencies, spectral tilts, durations of segments, etc. – are then collected in these segments and compared. The identification, segmentation and labeling of segments of interest may provide interesting insights in phonological (i.e. cross-categorical) vs. phonetic (i.e. intra-categorical) accommodation issues. However this distinction is often neglected and difficult to disentangle – particularly in studies involving dialectal variations (see for example **?** – by manual as well as automatic procedures. On the contrary, speaker recognition techniques often consider a global characterization of the phonetic space of each speaker without any a priori knowledge on the phonological variants used by speakers. We have demonstrated (**?**) that GMM-based speaker recognition scores correlates significantly with a more detailed analysis of the distributions of speaker-specific vocalic spaces. The correlation increases with the corpus size: we have found a significant correlation of .66 ($p < 0.01$) for the two objective measures of convergence in the case of large chains of 350 dominoes.

Subjective characterization

The AXB test introduced by **?** is the most widely used test for subjective characterization of phonetic convergence: listeners hear three versions of the same lexical items and judge which item produced by one talker, A or B, "sounds like a better imitation of" or "is more similar to" (**?**) the X item produced by another talker. A reduced perceptual distance between X and one of the A or B items is then interpreted as a convergence of A/B towards X. These results are often significant nevertheless, the size of the effects reported in the literature are often small with a preference for items produced in interaction with X rather than the ones produced with no interaction around 60% (e.g. reading). We tested real and synthetic convergence (created by adaptive synthesis with the harmonic plus noise model interpolating parameters at 0% and 20% between both speakers). Our conclusions about AXB tests are very disappointing. Subjects had trouble to remember A when hearing B even in the easiest case (e.g. when contrasting items with objective convergence rates of 0% versus 20%). This led them to develop strategies unrelated to the task – such as focusing on prosodic variations or background noises – to ease decision. The final results mirror this difficulty (see Figure 1). We have recently tested a novel

perceptual test that we named *speaker switching*[1]. This test consists of generating a continuous signal where we randomly switch between items uttered by two speakers in different conditions, e.g. in isolation, imitating or interacting with one another. The listeners' task is simply to press a key each time they perceive/suspect a speaker switch. We considered that a switch was detected when the key hit occurred between the onset of the current item and the onset of the next. We experimentally set the ISI at 1000 ms. This is a rather rapid but comfortable presentation rate that favors immediate on-line processing and provides much more information and control data than the AXB test. We report preliminary results of two *speaker switching* experiments (see Figures 2 & 3) where we switched between 4 conditions: items read in isolation by two speakers and items uttered by the same speakers during a domino game. The stimuli are the same as for the AXB test. All subjects reported that this task was much easier than the AXB decision task.

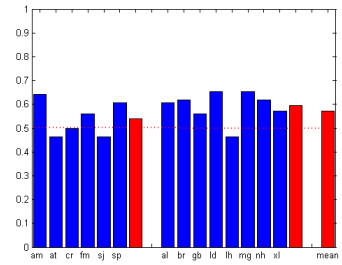


Figure 1. Results of an AXB test with 12 listeners. X are dominoes pronounced by speaker 1 and synthesized at 20% to speaker 2. A and B are the same dominoes pronounced by speaker 2 either synthesized at 0% or 20% to speaker 1. Only the last 8 listeners knew the speakers. Despite a large objective phonetic convergence, listeners have difficulties in rating closeness (ratings close to 50%)

Synthetic data: simulation of a symmetrical 20% convergence between speakers

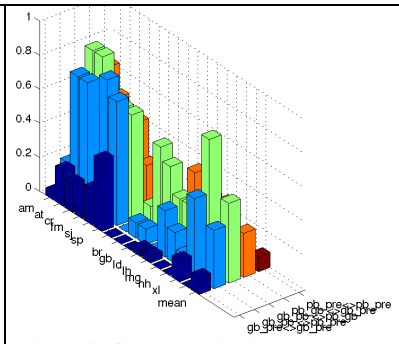


Figure 2. Speaker switching. Percentage of false detections for various transitions. Data from six listeners knowing the two speakers are displayed front. Data from other five listeners not knowing the speakers displayed at the back exhibit more confusion but display similar behavior: interaction increases misdetections.

Real data with a convergence between speakers close to 20%

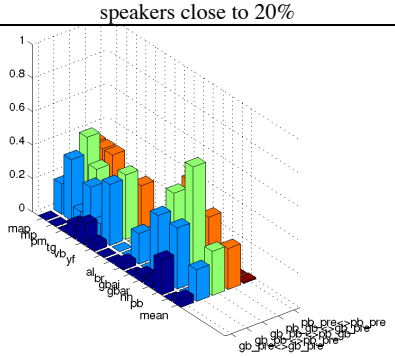


Figure 3. Same as Figure 2 but for real data. From left to right between: read items from speaker 1, interactive speech from speaker 1 and read items of speaker 2, interactive speech of both speakers, interactive speech from speaker 2 and read items of speaker 1 and read items from speaker 2. Data from six listeners knowing the two speakers are displayed front. Data from other six listeners not knowing the speakers displayed at the back.

References

Aubanel, V., & Nguyen, N. 2010. Automatic recognition of regional phonological variation in conversational interaction. *Speech Communication* 52, 577–586.

Delvaux, V., & Soquet, A. 2007. The influence of ambient speech on adult speech productions through unintentional imitation. *Phonetica* 64, 145–173.

Fowler, C. A., Sramko, V., Ostry, D. J., Rowland, S. A., & Hallé, P. 2008. Cross language phonetic influences on the speech of French-English bilinguals. *Journal of Phonetics* 36, 649–663.

Goldinger, S. D. 1998. Echoes of echoes? An episodic theory of lexical access. *Psychological Review* 105, 251–279.

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

Lelong, A., & Bailly, G. 2011. Study of the phenomenon of phonetic convergence thanks to speech dominoes In: Esposito, A., Vinciarelli, A., Vicsi, K., Pelachaud, C., & Nijholt, A. (eds), *Analysis of Verbal and Nonverbal Communication and Enactment: The Processing Issue*. Berlin: Springer Verlag, 280–293.

Lelong, A., & Bailly, G. 2012. Characterizing phonetic convergence with speaker recognition techniques. *Listening Talker Workshop*, Edinburgh.

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, 422–432.

[1]We acknowledge Jason A. Shaw from UWS for his fruitful suggestions concerning this idea.

A multicomponent approach to phonetic convergence

Natalie Lewandowski

Institute for Natural Language Processing, University of Stuttgart
natalie.lewandowski@ims.uni-stuttgart.de

Phonetic convergence is the conversational phenomenon of two speakers becoming more alike in terms of their segmental and suprasegmental pronunciation (Pardo, 2006). A controversial issue in convergence research concerns the social motivation vs. mechanistic nature of the accommodation in dialog. Although proposals for mixed models have also been made in the literature, most studies still contrast between automatic alignment processes and models which highlight that a speakers' convergence or divergence is consciously influenced, a.o. by social factors (as e.g., Communication Accommodation Theory).

The current study of talker behavior in native-nonnative dialogs sheds more light onto these highly discussed issues (Lewandowski, 2012). It fosters the assumption of a *hybrid model of convergence* with multiple operating mechanisms and pathways, and further strengthens the distinction between pure (conscious) *imitation* and largely subconsciously occurring *convergence* which appears to be tied to dialogic interaction.

Phonetic convergence was analyzed using an objective acoustic measurement – the comparison of *amplitude envelope signals* (Wade et al., 2010) at word level. Twenty speakers of German (the nonnative speakers, NNS) were involved in dialogs with two English native speakers (NS). The NNS were classified as either phonetically talented or less talented in a comprehensive test preceding the current study (Jilka, 2009). The dialogs were elicited by a picture matching game – the Diapix (Van Engen et al., 2010). The sessions were both preceded and followed by a simple word reading task, containing a.o. target words from the Diapix tasks. In addition to that, every German participant (the NNS) was asked to deliver a summary of the picture game in the end of each dialog (monologue). The German participants were not informed about the goal of the experiment, while both English native speakers were explicitly asked to maintain their own pronunciation style in order to avoid adapting to the nonnative speakers' accents in any way.

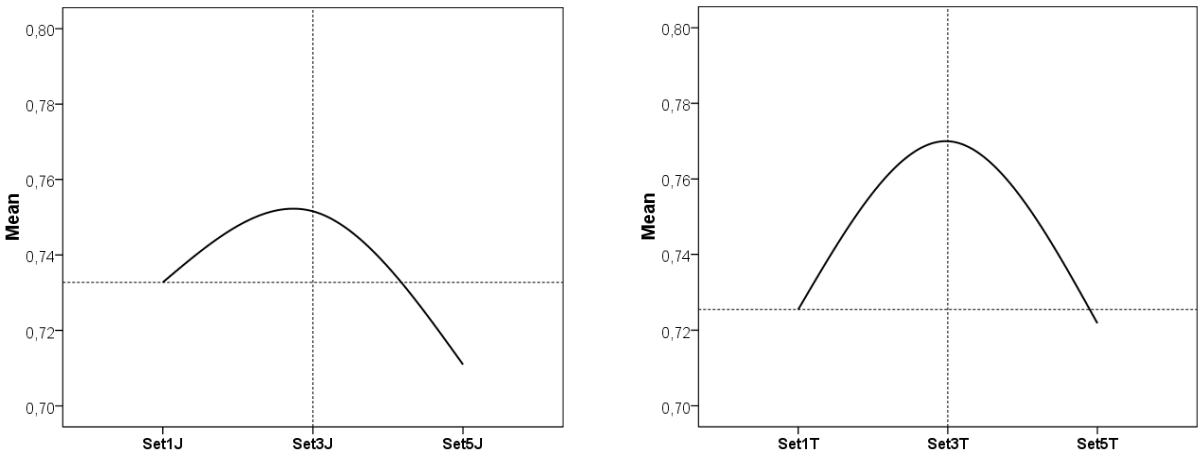
The acoustic measurement was done at word level, by extracting the target words and other frequently used content words at three time intervals within the dialogs – *early* (corresponding to the first third of the dialog), *late* (corresponding to the last third) and *summary*. The acoustic signal was converted to amplitude envelopes which represent a smoothed picture of the energy present in the underlying frequency bands and are said to contain crucial information used by listeners in speech perception (Wade et al., 2010). This allowed a global analysis without the necessity of pinning down any specific features where convergence could surface. The amplitude envelopes were first compared *between* dialog partners to reveal the dynamic patterns of mutual adaptation for the three extraction times – the convergence measurement (e.g., word *X* spoken by speaker *A* compared to another utterance of word *X* by the respective dialog partner *B* for different dialog times). Additionally, a *self-consistency* measurement of the talker's own pronunciation variation within the experiment was introduced (e.g., word *X* of speaker *A* compared to another utterance of word *X* also by speaker *A* across different dialog times) – monitoring how close to his or her own pronunciation a speaker stayed during the dialog.

It was confirmed that the German speakers (NNS) displayed phonetic convergence toward the English native speakers between an *early* and a *late* point in the dialogs ($p < .01$ for both conditions). It could also be shown that talent is significantly influencing the nonnative speakers' convergence – talented speakers converged more than less talented ones. An unexpected result was the significant convergence of the English NS toward the NNS ($p < .01$ in both conditions, $t = 3.166$ and $t = 4.342$), despite the clear instruction to restrain from any accommodative tendencies. Furthermore, the native speakers' adaptation happened without their conscious knowledge, since both participants were rather positive about not having altered their way of speaking. A look into the self-consistency data of the two NNS talent groups points to another novel fact: the less talented group who converged significantly less in the dialogs, still showed a comparable "perturbed" pattern of *self-*

consistency as the talented participants (no significant effect for the factor talent for the two NS conditions, ANOVA $p = .511$ and $p = .960$). This indicates that the less gifted talkers' pronunciation changed within the dialog, however, not in the direction of their conversational partners. It might thus be interpreted as a (subconscious) tendency to converge which could not be accomplished due to a lack of skills in the foreign language, resulting in a failure to approach the dialog partner?s pronunciation targets.

Further analyses of the summary and read speech pre- and post-tests seem indicative of a strong encapsulation of the convergence displayed within the dialogs. Neither the measurements of the read speech data nor, even more surprisingly, of the summaries produced by the NNS showed significant signs of convergence. The patterns in Figure 1a and 1b reveal a significant decrease in convergence between a late point and the summary ($p = .000$), which delimit the switch from *dialog* to *monologue*. The accommodation displayed within the dialog did not carry over to the narrative immediately following.

Figure 1a and 1b. The nonnative speakers' convergence towards both NS (J and T) in the dialog, compared for the times: early (Set1), late (Set3) and the summary (Set5). Set 5 includes a comparison of NNS summary items with NS late items. The Y axis shows the mean match values, the closer to 1 the more similar the compared amplitude envelopes.



The NS behavior and the NNS self-consistency data are convincing indicators for automatic, subconciously operating tendencies in phonetic convergence, with talent only influencing the eventual success of accommodation but not initializing the underlying mechanism. The subconscious nature of the basic mechanism stands in clear contrast to consciously controlled imitation. The relatively high variance in speaker behavior that cannot be accounted for solely by talent, however, still suggests that other factors (a.o. contextual and social) shape the *amount* of convergence displayed. The lack of carry-over effects from a dialogic to monologic style indicates that convergence here is limited to dialog (= to having a *dialog partner*) and phenomena observed within experiments on isolated word repetition or read speech might in fact have different underlying mechanisms.

References

Jilka, M. 2009. Assessment of phonetic ability. In: Dogil, G., & Reiterer, S. M. (eds), *Language Talent and Brain Activity*. Berlin: De Gruyter, 17–66.

Lewandowski, N. 2012. Talent in nonnative phonetic convergence. Dissertation. Universität Stuttgart.

Pardo, J. S. 2006. On phonetic convergence during conversational interaction. *Journal of the Acoustical Society of America* 119, 2382–2393.

Van Engen, K. J., et al. 2010. The Wildcat Corpus of Native-and Foreign-accented English: Communicative Efficiency across Conversational Dyads with Varying Language Alignment Profiles. *Language and Speech* 53, 510–540.

Wade, T., et al. 2010. Syllable frequency effects in a context-sensitive segment production model. *Journal of Phonetics* 38, 227–239.

Assessing phonetic compliance

Véronique Delvaux, Kathy Huet, Myriam Piccaluga, Bernard Harmegnies

Laboratoire de Phonétique, Université de Mons-Hainaut, Mons, Belgique
Veronique.Delvaux@umons.ac.be

This study is part of a broader project investigating the processes involved in the acquisition of new phonetic control regimes in L2 learning. Our main experimental paradigm consists in attempting to "shape" the speakers' productions in new ways based on what they are made to hear. Quite unsurprisingly, our first results ended up in a large amount of inter-individual variation, even though the participants were matched in several respects including linguistic experience. This is in agreement with evidence deriving from many studies focused on the "external factors" potentially determining individual variation in L2 phonetic learning, such as age of learning, age of arrival, or amount of L2 use (among others: Piske et al., 2001; Moyer, 2004). It is also to be linked with the results of studies focused on the effects of characteristics inherent to the learner, either in the cognitive domain (Golestani & Zatorre, 2009; Francis & Nusbaum, 2002), or in the broader psychological domain (motivation, personality factors, etc.) (Dörnyei, 2009; Cohen & Macaro, 2007). Moreover, there is a growing body of evidence in favor of the existence of a specific ability to produce and perceive foreign speech sounds, called "phonetic talent" (Jilka et al., 2007; Dogil & Reiterer, 2009). In its strict sense, phonetic talent denotes an innate, neurobiologically grounded, individual skill which is part of general language aptitude, but may be separated from other specific linguistic skills such as grammatical talent in L2. Although phonetic talent is an appealing concept, its objective assessment is hindered by the difficulty to depart between initial "talent" and the other interacting variables that have presided to each individual's language development and may still influence his productions in, e.g. an experimental imitation task. Also, in differential psychology, a separation is clearly made between "gift", an untrained and spontaneously expressed superior natural ability, and "talent", that progressively emerges from the transformation of this high aptitude into a well-trained and systematically developed skill. Moreover, to our knowledge, the only attempts reported in the literature to measure individual intrinsic abilities in dealing with foreign sounds are exclusively based on subjective (perceptual) data from native speakers of a different language to the individual's L1. Given these conceptual instabilities and methodological weaknesses, we will adopt here a pragmatic view, with no strong hypothesis about innateness. We then focus on the end result of the process, i.e. on the spontaneous ability of adult speakers to accurately produce speech sounds similar to models they are faced with. We call this ability "phonetic compliance". We posit that it varies among individuals and can be assessed in terms of gradient. This paper is a methodological account aiming both at studying the feasibility of phonetic compliance assessment (through 3 different quantification methods) and at preparing further research oriented towards a better understanding of the underlying mechanisms.

For this experiment, 10 Belgian French speakers (5 male, 5 female) have been submitted to an imitation task (instructions: "repeat as faithfully as possible, as if it was a sound from a foreign language") of 94 synthetic vowels regularly spaced in a Mel-scale F1*F2*F3 space (6 repetitions). They also produced 10 realizations of each of the 10 French oral vowels. Formant frequencies have been detected using Praat under the supervision of 2 trained phoneticians. Basically, the similarity between models given to the speaker and productions imitating the models can be estimated by the deviations of the productions from the targets. For a given speaker in an experiment with **S** vocoïd models, **P** productions of each stimulus, and provided that for each vowel, **I** (F_i) formants are taken into account for each production, the sum of the Euclidean distances target/realization for all stimuli and all productions can express globally the success of the speaker in the task. In formula 1, I1 tends towards zero when the productions of the speaker tend, globally, towards the target in the vocalic space. In other words, I1 is zero when compliance is maximal. Formula 2 is based upon the same principle, except that in this case, the *inverse* of the distance (-1/2 exponent) has been taken into account, in order to obtain a number with variations positively correlated with compliance. Furthermore, in this case, the speaker has been calibrated using his/her realizations of L1 vowels. The language has **V** vocalic phonemes and the speaker has realized **v** tokens of each. It is therefore possible to identify zones of the vowel space corresponding to usual productions of the speakers, and zones where he/she is not used to produce vocalic sounds. The idea in formula 2 is to give

higher reward to the success in imitating when imitation takes place in a region of the vocalic space the speaker does not use in his/her current practice of his/her L1. This is the reason for the weighting by the multiplicative term. It consists in the logarithm of the product of all the distances between a given production and each vowel's cluster centroïd: the multiplicative term tends toward zero when at least one distance production/centroïd tends toward zero. Thus, for a given realization, the product is large if the production resembles the target *and* if it is produced in a zone far from the ones corresponding with the speaker's L1. I2 may be viewed as a modulation of phonetic compliance assessment through individual L1 phonology. In formula 3 (where "**var**" stands for "variance"), the similarity between target and production is no more the main point, and the approach is more statistical: it is based upon the analysis of variability in the imitation task. When a speaker tries to imitate a model, he/she produces realizations that fall around it in the reference space. If the speaker's compliance is high, his/her variability around the model in the reference space is random, and if no other source of variance is active, the variability is constant whatever the stimulus. On the other hand, if the speaker is strongly influenced by his/her L1, one can suppose that his/her variability will vary from one stimulus to another, depending on whether the stimulus is close or not to a region of the vowel space present in L1. I3 should therefore tend toward zero (all variances equal) in a speaker with good compliance. Formula 3 uses individual variability in phonetic processing as a source of information on phonetic compliance. The 3 formulae have been currently applied to the productions of 4 speakers. Results show both convergence and specificities in how the indices characterize phonetic compliance. S4 is the most compliant since his productions are the closest to the models (I1), the furthest apart from his typical L1 realizations (I2), and the most homogenous in their variation (I3) (see Fig.1 for an illustration). S1 and S2 are the less compliant speakers, S2 being slightly better at leaving her L1 territory (I2) but, doing so, increasing greatly the variation of the variances associated with her realizations of the different models (I3). Based on the results collected on 10 speakers, we will discuss at the conference how to integrate the complementary information given by the 3 indices to further elaborate on the concept of phonetic compliance.

Table 1. Results

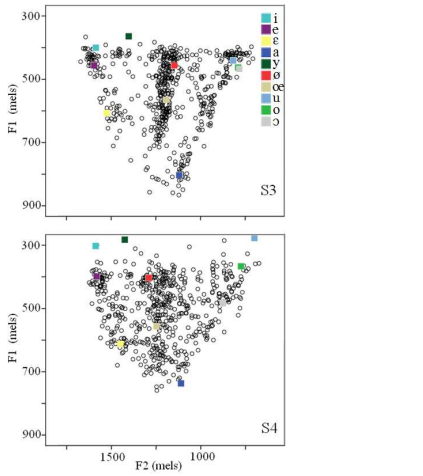
	I1		I2		I3	
Speaker	Value	Rank	Value	Rank	Value	Rank
S1	112709	3	34912	4	3872	3
S2	122050	4	42286	3	7475	4
S3	83570	2	45083	2	2457	2
S4	77422	1	49294	1	1552	1

$$I1 = \sum_{s=1}^{S}\sum_{p=1}^{P}\left[\sum_{i=1}^{I}(Fi_{ps} - Fi_{s})^2\right]^{1/2}$$

$$I2 = \sum_{s=1}^{S}\sum_{p=1}^{P}\left\{\prod_{v=1}^{V}\log\left[\sum_{i=1}^{I}(Fi_{ps} - \overline{Fi_{v}})^2\right]^{1/2}\left[\sum_{i=1}^{I}(Fi_{ps} - Fi_{s})^2\right]^{-1/2}\right\}$$

$$I3 = \mathrm{var}_{s}\left\{\sum_{p=1}^{P}\mathrm{var}_{p}\left(\left[\sum_{i=1}^{I}(Fi_{ps} - Fi_{s})^2\right]^{1/2}\right)\right\}$$

Figure1. F1 and F2 values (Mels) during the imitation task (black); centroïds of the 10 French oral vowels; for S3 and S4.



References

References

Cohen, A. D., & Macaro, E. 2007. Language Learner Strategies: Thirty Years of Research and Practice. Oxford: Oxford University Press.

Dogil, G., & Reiterer, S. 2009. *Language Talent and Brain Activity.* New York: Mouton de Gruyter.

Dörnyei, Z. 2009. *The Psychology of Second Language Acquisition.* Oxford: Oxford University Press.

Francis, A. L., & Nusbaum, H. C. 2002. Selective attention and the acquisition of new phonetic categories. *Journal of Experimental Psychology: Human Perception and Performance* 28, 349–366.

Golestani, N., & Zatorre, R. J. 2009. Individual differences in the acquisition of second language phonology. *Brain and Language* 109, 55–67.

Jilka, M., Anufryk, V., Baumotte, H., Lewandowska, N., Rota, G., & Reiterer, S. 2007. Assessing Individual Talent in Second Language Production and Perception *5th ISALSS*, Florianópolis, Brazil, 243–258.

Moyer, A. 2004. *Age, Accent, and Experience in Second Language Acquisition: An Integrated Approach to Critical Period Inquiry*. Clevedon: Multilingual Matters.

Piske, T., MacKay, I., & Flege, J. 2001. Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics* 29, 191–215.

Can a spoken dialog system be used as a tool to study convergence?

José Lopes[1,2]*, Andrew Fandrianto*[3]*, Maxine Eskenazi*[3]*, and Isabel Trancoso*[1,2]

[1]Instituto Superior Técnico, Lisboa, Portugal
[2]INESC-ID Lisboa, Portugal
[3]Language Technologies Institute, Carnegie Mellon University, Pittsburgh, PA, USA
jose.david.lopes@l2f.inesc-id.pt

The finding that people entrain to one another in a conversation (Brennan, 1996) has fostered much interest in this phenomenon within a variety of research communities, such as psychology. Members of the automatic speech processing community have viewed it as a potential functionality that, if present in human-machine interaction, could be capitalized upon to improve system performance (Lopes et al., 2011). Beyond the benefits to SDS research, we argue here that automated systems can, in turn, benefit research in other areas. We believe that, for the study of entrainment, SDS can provide platforms on which to run studies, offering more control over conditions in some ways that do human-human studies. We use the term entrainment here, from Brennan and others. This term may represent the action of one of the speakers. Assuming both speakers entrain, there should be convergence.

The literature does show that humans can be made to change their speech patterns to imitate the output of a spoken dialog system (SDS). Stoyanchev and Stent (2009) used a set of dialogs to study entrainment using two verbs and two prepositions as primes. They confirmed that callers can adapt their choice of terms to the terms used by the automated system. In Parent & Eskenazi (2010) the system primes were directly manipulated in the Let's Go spoken dialog system (real, not paid callers, Raux et al., 2005) and observed caller adaptation over time. The authors found that users do adapt and are more likely to do so in the first few turns following the first appearance of the prime. The same was done in European Portuguese with the Noctívago spoken dialog system (Lopes et al., 2011) with the result that the enlisted callers entrained to all of the primes that were proposed by the system.

Despite confirming the presence of entrainment, not all proposed primes were copied. Looking in more detail on the lexical and prosodic levels, both Lopes et al. (2011) and Parent & Eskenazi (2010) found differences in how often words were copied. Less frequent words were copied less frequently if they were new primes (and the system was already offering a very frequent prime, for example, "help" $>$ "assistance") and that conversely, if an infrequent word had been used and was replaced with a more frequent prime, the latter was easily copied ("start a new query" to "start a new request"). Lopes et al. (2011) observed in Noctívago that if a very frequent and contextually appropriate word had been used (like "agora", now) it would continue to be used whether the system still used it in its prompts or not. But the primes proposed here, for example "imediatamente" (immediately) and "neste momento" (right now), are all much longer and not necessarily more natural than "agora". Neither study found any influence of the part of speech on the likelihood to be copied. Both studies confirmed that continued exposure to the primes increases the likelihood of their uptake.

The individual choices may not, for some words, follow the lexical frequency in the language. This can be due to individual preference, local uses, professional uses or a myriad of other reasons. In a relatively short dialog, like the examples presented here, it would be difficult to adapt to these individual differences. If a dialog system was to be used by the same person over a longer period of time, this would be possible. And choices that may be made due to avoidance of difficult phonetic clusters (as in foreign words) can be dealt with automatically.

Entrainment on the prosodic level was further analyzed. In an attempt to get callers to stop shouting or hyperarticulating, the system spoke more softly or more slowly, respectively. It was observed that callers more frequently copied the first condition than the second. In this case, the system was adjusted to speak precisely 25% faster (measured in syllables per second) or 25% softer. This type of precise control would be difficult to obtain in human-human studies even if one speaker was instructed to speak 25% softer. Given the training and tuning, a limited domain speech synthesizer can vary elements like speaking rate, pitch variability and contour, rhythm, and intensity with great precision.

Interestingly, to our knowledge there has not yet been a study that compared entraining to an SDS and entraining to another human on a similar task with similar constraints. This could help further our understanding of

the differences in the two conditions. Armed with this knowledge, some studies could be carried out where the appearance of the prime is tightly controlled using an SDS and have some way to relate the findings above described to what humans might do when speaking to one another.

There are several other benefits to the use of an SDS in this area. Running studies on an SDS with real users reduces long term cost and increases scalability. While one laboratory study may painstakingly find 50 participants, running the study over a week or two, an SDS with real users, as in the case of the Let's Go platform, can get over 500 in the same time frame. These callers receive no remuneration other than getting bus scheduling information, thus additionally reducing costs.

We have seen that spoken dialog systems can offer controlled conditions for studies on how humans copy speech. We believe that these types of platforms should be considered as one of many tools that those who study entrainment can use.

References

Brennan, S. E. 1996. Lexical entrainment in spontaneous dialog. *Proceedings of the International Symposium on Spoken Dialog*.

Lopes, J., Eskenazi, M., & Trancoso, I. 2011. Towards Choosing Better Primes for Spoken Dialog Systems. *Proceedings ASRU 2011*.

Parent, G., & Eskenazi, M. 2010. Lexical Entrainment of Real Users in the Let's Go Spoken Dialog System. *Proceedings Interspeech 2010*.

Raux, A., Langner, B., Bohus, D., Black, A. W, & Eskenazi, M. 2005. Let's Go Public! Taking a Spoken Dialog System to the Real World. *Proceedings Interspeech 2005*.

Stenchikova, S., & Stent, A. 2007. Measuring adaptation between dialogs. *Proceedings SIGdial 2007*.

Individual differences in phonetic convergence

Alan C. L. Yu[1]*, Carissa Abrego-Collier*[1]*, and Morgan Sonderegger*[2]

[1]Phonology Laboratory, University of Chicago
[2]Department of Linguistics, McGill University
aclyu@uchicago.edu

Introduction: Recent studies have documented phonetic convergence (PC), but little is known about the great individual variability in the likelihood of convergence ubiquitous in PC studies. Understanding the source(s) of the individual differences is crucial for understanding sound change propagation, e.g., for identifying the characteristics of early adoptors of change. This study shows that the extent of PC depends on the speaker's disposition towards an interlocutor, as well as the speaker's personality traits and working memory capacity (WMC).

Methodology: The experiment contained three parts: a *baseline* production block and a post- training *test* block where subjects produced a list of 72 p/t/k-initial target words (randomized order) in a carrier phrase. In between the two production tasks was a *listening* block where subjects heard a constructed first-person narrative in which the same 72 p/t/k words were embedded. VOTs for the target words in the story were extended by 100% using Praat. The narrative described a male talker's blind date from the previous night and contained no other stressed syllable-initial voiceless aspirated stops aside from the target words. Two versions of the narrative were created: one in which the talker abandons his date and goes home alone ("negative" version), and one in which the date goes well and they leave together ("positive" version). For each version, there were two conditions: one in which the talker's date was female ("straight" condition), and one in which the talker's date was male ("gay" condition). All subjects also took the Automated Reading Span Task (RSPAN; a measure of working memory) and completed a series of on-line surveys, including the Big Five Inventory. 58 subjects (approximately evenly divided across conditions) participated in the study.

Analysis: Subjects' VOTs from the *baseline* and *test* blocks were analyzed using a linear mixed- effects model, which contained several types of predictors: **Time**: BLOCK (baseline vs. test), and TRIAL (the word's within-block position); **Linguistic**: CONSONANT the word began with, word length by SYLLABLES, log-transformed word frequency, and two speaking rate predictors; **Social**: subject GENDER, narrator SEXUALITY (gay vs. straight), subject ATTITUDE towards the talker (1–7), and narrative OUTCOME (positive vs. negative); **Cognitive**: RSPAN; **Personality**: "Big 5" scores (Agreeableness, Openness, Conscientiousness, Extroversion, Neuroticism). Fixed effects were included for the main effects of all predictors, as well as interactions of BLOCK with social, cognitive, and personality predictors. These interactions are the terms of primary theoretical interest, corresponding to what effects each of these types of predictors has on the degree of imitation. By-speaker and by-word random intercepts were also included, as well as all possible by-speaker random slopes.

Results: There is no main effect of block ($p > 0.25$): on average, subjects did not change VOT following exposure to the narrator's lengthened VOTs. The effect of block is strongly mediated by significant interactions. Subjects with a positive attitude towards the narrator converge, while subjects with a negative attitude diverge (BLOCK:ATTITUDE: $p < 0.001$). Male subjects converge to the (male) narrator, while female subjects diverge (BLOCK:GENDER: $p < 0.05$). Hearing the positive narrative was associated with divergence, and the negative narrative with convergence (BLOCK:OUTCOME: $p < 0.05$). A subject's openness to new experience (BLOCK:O: $p < 0.001$) and working memory capacity (BLOCK:RSPAN: $p < 0.01$) positively correlate with her degree of convergence. No effect of cognitive or personality predictors besides O and RSPAN was found.

Conclusion: Our results suggest that the dynamics of phonetic imitation is mediated by factors such as speaker attitude, which is constructed situationally, as well as an individual's personality and cognitive traits, such as openness and working memory capacity. These findings highlight the importance of considering individual-level characteristics in studying PC; whether an individual converges with their interlocutor depends not only on situationally-determined factors but also the personality and cognitive profile of the individuals involved.

Quantification of speech convergence through non-linear methods for the analysis of time-series

Leonardo Lancia, Susanne Fuchs, and Amelie Rochet-Capellan

leonardo_lancia@eva.mpg.de, fuchs@zas.gwz-berlin.de, ameliecapellan@free.fr

Since the first studies reporting phonetic convergence, this phenomenon has often been interpreted in theoretical frameworks characterized by a non segmental approach to speech production and perception as direct realist theory (Sancier & Fowler, 1997) or the exemplar theory (Goldinger, 1998). However objective methods currently used to detect and evaluate the amount of phonetic convergence are most of the time derived from a segmental (i.e. static) approach, and are based on measurements conducted at particular points in time or on temporal averages. This static approach is problematic for non segmental theories which 1) predict variability related to one single segmental slot to spread over larger portions of the speech chain and 2) give a central role to the evolution over time of the features characterizing the speech signal and to the trajectories corresponding to articulatory gestures. In our recent work, time-series comparison methods were adapted to various kinds of signals derived from speech production (eg. movements of the articulators, aerodynamic signals and acoustic signals) and tested with synthetic and/or natural signals whose variability could be controlled. Here, we will present three main approaches, which should make researchers able to analyze a wide range of signals, coming from several kinds of experimental tasks, without being bounded to a static characterization of the signals derived from speech.

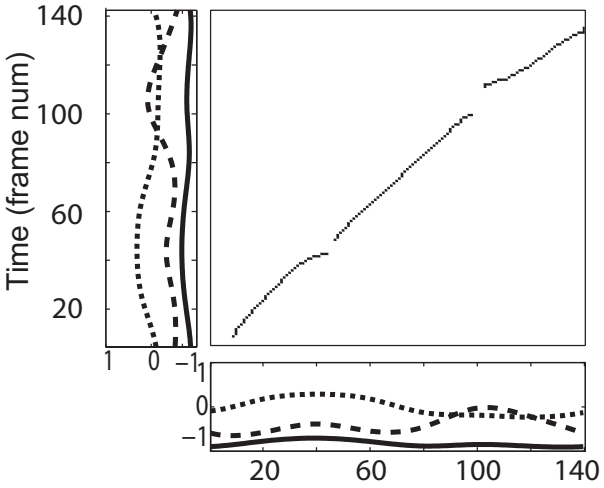
1. **Comparisons of mono dimensional trajectories**. A mono dimensional trajectory corresponds to a curve with the time on the abscissa and the value of one observable quantity on the ordinate (e.g. the fundamental frequency of a voiced signal). The main difficulty encountered when comparing time-varying trajectories is due to the lack of time alignment between corresponding events (peaks and valleys) on the different curves. It is often important to disentangle this temporal variability from the variability in the magnitude and in the shape of the corresponding events in the two trajectories. In the Functional Data Analysis framework (Lucero et al., 1997), this separation is achieved by representing each curve with a set of parameters shaping well known analytical functions (e.g. sinusoidal functions, Bsplines, discrete wavelets) and by aligning in time the functional representations through an optimization algorithm which proceeds by minimizing the differences between each trajectory and the average trajectory. Aligned trajectories can thus be compared and time varying measures of variability across trajectories can be obtained.

2. **Synchrony of mono dimensional cyclic trajectories**. Sometime, we are interested in the synchronization between two signals showing a roughly cyclical but non-stationary behavior. This may be the case when comparing the breathing patterns of speakers involved in a conversation. Spectral methods based on the wavelet transform can be used to obtain a cross-spectrum which represents the energy shared by the two signals at different frequencies and the relative phase between the signals at each frequency (Torrence & Compo, 1998). When the signals compared oscillate at the same frequency, this can easily be tracked on the cross-spectrum and the relative phase of the signals at this frequency is used to characterize their synchronization. The synchronization of signals which have different fundamental frequencies of oscillation is captured by the notion of generalized phase difference. For this kind of signals, application of spectral methods is problematic because the two frequencies of oscillation need to be estimated separately and the algorithms to perform this task are subject to errors. This issue could be solved by processing the two signals with a peak finding algorithm and correcting the results by hand. The information about the duration of each cycle was then used to inform an algorithm designed to find the main frequency component of each signal.

3. **Comparisons of multi dimensional trajectories**. Speech signals are often represented by multivariate trajectories (as for example, the joint motion of several articulators or the energy curves measured at different frequency bands in the spectrogram of acoustic signals). Multidimensional signals cannot be aligned with

a Functional Data Analysis approach because the time shift between the two signals can vary from one dimension to the other. Also popular methods like dynamic time warping are prone to errors, especially when the signals compared do not contain the same events in the same order. In such cases, a modified version of cross-recurrence analysis, a technique developed to compare the behavior of dynamical systems, can be used (Lancia & Tiede, 2012). The time scales of the two signals define the axes of a cross- recurrence plot which is populated by dark dots indicating the similar portions of the two signals (cf. Fig. 1). In the modified approach, the cross recurrence plot is submitted to a cleaning algorithm designed to remove those dark dots which are considered artifacts due to different rates or directions of change in the signals or to the presence in a given signal of repeated events. The similarity between the signals can then be quantified by counting the number of dots belonging to continuous lines, regardless of their slope and curvature. This measure is sensitive to differences in the shapes of the trajectories but not to differences in their time scales. Altogether the methods outlined above provide the means to measure the similarity between mono or multidimensional non stationary signals and to characterize inter-speakers, unpredictable variation.

Figure 1: Recurrence plot of the two multivariate signals on the x and the y axes, representing the joint motion of the tip of the tongue, the lower lip and the jaw during two repetitions of the utterance /tapa/. A dark dot at position i,j means that the first signal at the ith point in time is similar to the second signal at the jth point.



References

Goldinger, S. D. 1998. Echoes of echoes? An episodic theory of lexical access. *Psychological Review* 105, 251–279.

Lancia, L., & Tiede, M. 2012. A survey of methods for the analysis of the temporal evolution of speech articulator trajectories. In: Fuchs, Weihrich, Pape, Perrier (eds), *Speech Planning and Dynamics*. Frankfurt am Main: Peter Lang, 233–271.

Lucero, J., Munhall, K., Gracco, V., & Ramsay, J. 1997. On the registration of time and the patterning of speech movements. *Journal of Speech, Language, and Hearing Research* 40, 1111–1117.

Sancier, M. L., & Fowler, C. A. 1997. Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics* 25, 421–436.

Torrence, C., & Compo, G. P. 1998. A Practical Guide to Wavelet Analysis. *Bulletin of the American Meteorology Society* 79, 61–78.

The role of visual cues in imitating a new sound

Nancy Ward and Megha Sundara

University of California, Los Angeles
nancyward@ucla.edu, meghasundara@humnet.ucla.edu

Recent research shows that adults are sensitive to more than just the auditory cues when they are imitating a person implicitly. Adults shown a video of a model talker converged with the talker's articulations, even in the absence of auditory cues (Miller et al., 2010). Visual speech not only elicits convergence on its own, but convergence is greater for audiovisual speech than for auditory-only speech (Dias & Rosenblum, 2010). In this study, we investigate the mechanism by which visual cues facilitate convergence. Specifically, we test whether the presence of visual cues simply increases overall attention to the speech stimuli or whether the addition of visual cues to speech highlights visually salient aspects of speech sounds and thus, selectively enhances the imitation of those aspects.

We start from the premise that features of speech sounds can be more or less visually salient. For example, lip rounding is more easily comprehensible than vowel front/backness or openness in the visual modality. Studies on integration of auditory and visual speech cues have found that the use of each modality is dependent on the specific features present in the sound, such as whether these visually salient properties like lip rounding are present, and therefore these same findings may hold for imitating perceived sounds (Traunmüller & Öhrström, 2007). Our prediction in this study is that these visually salient aspects of the speech signal will be imitated more closely in a setting in which a subject receives audiovisual exposure, whereas auditorily salient aspects of the speech signal will be imitated to a similar degree across different modalities of exposure (auditory or audiovisual).

In this study, we are interested in not just how visual cues affect imitation, but how this contribution of the visual cues can aid in the process of acquiring a new sound. In order to test this, we will look at how imitation differs for familiar and unfamiliar sounds in the auditory and visual modalities. In order to establish how imitation differs between familiar and unfamiliar sounds, this experiment tests imitation of both types of sounds. Imitation is shown to be better for sounds with a larger accepted pronunciation range (Babel, 2009, 2010), but what about sounds that are not in the subject's language? Many studies on second- language learners show an advantage in performance on perception tasks when presenting the visual cues in addition to the auditory cues (Hardison, 2003; Navarra & Soto-Faraco, 2007). These studies on second-language acquisition clearly illustrate a facilitatory effect of audiovisual exposure on learning unfamiliar sounds, but this research has all focused on learning the ability to *perceive* the sounds, not the effects on learning to produce the new articulatory gestures.

Monolingual English-speaking adult subjects (n = 20/group) participated in a three-part study. The task used in this experiment was a modified version of the word-naming imitation paradigm (Goldinger, 1998; Nielsen, 2007, 2008, 2011). In the first part, the *baseline phase*, subjects heard the model talker producing a set of target words and were asked to repeat the word they heard. Following this, they participated in an *exposure phase*, in which they either **heard** or **heard and saw** the model talker producing multiple repetitions of a subset of the words heard in the *baseline phase*. Finally, they participated in a *post-exposure test phase* in which they repeated the same set of words from the *baseline phase*. We are testing whether subjects improve in their imitation between the *baseline* and *post-exposure test phases* according to whether they **heard** or **heard and saw** the talker in the *exposure phase*.

The target word stimuli used in this experiment were invented words with French vowels. They were produced by a native French speaker and modeled after Traunmüller & Öhrström (2007). They contain vowels differing in both acoustically-salient (backness and height) and visually-salient (rounding) features. We are analyzing the acoustic qualities of the vowels produced by the subjects, in comparison with the vowels from the model talker, in a manner following Babel (2009, 2010). The vowel formants, duration, and pitch will be measured in the baseline and post-exposure productions and the two sets of measurements will be compared to each other, as well as to the formants of the model talker's vowels, using the Euclidean distance measurement. The first three formants will be used to capture information about aperture, backness, and rounding. If a subject imitated

the model talker, then the acoustic properties of the vowels will be more similar to the model talker's vowels in the post-exposure test phase than in the baseline phase. This analysis will be statistically confirmed with a mixed-model ANOVA.

We hypothesize that monolingual English-speaking adults will show differences in their imitation of acoustically and visually salient vowel features, depending on the mode of exposure in the experiment. In particular, we predict that imitation of vowel rounding will be improved in the audiovisual condition, but imitation of vowel backness and openness should not differ to the same degree between the auditory and audiovisual conditions. We also hypothesize that there will be differences in imitation of unfamiliar versus familiar sounds; perhaps the contribution of the visual signal will facilitate greater imitation of unfamiliar sounds due to that there is no established pronunciation range.

Preliminary pilot data from six English-speaking adults suggest that there are differences in imitation according to the type of exposure, but that convergence overall is increased for audiovisual over auditory exposure (confirming previous results showing that visual cues increase imitation). When the results were broken down by vowel feature, there was little effect of the visual cues on convergence to the openness dimension, but a larger contribution of the visual cues to the backness and rounding dimensions. Within the results for the rounding dimension, there was an increased contribution from the visual cues for unfamiliar sounds compared to familiar sounds, i.e. visual cues aided in subjects' learning the vowel roundedness cue. More comprehensive results are forthcoming.

In summary, the focus of this study is to understand the role of visual cues in imitating a new sound. In these experiments, we are looking to get a sense of how a speaker presented with new sounds may best learn to produce them, and whether visual cues help in this process.

References

Babel, M. 2009. *Phonetic and Social Selectivity in Speech Accommodation*. Doctoral Dissertation, University of California Berkeley.

Babel, M. 2010. Dialect convergence and divergence in New Zealand English. *Language in Society* 39, 437–456.

Dias, J. W., & Rosenblum, L. D. 2010. Visual influences on interactive speech alignment. Poster presented at the *Second Pan-American/Iberian Meeting on Acoustics*.

Goldinger, S. D. 1998. Echoes of echoes? An episodic theory of lexical access. *Psychological Review* 105, 251–279.

Green, J. R., Nip, I., Wilson, E. M., Mefferd, A. S., & Yunusova, Y. 2010. Lip movement exaggerations during infant-directed speech. *Journal of Speech, Language, and Hearing Research* 53, 1529–1542.

Hardison, D. 2003. Acquisition of second-language speech: effects of visual cues, context, and talker variability. *Applied Psycholinguistics* 24, 495–522.

Miller, R. M., Sanchez, K., & Rosenblum, L. D. 2010. Alignment to visual speech information. *Attention, Perception, & Psychophysics* 72, 1614–1625.

Navarra, J., & Soto-Faraco, S. 2007. Hearing lips in a second language: visual articulatory information enables the perception of second language sounds. *Psychological Research* 71, 4–12.

Nielsen, K. 2011. Specificity and abstractness of VOT imitation. *Journal of Phonetics* 39, 132–142.

Nielsen, K. Y. 2007. Implicit phonetic imitation is constrained by phonemic contrast. *Proceedings of ICPhS 2007*, 1961–1964.

Nielsen, K. Y. 2008. *The Specificity of Allophonic Variability and its Implications for Accounts of Speech Perception*. Doctoral Dissertation, University of California, Los Angeles.

Traunmüller, H., & Öhrström, N. 2007. Audiovisual perception of openness and lip rounding in front vowels. *Journal of Phonetics* 35, 244–258.

Sensory-motor maps of speech: Functional coupling and sensory-motor predictions during vowel perception and production

Krystyna Grabski and Marc Sato

Gipsa-Lab, Département Parole & Cognition, CNRS UMR 5216 & Grenoble Université
krystyna.grabski@gips-lab.inpg.fr, marc.sato@gipsa-lab.inpg.fr

Do speech representations draw on procedural knowledge and sensory-motor experience? Although a functional distinction between frontal motor sites for speech production and temporal auditory sites for speech perception has long been postulated, some recent neurobiological studies on mirror mechanisms and motor control rather argue for sensory-motor interactions in both speech production and perception. During speech production, decreased neural responses observed within the auditory and somatosensory cortices are thought to reflect feedback control mechanisms in which sensory consequence of the speech-motor act are evaluated with actual sensory input in order to further control production. Conversely, motor system activity observed during speech perception has been proposed to partly constrain phonetic interpretation of the sensory inputs through the internal generation of candidate articulatory categorizations.

In this framework, we performed two independent sparse sampling functional magnetic resonance imaging (fMRI) studies to compare the neural correlates of French vowel perception and production (i/, /y/, /u/, /e/, /ø/, /o/, /ɛ/, /œ/ and /ɔ/ vowels, auditory and visually presented, respectively). The aim of the first study was to identify the neural networks underlying vowel perception and production and to investigate whether sensory-motor brain areas might commonly participate in both systems. In order to further investigate/identify possible sensory-motor loops involved in predictive coding during vowel perception and production, we carried out a second sparse sampling fMRI study with new participants using an adaptation (or repetition-suppression, RS) paradigm. This method is based on suppression of neural activity caused by a repeated presentation of a stimulus or a feature, and can be interpreted as the ability to predict specific features the adapting neural population is specifically sensitive to.

Results from both experiments and whole-brain analyses demonstrate that vowel representations are largely distributed over left sensory-motor brain areas (the posterior inferior frontal gyrus, the supramarginal gyrus and the superior temporal gyrus/sulcus) in both vowel perception and production. Furthermore, these brain areas also appear sensitive to adaptation effects in both tasks and are likely involved in sensory-motor predictions. These results appear in line with recent neurobiological models of speech perception and production that postulate a crucial role of these regions in sensory-to-motor and motor-to-sensory speech interactions. Taken together, these results strongly suggest a functional coupling of vowel perception and production and provide evidence for sensory-motor mechanisms involved in predictive coding of vowels.

Bio-acoustic fingerprints of individual differences in bilingual speech imitation ability: neuro-imaging and spectral analysis

Susanne M. Reiterer[1,2]*, Xiaochen Hu*[2,3,4]*, Sumathi T. A.*[5]*, and Nandini C. Singh*[5]

[1]University of Vienna, Centre for Language Learning and Teaching Research (FDZ), Vienna, Austria
[2] University Clinic Tübingen, Tübingen, Germany
[3] University of Tübingen, Hertie Institute for Clinical Brain Research, (Hertie, CIN), Tübingen, Germany
[4] University of Bonn, Clinic for Psychiatry and Psychotherapy (KBFZ), Bonn, Germany
[5] National Brain Research Centre (NBRC), Manesar, India
Susanne.Reiterer@univie.ac.at

1. Introduction

Speech sound imitation is a pivotal learning mechanism for humans. Individuals differ greatly from each other in their aptitude, ability and success in sound imitation learning. This is especially evident when it comes to the acquisition of a second language sound system. An unanswered scientific question is why individuals show these differences in their ability to produce and imitate foreign speech sounds. Purely cognitive explanations are discussed controversially and very little is known about the neuro-biological/neuro-cognitive correlates of this special aptitude There is increasing interest in individual differences in skill acquisition, including second language learning skills in multilinguals. However, one variable causing individual differences in language learning has been largely neglected so far: the aspect of "ability" or language aptitude. Apart from a few exceptions (e.g. Golestani et al., 2007), language aptitude has not been investigated by means of brain imaging tools or psycho-acoustic measurements so far. We therefore investigated language aptitude from a neuro-psycholinguistic perspective, focusing on the outstanding skill of "pronunciation or acoustic-phonetic imitation ability" with respect to a second language.

2. Methods

2.1. Subjects and Behavioural Testing

At first, 138 German-speaking subjects were recorded during imitation of Hindi (L0) sentences, i.e., a language they had no experience with (internet rating database, perceptual evaluation by 30 native Indian raters, compare Reiterer et al., 2011). Hindi was chosen to eliminate confounds with previous language experience. From these subjects low scoring, midrange scoring and high scoring individuals (each N=20) were chosen to form 3 balanced groups over all ability levels for further investigations with brain imaging and acoustic signal analysis tools. Two groups of each 9 (mean age: 28 yrs, 4 females each, right-handed, mother tongue (L1) German, onset of L2 (English) learning at 10 yrs of age, late learners) with either high or low speech imitation/pronunciation abilities – as evaluated by the repetition of Hindi test materials (unknown language) – participated then in the fMRI and acoustic experiments.

2.2. Method 1 - Functional magnetic resonance imaging (fMRI)

Altogether, 60 participants underwent fMRI scanning (1.5 T scanner, sparse sampling, TR: 12 s, TA: 3 s, SPM5, flexible factorial ANOVA, random effects analysis) during a sentence reading task in 3 different sub-conditions: A) reading native language German (L1); B) reading second language English (L2); C) reading German sentences with a "fake" English accent (L1 ACC). The two extreme groups of high and low ability (top and bottom 15% of the group of 60 subjects who had participated in fMRI measurements based on their Hindi score) were compared to each other by means of fMRI (whole-brain corrected for multiple comparisons at $p < 0.05$, cluster level).

2.3. Method 2 - Modulation Spectrum Analysis (MSA)

Amplitude modulations of the speech envelope encode characteristic articulatory features. Here we used novel methods of spectral analysis to construct a speech modulation spectrum, which is a probability distribution of the different spectral, temporal and spectro-temporal modulations of the amplitude envelope (compare Singh & Singh, 2008). The presence of energy fluctuations across a frequency spectrum at particular times are called spectral modulations (ω_x) and temporal modulations (ω_t). Based on the specific time scales, the temporal modulations provide both segmental and suprasegmental information, whereas the spectral modulations provide

information about harmonic and formant structure. Speech rhythm for instance is encoded in low temporal modulation while suprasegmental information like formant transitions and voice onset times are encoded in higher temporal modulations. Since the spectro-temporal modulations represent different articulatory features, the 2-D energy distributions of the spectro-temporal modulations is called the "articulation space". The segmental and suprasegmental features of differently skilled accent imitators were obtained and compared with native speakers' utterances for 2 different types of long utterances (1= imitate Hindi sentences, 2=read aloud sentences in L1 German (condition A), L2 English (cond. B), and German with a "fake" English foreign accent (cond. C). Thus we obtained characteristic speech modulation spectra of our participants.

3. Results

3.1. fMRI (Method 1)

Both groups showed hemodynamic activation of widely distributed bilateral language networks (including inferior-frontal premotor areas, sensorimotor cortex, temporo-parietal, visual regions, insulae, basal ganglia and the cerebellum. Generally, subjects with low pronunciation skills displayed significantly higher and more widespread activations when compared to the High ability group – especially when they had to fake a foreign accent. The overall activation increases of the Low ability group in all reading tasks comprised both hemispheres but peak activation was in the left inferior parietal (supramarginal gyrus, BA 40) and postcentral areas, which integrate aspects of speech perception and production, possibly (in combination with premotor areas) a correlate for acoustic speech imitation ("acoustic mirror neuron system").

3.2. Modulation Spectra (Method 2)

Articulation space (the area covered) correlated positively with imitation skill (e.g. Pearson's r=0.7, $p<0.01$ in condition C, when faking the foreign accent, see also Fig. 1).In each condition (A,B,C) the high ability imitators (blue bars in Fig. 2) had significantly larger articulation areas. Our results suggest that skilled accent imitators have a larger articulation space as compared to poor imitators.

Figure 1

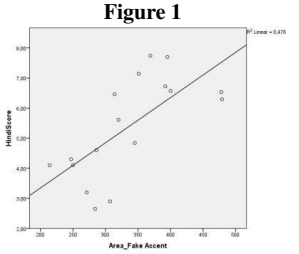
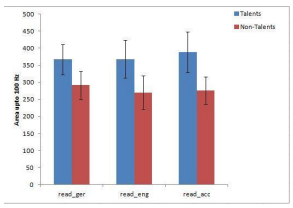


Figure 2



Our working hypothesis suggests that this extension in space might provide access to a larger repertoire of sounds, which in turn could possibly provide skilled imitators greater flexibility in pronunciation. This might confirm our hypothesis that even late but highly skilled L2 speakers who are good at accent imitation in general keep their phonetic categories more flexible and open for being exposed to new sounds without confining their articulatory repertoire to the mother tongue speech sound processing schemes.

4. Conclusion

Our data provide evidence that individual differences in speech imitation ability/aptitude play a decisive role in speech production which sometimes even override differences between languages and can be visualized and quantified by various methods of signal analysis. Regarding speech imitation ability on a neuro-cognitive scale, we confirmed the theory of cortical processing effort by showing that increase in intensity as well as extensity of cortico-subcortical speech relevant areas can be shown as a function of speech imitation ability even in the case of the mother tongue. Poorer skills are always associated with higher amounts of consumption of neural workspace. With regard to acoustic measures of speech output we found a larger articulation space to be a possible "marker" of high ability in L2 speech imitation skills.

References

Golestani, N. et al. 2007. Brain structure predicts the learning of foreign speech sounds. *Cerebral Cortex* 17, 575–582.

Reiterer, S., et al. 2011. Individual differences in audio-vocal speech imitation aptitude in late bilinguals: functional neuro-imaging and brain morphology. *Frontiers in Psychology (Language Sciences)* 2, 271.

Singh, N. C., & Singh, L. 2008. The development of articulatory signatures in children. *Developmental Science* 11, 467–473.

The sound of your lips: haptic information speeds up the neural processing of auditory speech

Avril Treille, Camille Cordeboeuf, Coriandre Vilain, & Marc Sato

Gipsa-Lab, Département Parole et Cognition, CNRS UMR 5216 & Grenoble Université
{avril.treille,camille.cordeboeuf,coriandre.vilain,marc.sato}@gipsa-lab.inpg.fr

While speech perception has long been thought as a mere auditory process, the human ability to follow speech gestures through other sensory modalities can be considered as a core component of speech perception. Remarkably, speech can be perceived not only by the ear and by the eye but also by the hand. Strong evidence for manual tactile speech perception mainly derives from researches on the Tadoma method that has evolved within the deaf-blind community. In this method, sometimes referred to as "tactile lipreading", speech is received by placing a hand on the face of the talker and monitoring facial movements. Although years of training are required to learn the Tadoma method, remarkable performance and almost normal communication can be achieved by some experienced deaf-blind users. Crucially, a few studies also provides evidence for audio-tactile speech interactions in naive and untrained normally sensed adults with inexperienced participants presented with syllables heard and felt from manual tactile contact with a speaker's face.

Given the multisensory nature of speech perception, one fundamental question is whether sensory signals are integrated in the speech processing hierarchy and may reflect predictive sensory-motor, anticipatory, mechanisms. Despite no current agreement between theoretical models of audiovisual speech perception regarding the processing level at which the acoustic and visual speech signals are integrated, several magnetoencephalographic (MEG) and electroencephalographic studies (EEG) of audiovisual speech perception suggest that visual speech input modulates activity in the primary and secondary auditory cortices at an early stage in the cortical speech processing hierarchy.

The present EEG study aimed at further investigating early cross-modal interactions in speech perception. To this aim, we compared auditory-evoked N1 and P2 responses (appearing 100 ms and 200 ms from the acoustic onset) from fourteen participants during auditory, audio-visual and audio-haptic perception of /pa/ and /ta/ syllables. Participants were seated at arm's length from a female experimenter. In the auditory condition, they were instructed to listen to the produced syllables with their eyes closed. In the audio-visual condition, they were asked to look at the experimenter's face. In the audio-haptic condition, they were asked to keep their eyes closed with their right hand placed on the experimenter's face (the thumb vertically against the experimenter's lips and the other fingers horizontally along the jaw line in order to to help distinguishing both lip and jaw movements related to /pa/ and /ta/ syllables).

In line with previous studies, auditory-evoked N1 amplitude was attenuated during audio-visual compared to auditory speech perception for frontal, central and parietal electrodes, as well as to audio-haptic speech perception. Crucially, shortened latencies of N1 responses were observed during audio-haptic and audio-visual speech perception, compared to auditory speech perception for frontal, central and parietal electrodes.

Altogether, these results suggest some early integrative mechanisms between auditory, visual and haptic modalities in speech perception as well as a predictive role of haptic and visual information in auditory speech processing.

Synchrony and convergence of pauses in spontaneous conversation

Kristina Lundholm Fors

kristina.lundholm@gu.se

1 Introduction and background

A pause is a silence that occurs within a speaker's turn, and it can, but does not have to, coincide with a transition relevance place (TRP, Sacks et al., 1974). The placement and length of pauses are highly important when it comes to turn taking in conversation. If speakers do not reach an implicit agreement on the tolerated length of pauses, the dynamics of the conversation will be considerably affected. Tannen (1985) gives several examples of how we react negatively to pauses that are longer or shorter than we expect, and shows that a more "fast-paced" person, i.e. someone who does not tolerate long silences, will tend to dominate the conversation. Furthermore, long silences can be interpreted by listeners as a sign of trouble in the conversation (Roberts et al., 2006).
Because of the importance of pause lengths in conversation, we are interested in finding out how speakers adapt to each other with regards to pauses. Speakers ordinarily grow more similar to each other when speaking to each other, and this adaption process is often referred to as entrainment. It is believed that entrainment is present on all levels of communication, and that this process helps us understand each other (Pickering & Garrod, 2004). Entrainment has been investigated in a number of studies and evidence has been found for for example lexical entrainment (Brennan, 1996), phonetic entrainment (Pardo, 2006) and acoustic-prosodic entrainment (Levitan & Hirschberg, 2011). Edlund et al. (2009) analyzed pause and gap lengths in dialogues, and found indications of entrainment, albeit the results were not consistent across all dialogues. In this study, we explored the entrainment of the length of pauses (intra-turn silences) in spontaneous dialogues.

2 Method and material

The data consisted of 6 dialogues, each approximately 10 minutes. The speakers were 5 Swedish females, and they were recorded in a recording studio while engaged in face-to-face interaction. After the recording, pauses were annotated manually based on the acoustic signal. To investigate entrainment, we used the method proposed by Edlund et al. (2009), with some slight variations.

3 Results

Correlations between pause lengths in each of the 6 dialogues were calculated using Pearson's r and are presented in Table 1. We found that four of the dialogues exhibited a significant correlation between the pause lengths of the speakers in the dialogue, which is evidence for entrainment in the form of synchrony: in dialogue 1 and 6 there was a strong positive relationship between pause lengths, and in dialogue 3 and 4 there was a moderate positive relationship between pause lengths. In dialogue 5 we found no significant correlation between the pause lengths of the speakers, and in dialogue 2 there was a strong negative correlation between the two speakers' pause lengths. For dialogue 5, we decided to explore whether grouping the pauses based on whether they coincided with TRPs would affect the outcome of the analysis. For the pauses that coincided with TRPs, we found a significant positive correlation (r=0.466, p=0.039), whereas the pauses that did not occur at TRPs showed a significant negative correlation (r=–0.530, p=0.001).

We also analyzed all the speakers' pauses to see whether they become closer in length over the course of the dialogue (if they converge). This was done by examining the relationship between difference in pause lengths between the speakers in the dialogue and time: if there is convergence there should be a significant negative correlation between difference in pause lengths and time. Three of the dialogues (D3, D4 and D6) show indications of convergence.

Table 1: Relationships between pause lengths in D1-D6.

	Synchrony	Convergence/divergence
D1	0.621**	–0.223 (p=0.087)
D2	–0.635**	0.717**
D3	.0327*	–0.613**
D4	.0391**	–0.306*
D5	–0.175 p=0.132	0.059 (p=0.666)
D6	0.405**	-0.317**

*Significant correlation at the .05 level; **Significant correlation at the .01 level

4 Discussion

In our data we found evidence of pause length entrainment in the form of synchrony in 4 of the 6 dialogues, and in the form of convergence in 3 of the 6 dialogues. Dialogue 2 stands out from the others by showing significant strong evidence of asynchrony and divergence in the speakers' pause lengths: we plan to examine this dialogue more closely and see whether there are other signs of disentrainment, and how it differs from the other dialogues. In this study we have predominantly treated pauses as one group, whether or not they occur at TRPs. However, in future studies it would be advisable to also analyze pauses as two groups, as they might behave in different ways (which we saw in dialogue 5).

References

Brennan, S. 1996. Lexical entrainment in spontaneous dialog. *Proceedings of ISSD*, 41–44.

Edlund, J., Heldner, M., & Hirschberg, J. 2009. Pause and gap length in face-to-face interaction. *Tenth Annual Conference of the International Speech Communication Association*, 2779–2782.

Levitan, R., & Hirschberg, J. 2011. Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions. *Twelfth Annual Conference of the International Speech Communication Association*, 3–6.

Pardo, J. 2006. On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America* 119, 2382–2393.

Pickering, M. J., & Garrod, S. 2004. Toward a mechanistic psychology of dialogue. *Behavioural and Brain Sciences* 27, 169–226.

Roberts, F., Francis, A., & Morgan, M. 2006. The interaction of inter-turn silence with prosodic cues in listener perceptions of "trouble" in conversation. *Speech Communication* 48, 1079–1093.

Sacks, H., Schegloff, E., & Jefferson, G. 1974. A simplest systematics for the organization of turn-taking for conversation. *Language* 50, 696–735.

Tannen, D. 1985. Silence: anything but. In: Tannen, D., & Saville-Troike, M. (eds), *Perspectives on Silence*. Norwood, NJ: Ablex.

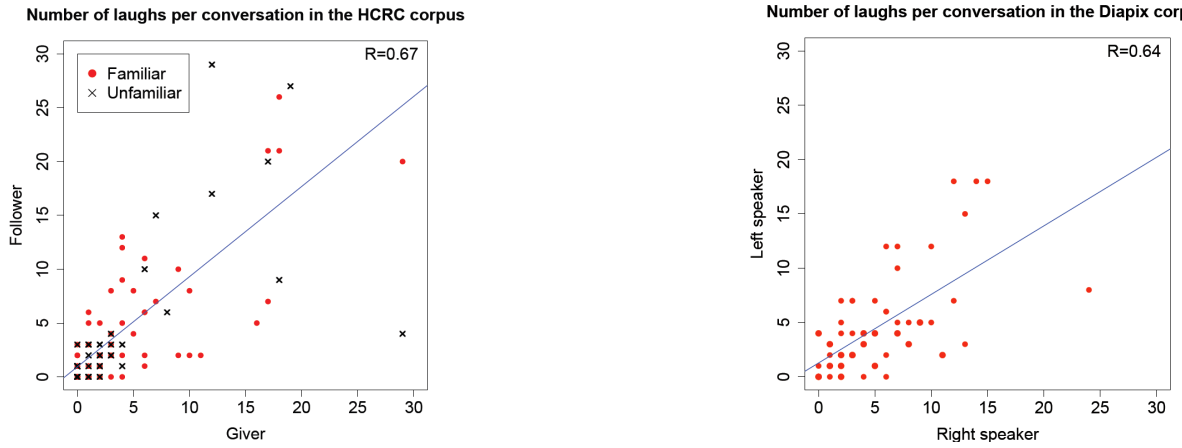
Convergence of laughter in conversational speech: effects of quantity, temporal alignment and imitation*

Jürgen Trouvain[1] *and Khiet P. Truong*[2]

[1]Saarland University
[2]University of Twente
trouvain@coli.uni-saarland.de, k.p.truong@utwente.nl

A crucial feature of spoken interaction is joint activity at various linguistic and phonetic levels that requires fine-tuned coordination. This study gives a brief overview on how laughing in conversational speech can be phonetically analysed as partner-specific adaptation and joint vocal action. Laughter as a feature of social bonding leads to the assumption that when laughter appears in dialogues it is performed by both interlocutors. One possible type of convergence is when the conversational partners adapt their **amount of laughter** during their interaction. This partner-specific adaptation for laughter has been shown by Campbell (2007a). Persons, initially unknown to each other and without any negative attitude to the unknown partner, had to talk in ten consecutive 30-min conversations (interval of one week). With each conversation the level of familiarity increased which was also reflected by the increasing number of their laughs. Smoski & Bachorowski (2003) also showed that familiarity plays a big role for the number of laughs: friends laugh more often together than strangers do. But there is also evidence that the level of social distance plays a role for phonetic convergence/divergence in speech in terms of extended voice onset time in stop consonants (Abrego-Collier et al., 2011). Figure 1 illustrates the convergence effect in terms of the number of laughs for two speech corpora of task-based dyadic conversations (Anderson et al., 1991 for a map task; Baker & Hazan, 2011 for a spot-the-difference game) with rather high correlation values. However, the familiarity effect based on the experimental data of Smoski & Bachorowski (2003) could not be confirmed with the conversational data of the Map Task Corpus (Anderson et al., 1991).

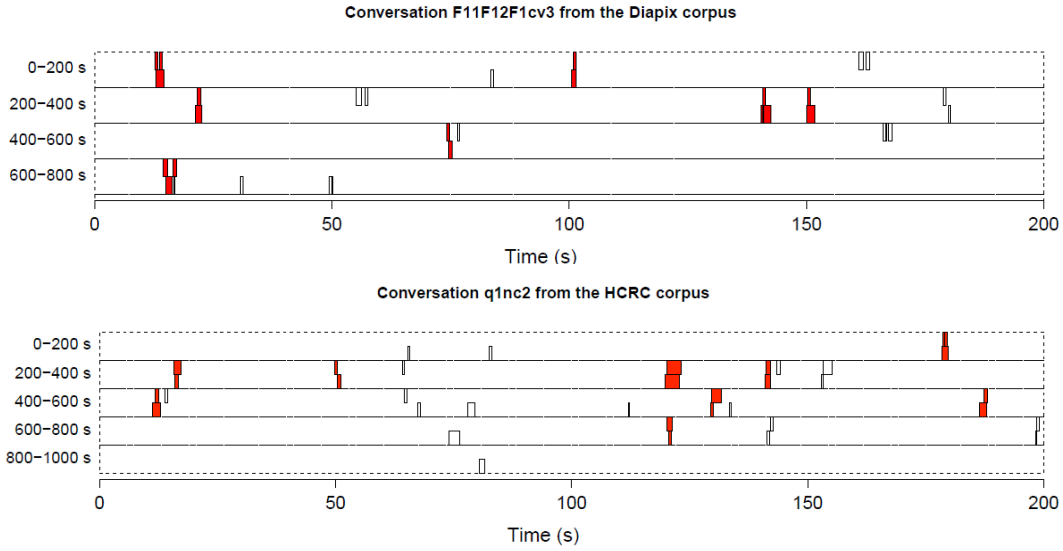
Figure 1: Correlations of number of laughs in the conversations of the HCRC Map Task Corpus divided into conversational partners who were familiar with each other or not (left) and the Diapix Lucid Corpus with friends only (right). Multiple occurrences of combinations are not visible here.



An even more partner-specific adaptation is the **temporal alignment** of laughter in conversations. In conversations the paradigm of "one speaker at a time" seems valid, for instance in a larger cross-linguistic study Stivers et al. (2009) show "that all of the languages tested provide clear evidence for a general avoidance of overlapping talk". But there are also studies on conversational speech observing a substantial amount of overlapping vocalization, mainly as "cross-talk" (e.g. Campbell, 2007b or Heldner & Edlund, 2010). But particularly laughter has a tendency to overlap with laughter as could be shown by Laskowski & Burger (2007), Truong & Trouvain (2012b) and also Smoski & Bachorowski (2003). Obviously laughter seems to represent an optimal opportunity for joint vocalization. Such a temporal alignment can sometimes also be observed in spontaneous speech where we can find collaborative completions (Local, 2005) as continuations of the conversational partner with matching prosodic features. This type of emergent coordination is probably less often observed in contrast to planned vocal coordination in choir singing, ritualized community talking in church (e.g. common praying) and experiments with synchronous reading (Cummins, 2007). Figure 2 gives two examples for the close temporal vicinity of laughs in conversations which often lead to partial overlap of laughs.

*This work was partly supported by the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 231287 (SSPNet) and the UT Aspasia Fund.

Figure 2: Laugh activity plot for the conversation F11F12F1cv3 (Diapix) and q1nc2 (Map Task). In each track of 200 sec each bar stands for a laugh (red: both speakers overlap). Bar width represents duration of the laugh.



Laughter also seems to represent a good candidate for **phonetic imitation** when both interlocutors are laughing synchronously. In two recent studies (Truong & Trouvain, 2012a,b) we could show for various corpora of conversational speech that overlapping laughs are stronger prosodically marked than non-overlapping ones, in terms of higher values for duration, mean F0, mean and maximum intensity, and the amount of voicing. This effect is intensified by the number of people joining in the laughter event. We also found that group size affects the amount of overlapping laughs which illustrates the contagious nature of laughter and which could be interpreted as entrainment at group level.
In summary, laughter as a cue for entrainment/convergence is mirrored by the number of laughs of conversational partners and especially by their temporal alignment resulting in overlapping laughs. Thus, laughing in social interactions is a joint vocal action *par excellence* which is also reflected by its acoustic forms. Future research has to show the fine-grained mechanisms of the temporal and acoustic interplay of speakers laughing together and how this interplay is perceived by listeners.

References

Abrego-Collier, C., Grove, J., Sonderegger, M., & Yu, A. C. L. 2011. Effects of speaker evaluation on phonetic convergence. *Proceedings of ICPhS 2012*, Hong Kong, 192–195.

Anderson, A. H., et al. 1991. The HCRC Map Task Corpus. *Language and Speech* 34, 351–366.

Baker, R., & Hazan, V. 2011. DiapixUK: task materials for the elicitation of multiple spontaneous speech dialogs. *Behavior Research Methods* 43, 761–770.

Campbell, N. 2007a. Whom we laugh with affects how we laugh. *Proc. Interdisciplinary Workshop on The Phonetics of Laughter*, Saarbrücken, 61–65.

Campbell, N. 2007b. Approaches to conversational speech rhythm: speech activity in two-person telephone dialogues. *Proc. Int. Congress of the Phonetic Sciences*, Saarbrücken, 343–348.

Cummins, F. 2007. Speech synchronization: Investigating the links between perception and action in speech production. *Proc. Int. Congress of the Phonetic Sciences*, Saarbrücken, 529–532.

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

Laskowski, K., & Burger, S. 2007. Analysis of the occurrence of laughter in meetings. *Proc. Interspeech*, 1258–1261.

Local, J. 2005. On the interactional and phonetic design of collaborative completions. In: Hardcastle, W., & Beck, J. (eds), *A Figure of Speech: a Festschrift for John Laver*. New Jersey: Lawrence Erlbaum, 263-282.

Smoski, M. J., & Bachorowski, J. A. 2003. Antiphonal laughter in developing friendships. *Annuals of the New York Academy of Sciences* 1000, 300–303.

Stivers, T., et al. 2009. Universals and cultural variation in turn-taking in conversation. *Proc. National Academy of Sciences of the United States of America* 106, 10587–10592.

Truong, K. P., & Trouvain, J. 2012a. Laughter annotations in conversational speech corpora possibilities and limitations for phonetic analysis. *Proc. Workshop on Corpora for Research on Emotion Sentiment and Social Signals*, Istanbul.

Truong, K. P., & Trouvain, J. 2012b. On the acoustics of overlapping laughter in conversational speech. *Proc. Interspeech*, to appear.

Speech imitation between speakers influences the realization of initial rises in French intonation

Amandine Michelas and Noël Nguyen

Laboratoire Parole et Langage, Aix-Marseille Université & CNRS, Aix-en-Provence
michelas@lpl-aix.fr, noel.nguyen@lpl-aix.fr

Over the last few years, interest in imitation has widened across many disciplinary fields, including phonetics and phonology. Within these fields, a growing number of studies have shown that the tendency of speakers to imitate each other during a conversational exchange affects not only segmental features but also suprasegmental attributes such as rate of speech and silent pauses (**?**), vocal intensity (**?**) or pitch (**???**, *inter alia)*. However, little is known about how imitation affects intonation, especially in a language like French that has no lexical stress.

In French, stress is postlexical and pertains to a phrasal domain which is smaller than the intonation phrase. In Jun & Fougeron's autosegmental-metrical model of French intonation (**?**) casted within the autosegmental-metrical framework of intonation, this domain is called the Accentual Phrase or AP. The AP is characterized by the presence of a typical final f0 rise (LH*) on the last syllable of the phrase which is lengthened. In addition, an optional initial rise (LHi) may appear near the beginning of the AP (i.e. generally on the first syllable of the first content word occurring in the phrase). Figure 1 shows the Noun Phrase "la maison de Monet" *Monet's house* where the first AP "La maison" can be either pronounced with only a final rise (LH*, left) or with an additional initial High tone (LHiH*, right). Both the early and late rises can be described as a sequence of Low and High tones, but only the final rise is a pitch accent associated with a metrical strong syllable. When all four tones are produced, the tonal pattern of the AP is LHiLH*. Other attested sequences can be formed by the absence of one or more tones (LLH*, LHiH* and HiLH* and LH*). The factors favouring the production of the initial rise are not entirely clear, but have been claimed to include a large number of syllables in the phrase and a slow speaking rate (**???**, *inter alia)*. In this study, we argue that speech imitation between speakers is another factor influencing the realization of initial rises in French. Specifically, we hypothesized that participants in a shadowing task would produce more initial H tone when they heard a stimulus including both initial and final H tones than when only a final H tone was present in the auditory stimulus.

21 pairs of noun phrases whose segmental structure was identical but differing in the potential placement of an initial High tone near the beginning of the first AP were presented to listeners (Figure 1). In a shadowing task, 6 native speakers of French (3 males and 3 females) listened to target phrases and were instructed to repeat them, first, without any instructions to imitate stimuli (repetition task), and then with explicit instructions to imitate the speaker's way of producing the stimuli (imitation task).

We employed a mixed logit model to examine if auditory stimuli influenced the realization of initial H tones in the tonal patterns produced by speakers. The results showed that the tonal pattern of the auditory stimuli had a significant effect on tonal patterns produced by participants (β=2.8, se=0.49, z=5.79, p$<$0.0001) while the kind of task (repetition vs. imitation) had no significant effect (β =1.0, se=0.5147, z=1.867, p=0.052). Percentages of tonal patterns produced by participants which include either only a final H target (without Hi Response) or an additional/initial H target (Including Hi Response) during the two tasks for both Without Hi Stimuli and Including Stimuli are shown in Figure 2.

First, in contrast with previous studies on French intonation suggesting that initial accents are difficult to identify (**??**), our results indicate that participants readily differentiated among the stimuli including only a final H tone and stimuli including an additional/optional initial H tone. Secondly, we found that participants produced more initial H tones when they heard a stimulus including both initial and final H tones than when only a final rise was present; this was true for both tasks. These findings suggest that between-speaker speech imitation influences the realization of initial rises in French intonation and highlights the need to include a between-speaker accommodation mechanism in models of speech production.

References

Astésano, C. 2001. *Rythme et Accentuation en Français*. Paris: L'Harmattan.

Figure 1: f0 contour for two noun phrases "La maison de Monet" pronounced in isolation (as an Intonation Phrase) with only a final rise (LH*, left) and with an additional/optional initial high tone (LHiH*, right).

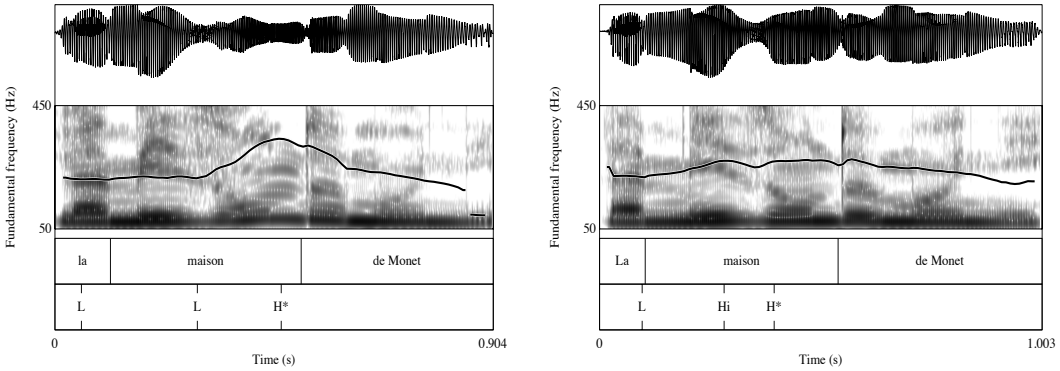
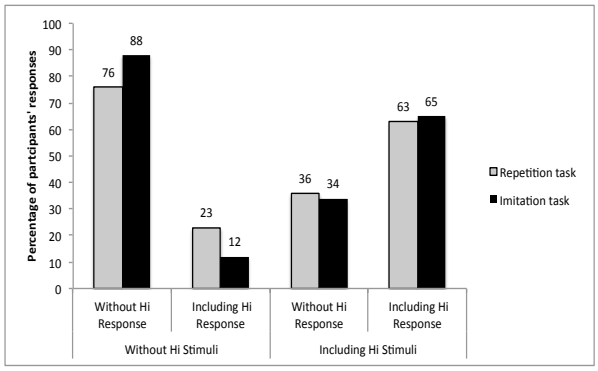


Figure 2: Percentages of responses including only a final H tone (Without Hi responses) or with an additional initial H tone (Including Hi responses) for stimuli including only a final H tone (Absence of Hi in the stimuli) or an additional initial H tone (Presence of Hi in the stimuli) for both repetition and imitation tasks.



Braun, B., Kochanski, G., Grabe, E., & Rosner, B. S. 2006. Evidence for attractors in English intonation. *J. Acoustical Society of America* 119, 4006–4015.

Dilley, L., & Brown, M. 2007. Effects of pitch range variation on F0 extrema in an imitation task. *J. Phonetics* 35, 523–551.

Giles, H., Coupland, N., & Coupland, J. 1991. Accommodation theory: Communication, context, and consequence. In: Giles, H., Coupland, N., & Coupland, J. (eds), *Contexts of Accommodation: Developments in Applied Sociolinguistics*. Cambridge: Cambridge University Press, 1–68.

House, D., Hermes, D., & Beaugendre, F. 1997. Temporal alignment categories of accent-lending rises and falls. *Proceedings of the 5th European Conference on Speech Communication and Technology* 2, 879–882.

Jun, S. A., & Fougeron, C. 2000. A phonological model of French intonation. In Botinis, A. (ed), *Intonation: Analysis, Modelling and Technology*. Boston: Kluwer Academic Publishers, 209–242.

Michelas, A. 2011. *Caractérisation phonétique et phonologique du syntagme intermédiaire en français: de la production à la perception.* Thèse de Doctorat de l'Université d'Aix-Marseille I.

Natale, N. 1975. Convergence of mean vocal intensity in dyadic communication as a function of social desirability. *J. Personality and Social Psychology* 37, 790–804.

Pierrehumbert, J., & Steele, S. A. 1989. Categories of tonal alignment in English. *Phonetica* 46, 181–196.

Vaissière, J. 1997. Langues, prosodies et syntaxe. *Traitement Automatique des Langues* 38, 53–82.

Welby, P. 2006. French intonational structure: Evidence from tonal alignment. *J. Phonetics* 34, 343–371.

The bilingual advantage in phonological learning

Laura Spinu and Yulia Kondratenko

Classics, Modern Languages, and Linguistics – Concordia University Montreal, QC, Canada
lspinu@alcor.concordia.ca

Recent studies have shown that bilinguals tend to be better at learning foreign languages in adulthood, and generally outperform monolinguals on certain types of non-linguistic tasks, e.g. those involving selective attention (Kovacs & Mehler, 2009; Costa et al., 2008; Colzato et al., 2008; Bialystok et al., 2005), as well as linguistic tasks, e.g. manipulating language in terms of discrete phonemic units (Bialystok et al., 2005; Bruck & Genesse, 1995) and novel word acquisition (Kaushanskaya & Marian, 2009). Focusing specifically on foreign language learning in adulthood, one of the most difficult aspects to master with native-like proficiency is represented by the phonological and phonetic properties of the target language. The concept of "phonological deafening" (Aslin et al., 1998; Polka & Werker, 1994; Werker & Lalonde, 1988; Werker & Tees, 1984), referring to the gradual loss, in early childhood, of the ability to distinguish phonemic contrasts not present in one's native language, illustrates this difficulty. Furthermore, the existence of accents, defined as a deviation from native speaker norms in the production of L2 sounds and sound combinations, also supports this claim.

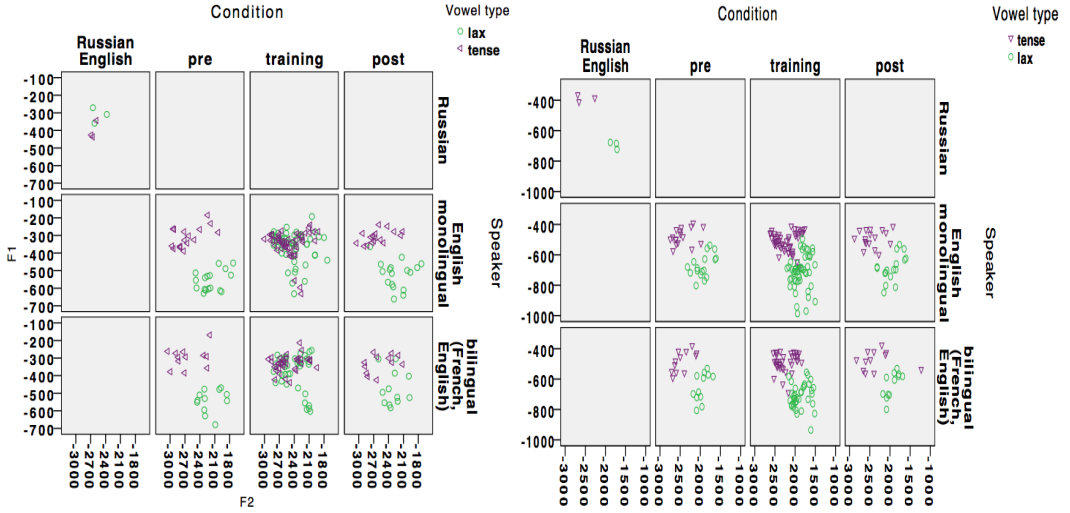
In light of these observations, we address the question of whether the "bilingual advantage" also applies to phonological learning. We compared monolinguals and bilinguals in a production experiment with two tasks: imitation and spontaneous reproduction of a novel foreign accent. Despite criticism according to which imitation does not reflect linguistic skill, producing behavior with no "carry-over into the post-imitative tasks" (Barry, 1989), we opted for the inclusion of this condition as recent studies have demonstrated that imitation alone improves the comprehension of a novel accent (Adank et al., 2010), and following the example of other studies which have used direct imitation in language tests (Delvaux et al., 2011; Markham, 1997; Neufeld, 1987).

We collected experimental data from 29 speakers (17 monolingual English and 12 French-English bilinguals from Quebec, Canada) who were trained to produce English with a Russian accent. A preliminary analysis addressed subjects' ability to neutralize the tense-lax vowel contrast in reproducing the Russian English accent, e.g. producing words like *beat* and *bit* with the same vowel sound, specifically [bit]. While tense-lax neutralization is a strong marker of a Russian accent in English, in the accent we employed, this was the case for only 3 of the 4 tense-lax vowel pairs of English, as the Russian speaker we recorded produced a consistent tense-lax distinction with mid front vowels (*bait* and *bet*), as reflected by F1/F2 values. All other tense-lax pairs (high front, high back, and mid back), were neutralized to tense in her speech.

F1 and F2 measurements were obtained for two vowel pairs (high front and mid front), with the remaining two vowel pairs (high back and mid back) and overall duration measurements currently being under way. Preliminary results show that none of the subjects neutralized the tense-lax distinction in mid front vowels. For high front vowels, both monolinguals and bilinguals were capable of native-like production (that is, complete neutralization of the distinction when asked to imitate sentences spoken in this accent), but the bilinguals, as a group, were closer to this pattern when asked to spontaneously produce novel sentences without prior audio prompts. While prior to accent training the bilinguals' F2 values were significantly different in *heat* versus *hit*, this was no longer the case when they spontaneously imitated the Russian English accent. Monolinguals produced significant differences in both conditions. As for F1, both monolinguals and bilinguals produced significant differences between the vowels in *heat* versus *hit* across the board. Figure 1 displays F1/F2 vowel plots for the Russian speaker, as well as the monolingual and bilingual subjects in three conditions: (i) *pre*: prior to accent exposure, that is, the native production of these vowels, (ii) *training*, during which the subjects listened to and immediately imitated sentences spoken with a Russian English accent, and (iii) *post*, with the subjects producing spontaneously sentences they had not previously heard, and trying to reproduce the Russian English accent to the best of their ability.

To summarize, we have found differences between tense-lax neutralization patterns in high front as compared to mid front vowels, showing that the subjects paid attention to the particularities of the accent to which they had been exposed. Differences were also found between imitation (training) and spontaneous reproduction of

Figure 1: Scatterplot of F1/F2 values for tense and lax high front vowels, as in *beat* and *bit* (left) and mid front vowels, as in *bait* and *bet* (right).



the new accent, with both monolinguals and bilinguals successfully producing the Russian pattern in imitation, but not (always) in spontaneous reproduction. Most notably, a difference was noted in the behavior of monolinguals as compared to bilinguals, the former not having been successful at spontaneously reproducing the new pattern, while the latter produced partial neutralization (an intermediate form between [i] and [ɪ]), thus showing stronger learning effects. This intermediate form shows the process of phonological learning *at work* and, very importantly, the initial learning that occurred was achieved through *imitation*.

To account for these findings, we discuss the concept of echoic memory (Calabrese, 2011), a mechanism by which sensory representations of speech uttered by others can be stored and checked against different mental representations, until the acoustic patterns stored in echoic memory can either be ascribed to existing phonological representations (e.g. in the case where one becomes able to parse correctly mispronunciations due to a speech defect), or be converted in licit phonological representations (e.g. when an L2 learner acquires a non-native sound). From this perspective, bilinguals' echoic memory strategies may differ from those of monolinguals (having been employed more intensively in the early/simultaneous acquisition of two languages), such that novel intermediate phonological representations are arrived at more rapidly. Our study adds to the body of work suggesting that there is an advantage of bilingualism in foreign language learning in adulthood, and offers an explanation in terms of perceptual strategies in which echoic memory is involved.

References

Adank, P., Hagoort, P., & Bekkering, H. 2010. Imitation Improves Language Comprehension. *Psychological Science* 21, 1903–1909.

Aslin, R. N., Jusczyk, P. W., & Pisoni, D. B. 1998. Speech and auditory processing during infancy: Constraints on and precursors to language. In: Kuhn, D., & Siegler, R. (eds), *Handbook of Child Psychology: Cognition, Perception, and Language*. New York: Wiley.

Barry, W. 1989. Perception and production of English vowels by German learners: Instrumental-phonetic support. *Phonetica* 46, 155–168.

Bialystok, E., Martin, M. M., & Viswanathan, M. 2005. Bilingualism across the lifespan, the rise and fall of inhibitory control. *International Journal of Bilingualism* 9.

Bruck, M., & Genesse, F. 1995. Phonological awareness in young second language learners. *Journal of Child Language* 22, 307-324.

Calabrese, A. 2011. Auditory representations and phonological illusions: A linguist's perspective on the neuropsychological bases of speech perception. *Journal of Neurolinguistics Online*, publication date: 1-Jun-2011.

Costa, A., Hernandez, M., & Sebastian-Galles, N. 2008. Bilingualism aids conflict resolution, Evidence from the ANT task. *Cognition* 106.

Kovacs, A., & Mehler, J. 2009. Cognitive gains in 7-month-old bilingual infants. *Proc. National Academy of Sciences* 106, 16.

Colzato, L. S., Bajo, M. T., van den Wildenberg, W., Paolieri, D., Nieuwenhuis, S., La Heij, W., & Hommel, B.2008. How does bilingualism improve executive control? A comparison of active and reactive inhibition mechanisms. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 34.

Kaushanskaya, M., & Marian, V. 2009. The bilingual advantage in novel word learning. *Psychonomic Bulletin and Review* 16, 705–710.

Functional causality of the dorsal stream in sensorimotor integration of speech repetition

Takenobu Murakami[1,2]*, Yoshikazu Ugawa*[1]*, & Ulf Ziemann*[2]

[1]Department of Neurology, Fukushima Medical University, Fukushima, Japan
[2]Department of Neurology, Goethe-University, Frankfurt am Main, Germany
maaboubou@gmail.com,ugawa-tky@umin.net,u.ziemann@em.uni-frankfurt.de

Introduction

Acquisition of language skills in humans involves social interactions in earliest infancy, and speech repetition plays a fundamental role by mapping auditory speech input onto matching speech output (Hickok & Poeppel, 2004; Wernicke, 1874). At the level of neuronal networks in the brain, speech repetition occurs within the auditory dorsal stream, which is constituted by a temporo-parieto-frontal network formed by the posterior part of superior temporal sulcus (pSTS) representing a sensory phonological processing area, the posterior inferior frontal gyrus (pIFG) involved in motor articulation, and the temporo-parietal junction (area Tpj) forming sensorimotor integration of speech processing (Buchsbaum et al., 2011; Hickok et al., 2011; Hickok & Poeppel, 2007). However, it is less known which component of the dorsal stream plays an essential role in the modulation of speech repetition. To investigate the issue of causality of the dorsal stream in sensorimotor integration of speech repetition, we employed continuous theta-burst transcranial magnetic stimulation (cTBS) in a "virtual lesion mode" (Ziemann, 2010) to disrupt neuronal activity in the stimulated areas.

Methods

Nineteen right-handed German volunteers participated in this study. Four behavioral speech tasks were given after cTBS. (1) Word-picture matching test: subjects listened to a German noun and were asked to choose a picture which fit to the word heard from four different pictures; the target picture, pictures implying a phonemic or semantic error, and an unrelated picture. This matching test was performed at three different auditory noise levels. (2) Syllable repetition test: subjects listened to one of six different syllables (Ba, Da, Ga, Ka, Pa, Ta) under three noise levels and repeated the perceived syllable immediately. (3) Pseudo-word repetition test: subjects listened to one of 15 meaningless pseudo-words under three noise levels and immediately repeated it. (4) Sentence repetition test: subjects listened to one of 30 German sentences under three noise levels and immediately repeated it. In all tests, error rates (ERs) and reaction time (RTs) at each noise level were calculated. Inhibitory cTBS was applied over the individual activated regions of pIFG, Tpj, pSTS of the left hemisphere by using an fMRI-guided TMS neuronavigation system. Left middle occipital gyrus (MOG) was defined as a control region to clarify the topographical specificity of cTBS effects because it is assumed that the MOG is concerned with visual processing (Restle et al., in press). The behavioral speech tests were performed after cTBS and the effects of cTBS sites (four levels) on test performance were tested by repeated measures ANOVAs.

Results

(1) In word-picture matching test, cTBS of Tpj and pSTS increased phonemic errors at middle-level noise and cTBS of Tpj and pIFG increased phonemic errors at high-level noise when compared to cTBS of MOG. However, the other types of errors (semantic, unrelated errors, all kinds of errors, and no-response) showed no significant difference between sites of cTBS. (2) ERs of syllable repetition after cTBS of Tpj, pSTS and pIFG were larger than those after cTBS over MOG. (3) Pseudo-word repetition test demonstrated that ERs increased after cTBS of Tpj and pSTS when compared to those after cTBS of MOG and of pIFG at low-level noise. At middle noise level ERs after cTBS of Tpj were larger than MOG. These results indicated that Tpj and pSTS are situated at hierarchically higher level and integrate phonological perception onto motor articulation through pIFG. (4) In the sentence repetition test, no differences of ERs were observed between sites of cTBS, although

a significant noise level effect was shown. Throughout all the behavioral testing, RTs were not significantly different between sites of cTBS.

Discussion

Disruption of the dorsal stream by inhibitory cTBS led to increases of phonemic errors in word-picture matching test and to increases of errors of syllable and pseudo-word repetitions, demonstrating that the dorsal stream plays a crucial role in sensorimotor integration of phonological processing. Perception/repetition of native sentences was not altered, largely excluding modulation of working memory. The lack of cTBS effects on native sentences was explained by processing of lexical/semantic material in the ventral auditory stream (e.g. Hickok & Poeppel, 2004). The sensory phonological system performs critical modulation of subsequent articulation in the motor system via a sensorimotor interface, supporting evidence that the sensory system is situated hierarchically at higher level to modulate the motor articulation.

References

Buchsbaum, B. R., Baldo, J., Okada, K., Berman, K. F., Dronkers, N., D'Esposito, M., & Hickok, G. 2011. Conduction aphasia, sensory-motor integration, and phonological short-term memory - an aggregate analysis of lesion and fMRI data. *Brain and Language* 119, 119–128.

Hickok, G., & Poeppel, D. 2004. Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition* 92, 67–99.

Hickok, G., & Poeppel, D. 2007. The cortical organization of speech processing. *Nature Reviews Neuroscience* 8, 393–402.

Hickok, G., Houde, J., & Rong, F. 2011. Sensorimotor integration in speech processing: computational basis and neural organization. *Neuron* 69, 407–422.

Restle, J., Murakami, T., & Ziemann, U. 2012. Facilitation of speech repetition accuracy by theta burst stimulation of the left posterior inferior frontal gyrus. *Neuropsychologia*, in press.

Wernicke, C. 1874. *Der Aphasische Symptomencomplex*. Breslau: Max Cohn and Weigert.

Ziemann, U., 2010. TMS in cognitive neuroscience: Virtual lesion and beyond. *Cortex* 46, 124–127.

Perceptually induced speech motor representations

C. Neufeld[1], R. Craioveanu[2], F. Rudzicz[3], W. Wong[4], and P. van Lieshout[1]

University of Toronto
[1]Speech-Language Pathology, [2]Linguistics, [3]Computer Science, [4]Electrical and Computer Engineering
christopher.neufeld@utoronto.ca

Recent neurological research has shown that motor areas associated with the movement of speech articulators are activated by the mere presentation of speech stimuli in both auditory and visual modalities (Fadiga et al., 2002; Pulvermüller et al., 2006; Watkins et al., 2003). Since the activity of these brain regions typically results in the overt movement of speech articulators, this invites us to question how perceptually induced speech-motor representations (PISR) differ from explicit motor plans. Listeners refrain from compulsively moving their articulators in response to speech stimuli, suggesting several possible explanations for the cortical motor activity observed in speech perception in the absence of explicit movement. Either PISR are a fundamentally different kind of motor representation from typical motor-plans, and the findings cited above simply reflect shared neural resources, rather than any cognitive overlap. Alternatively, the activation of speech-motor areas may reflect a priming effect, which is speech-specific, but not necessarily a detailed motor-plan. Or, finally, PISR may be fine-grained motor-plans which are filtered or damped by some antagonistic mechanism to prevent the involuntary movement of speech-articulators during speech perception. This latter hypothesis would provide a explanatory mechanism for convergence in speech as a form of entrainment: the speech motor plans of speakers are neurologically represented as such in listeners, and affect the fine-grained structure of the listener's motor plans when it is their turn to speak (Pickering & Garrod, 2004).

To distinguish these three possibilities, an experiment was devised which engages both explicit and tacit speech motor streams. Subjects are presented with prompts which instruct them either to produce two different **stimuli types** /ma/ or /na/ (labial or lingual) without vocalizing, to the beat of a visual metronome. There are three different **stimuli rates**: 2 Hz, 3 Hz, and 4.5 Hz. After several seconds of following the visual metronome, a rhythmic audio distractor is played over headphones. There are three **distractor rates**: a slower rhythm with respect to the visual metronome, on-beat or faster. Slow and fast rhythms are the target speed divided by, or multiplied by 1.5 respectively. Finally, there are three **distractor types**: nonspeech, matched and mismatched. Matched and mismatched distractors were created by recording the first author repeating /ba/ or /da/ at various speeds. Matched distractors have the same place of articulation as the target gesture (/ba/ in the case of /ma/, /da/ in the case of /na/), and mismatched distractors have the opposite place of articulation (/da/ in the case of /ma/). Speech distractors were automatically realigned to be exactly on the desired beat, and normalized for pitch and amplitude. Nonspeech distractors were created by superposing pure sine waves at the frequencies and amplitudes of F0-F3 of the speech distractors and multiplying the resulting complex wave by the amplitude envelope of speech distractors.

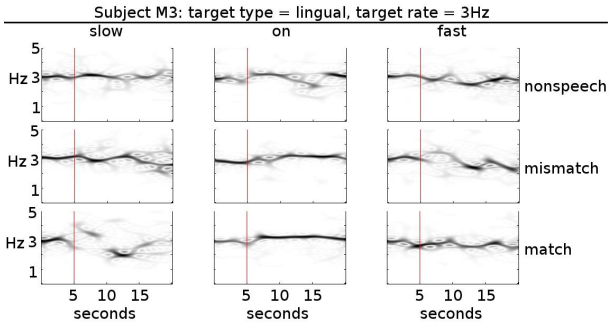
If PISR are fine-grained motor plans, we should expect that off-beat, matched distractors will be most disruptive to the accurate maintenance of the target frequency. For example, if a subject is attempting to produce a labial gesture at 3 Hz, and is suddenly presented with the audio of a labial speech gesture at 2 Hz, motor-areas responsible for driving the oscillations of the lips will be simultaneously representing 2 and 3 Hz, and this conflict should be detectable at the level of oral motor output. Mismatched off-beat speech distractors will be somewhat disruptive since there is some movement of lips and tongue for both labial and lingual gestures, but should not be as disruptive as matched off-beat speech distractors. Finally, nonspeech, off-beat distractors should be the least disruptive. If PISR are just a generic priming of the speech-motor system, organ matched and mismatched off-beat distractors should be equally distracting but more distracting than off-beat nonspeech, and speech on-beat speech distractors should be equally facilitative, but more facilitative than nonspeech. If PISR reflect shared neural, but not cognitive resources, we would predict that there should be no difference across distractor type conditions: off-beat distractors should all be equally distracting, regardless of type, and on-beat distractors should be equally facilitating, regardless of type.

Articulatory data was collected from 15 subjects with a 12-channel 3D Electromagnetic articulograph (EMA) with a sampling rate of 200 Hz. Gestures were derived from raw EMA data by calculating the Euclidean

distance between the tongue-tip and the nose, for lingual targets, and the distance between upper and lower lip, for labial targets. Figure 1 shows some data from one subject. Each panel is a time-frequency representation of the trajectory of the tongue-tip with a target frequency of 3 Hz.

Each column of panels represents a distractor rate, and each row a distractor type. The vertical lines indicate the onset of the auditory distractor. Disorder in the signal can be observed where the time-frequency representation is diffuse and has no strong single peak frequency, or where the peak frequency changes rapidly over short time spans. It can be seen that on-beat distractors facilitate the maintenance of a 3 Hz rhythm. However, the signal appears slightly more disordered for the nonspeech distractors than for speech in the on-beat condition, and the matched speech distractor shows the greatest degree of facilitation: a highly ordered signal with almost all its energy at 3 Hz. The most extreme instance of distraction occurs with the slow matched distractor. After the onset of the auditory distractor, the signal becomes disordered, with more energy in higher frequencies. At around 10 seconds, the signal is orderly once more, but most of its energy is around 2 Hz: the distractor frequency. From 15-20 s, the signal becomes highly disordered again, without a strong peak in any frequency band. By contrast, for mismatch and nonspeech, the slow auditory distractors do not appear to interfere with the maintenance of a 3 Hz rhythm to nearly so dramatic an extent. The signal is more disordered than for on-beat distractors, but only marginally so, and the signal is generally concentrated around 3 Hz.

Figure 1: Time-frequency representations of speech gestures.



To quantify the disorder of these time series, 2^{14}-point spectrograms were calculated from the gestural time series using a 1.25 second Hamming window. At each time step, the frequency bin with the highest amount of energy was calculated, producing a peak-frequency track. The absolute value of the derivative of this frequency track was used as a measure of disorder: for disordered signals, the peak-frequency changes rapidly (such as can be seen in the bottom left panel of Figure 1, and for ordered signals, the peak frequency changes slowly or not at all (such as can be seen in the bottom middle panel).

A 4-way ANOVA (stimuli type $\times$ stimuli rate $\times$ distractor type $\times$ distractor rate) was calculated, and showed significant main effects for all independent variables ($p < 0.05$). Post-hoc Tukey tests showed that matched distractors caused significantly more disordered gestures than either mismatched or nonspeech distractors, with no significant differences between the latter. Post-hoc tests also showed that fast distractors caused significantly more disorder than slow distractors, both of which cased significantly more disorder than on-beat distractors. The 4.5 Hz target signals were significantly more disordered than either the 3 Hz target signals or 2 Hz distractor signals. Finally, lingual gestures were significantly more disordered than labial disorders.

These preliminary observations suggest that the paradigm presented here is a productive research tool. The fact that off-beat distractors disrupt the accurate maintenance of a rhythmic speech gesture indicates that asynchronous auditory input does impact speech-motor output. Thus, there is support for hypothesis that PISR are organ-specific motor plans, since matched distractors cause significantly more short-term volatility in peak frequency than either mismatched or nonspeech distractors, while mismatched distractors are not significantly different from one another, arguing against the priming hypothesis.

References

Fadiga, L., Craighero, L., Buccino, G., & Rizzolatti, G. 2002. Speech listening specifically modulates the excitability of tongue muscles: a TMS study. *Eur. J. Neurosci.* 15, 399–402.

Pickering, M. J., & Garrod, S. 2004. Toward a mechanistic psychology of dialogue. *Behavioural and Brain Sciences* 27, 169–226.

Pulvermüller, F., Huss, M., Kherif, F., Martin, F. M. d. P., Hauk, O., & Shtyrov, Y. 2006. Motor cortex maps articulatory features of speech sounds. *Proc. Nat. Acad. Sci.* 103, 7865–7870.

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.

Plasticity of sensory-motor goals in speech production: Behavioral evidence from phonetic convergence and speech imitation

Marc Sato[1], *Krystyna Grabski*[1], *Maëva Garnier*[1], *Lionel Granjon*[1], *Jean-Luc Schwartz*[1], *Noël Nguyen*[2]

[1]Gipsa-Lab, CNRS & Grenoble Universités, France
[2]Laboratoire Parole & Langage, CNRS & Aix-Marseille Université
marc.sato@gipsa-lab.inpg.fr

Imitation is one of the major processes by which humans develop social interactions. In speech communication, imitative processes are used from birth to adulthood, as highlighted by children's mimicking abilities and by adult's tendency to automatically "imitate" a number of acoustic-phonetic characteristics in another speaker's speech. These adaptive changes are thought to play a key role in speech development/acquisition and to facilitate conversational exchange by contributing to setting a common perceptuo-motor link between speakers. Based on acoustic analyses of speech production in various laboratory tasks, the present study aimed to better characterize sensory-to-motor adaptive processes involved in unintentional as well as voluntary speech imitation, and to test possible motor plastic changes due to auditory-motor recalibration mechanisms.

Methods

Three groups of participants involved in speech production or imitation tasks were exposed via loudspeakers to vowel utterances spoken by different speakers. The first task was designed to induce unintentional imitation of acoustically presented vowels and to measure the magnitude of imitative changes in speech production as well as possible motor after-effects. To this aim, participants were instructed to produce vowels according to either an orthographic or an acoustic cue, without any instructions to repeat or to imitate the acoustic cues. A block design was used where participants produced a vowel target according first to an orthographic cue (baseline), then to an acoustic cue (phonetic convergence) and finally to an orthographic cue (motor after-effect). To compare phonetic convergence and voluntary imitation of the acoustic vowels, we asked the second group of participants to imitate the acoustically presented vowels. In a third task, we tested whether motor after-effects can also occur without prior unintentional or voluntary vowel imitation but only after auditory exposure of the acoustic targets.

The three tasks were performed in a soundproof room using the same experimental setting and participants' productions were recorded for offline analyses. A semi-automatic procedure was first devised for segmenting participants' recorded vowels (around 10000 utterances). For each participant, the procedure involved the automatic segmentation of each vowel based on an intensity and duration algorithm detection. The algorithm automatically identified pauses (with minimal duration of 1000 ms and low intensity energy inferior to 55 dB) between each vowel by marking boundaries. If necessary, these boundaries were hand-corrected, based on waveform and spectrogram information, so as to correctly mark the onset and offset of vowels. After individual sound file extraction of each vowel, omissions, wrong productions and hesitations were manually identified and removed from the analyses. Finally, for each vowel, F0 and F1 values were calculated from a period defined as $\pm$ 25 ms of the maximum peak intensity of the sound file. For each participant, median F0 and F1 values were first computed for each vowel and expressed in bark. For each experiment, median F0 and F1 exceeding $\pm$ 2 standard deviations from the mean were removed from the analyses.

Results

Phonetic convergence and imitation (Experiments A and B): For each participant and vowel, median F0- and F1-responses observed during the presentation of the acoustic cue were subtracted from the preceding baseline (i.e., median F0- and F1-responses observed in the preceding sub-block during the presentation of the orthographic cues). These values were then correlated with F0 and F1 values of the respective acoustic cue subtracted from the preceding baseline. Single subject correlation coefficients were calculated for both F0 and F1 and entered

into analyses of variance (ANOVA) with the experiment (phonetic convergence, imitation) as a between-subject variable. In addition, individual one-tailed t-tests (with Bonferroni corrected p-value) were performed for each experiment in order to test significant correlation coefficients (compared to zero).

ANOVA on single subject correlation coefficients for F0 show a significant effect of the task. In addition, correlation coefficients differed significantly from zero in both Experiment A and Experiment B. For F1, there was no significant effect of the task. Correlation coefficients also differed significantly from zero in both Experiment A and Experiment B.

After-effects (Experiments A, B and C): For each participant and vowel, median F0- and F1-responses observed during the second presentation of the orthographic cue were subtracted from the preceding baseline. These values were then correlated with F0 and F1 values of the respective acoustic cue subtracted from the preceding baseline. As previously, single subject correlation coefficients were calculated for both F0 and F1 and entered into analyses of variance (ANOVA) with the experiment (phonetic convergence, imitation, perceptual categorization) as a between-subject variable. In addition, individual one-tailed t-tests (with Bonferroni corrected p-value) were performed for each experiment in order to test significant correlation coefficients (compared to zero).

ANOVA on single subject correlation coefficients for F0 showed no significant effect of the task. In addition, correlation coefficients differed significantly from zero in both Experiment B and Experiment C but not in Experiment A. For F1, there was no significant effect of the task. Correlation coefficients did not differ significantly from zero in Experiments A, B and C.

Conclusion

These results demonstrate automatic imitative processes during speech communication even at a fine-grained acoustic-phonetic level and highlight the online plasticity of phonemic sensory-motor goals during speech production, although only for F0 acoustic parameter. They will be discussed in relation with forward and inverse internal models of speech production in which feedback control mechanisms allow evaluating the sensory consequence of the speech-motor act with actual sensory input in order to further control production.

Role of motor representations in perception and imitation of singing

Yohana Lévêque[1]*, Antoine Giovanni*[1]*, and Daniele Schön*[2]

[1]Laboratoire Parole et Langage, CNRS & Aix-Marseille Université
[2] Institut de Neuroscience des Systèmes, INSERM & Aix-Marseille Université
yohana.leveque@lpl-aix.fr,daniele.schon@univ-amu.fr,antoine.giovanni@mail.ap-hm.fr

Hypothesis of an internal simulation process as a basis for imitation and understanding of observed actions has received great attention from different disciplinary fields (Gallese & Goldman, 1998; Iacoboni, 2009). A set of studies about speech perception in particular have explored the question of to what extent motor representations are recruited in speech processing (Galantucci et al., 2006, for a review). Neuroimaging brought evidence that premotor areas could be activated during speech perception, and this in a somatotopic way (D'Ausilio et al., 2011), but not in every perceptual context (Sato et al., 2009; Osnes et al., 2011). We propose to explore mechanisms and conditions of this "motor resonance" during auditory perception using singing voice. Singing is a vocal behavior which shares common features with speech, but, while the articulatory part of the signal is not as important as in speech, the phonatory dimension is more important than in speech. Singing voice can thus be used to investigate the auditory-vocal loop with minimal linguistic issues, but directly addressing the question of the auditory-motor processes underlying vocal control.

We present here a series of three studies about motor representations in singing voice perception and imitation. More precisely, we studied the influence of timbre humanness and singing skill on motor activity. In the first study (Lévêque et al., 2012), adult participants with normal or poor singing skills were asked to sing a pitch after a natural vocal model or a synthesized complex sound sharing spectral similarities with voice. We found that poor singers only were affected by the timbre model, singing more accurately after a voice than after a synthesized sound.

To find out whether this advantage for human voice was due to motor activations, we carried out a second study using electroencephalography (EEG). Twenty participants were asked to listen to computer-generated or sung melodies, and then to vocally reproduce each sequence, while EEG was recorded. Analysis of beta-motor (20 Hz) and mu (10 Hz) brain rhythms during the perception periods showed that sung melodies induced a stronger motor activity than computer-generated melodies.

Furthermore, we found that the motor resonance was inversely proportional to participants' vocal accuracy. Results of both studies suggest that the mirror system is activated early during auditory perception of singing voice, more strongly for natural voice and in participants experiencing difficulties to execute the vocal task (poor singers). Poor singers may rely more on biomechanical representations linked to voice production than good singers when encoding the auditory target.

In the last study, we used a sound categorization task in a Transcranial Magnetic Stimulation protocol to study the involvement of motor representations in a nonimitative perceptual context. Participants received a continuous Theta Burst Stimulation over the right premotor larynx area (experimental group) or vertex (control group). Their performances in an auditory categorization task involving natural and distorted singing voice sounds were evaluated before and after stimulation.

Despite a general shortening of response times after stimulation, we observed longer response times for natural voice sounds compared to distorted sounds only after stimulation over larynx area. This result is in line with our hypothesis of an involvement of the premotor cortex in voice perceptual processing, at least in this context of timbre discrimination.

These studies provide some evidence of an activation of motor representations in perception of non-speech vocal stimuli, within imitative and non-imitative perceptual contexts. They suggest that articulatory movements are not the only action likely to induce a motor resonance in the listener, given that articulation was strongly

reduced in the sung stimuli we used. Phonatory gesture is by itself an action that can be mapped into bodily representations. D'Ausilio et al. (2011)'s recent study provides supporting data. Contribution of our results to the debate on the functional role of this mirror-like brain activity is discussed.

References

D'Ausilio. A., Pulvermüller, F., Salmas, P., Bufalari, I., Begliomini, C., & Fadiga, L. 2009. The motor somatotopy of speech perception. *Current Biology* 19, 381–385.

D'Ausilio, A., Bufalari, I., Salmas, P., Busan, P., & Fadiga, L. 2011. Vocal pitch discrimination in the motor system. *Brain and Language* 118, 9–14.

Gallese, V., & Goldman, A. 1998. Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences* 2, 493–501.

Galantucci, B., Fowler, C. A., & Turvey, M. T. 2006. The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review* 13, 361–377.

Iacoboni, M. 2009. Imitation, empathy, and mirror neurons. *Annual Review of Psychology* 60, 653–670.

Lévêque, Y., Giovanni, A., & Schön, D. 2012. Pitch-matching in poor singers: Human model advantage. *Journal of Voice* 26, 293–298.

Osnes, B., Hugdahl, K., & Specht, K. 2011. Effective connectivity analysis demonstrates involvement of premotor cortex during speech perception. *NeuroImage* 54, 2437–2445.

Sato, M., Tremblay, P., & Gracco, V. L. 2009. A mediating role of the premotor cortex in phoneme segmentation. *Brain and Language* 111, 1–7.

Prosodic matching as a sequential resource in naturally occurring interaction

Beatrice Szczepek Reed

University of York, UK
beatrice.szczepek.reed@york.ac.uk

Previous work on naturally occurring talk has shown that one of the primary concerns for conversational participants is the appropriate, and appropriately timed display of orientation to prior talk. This becomes most obvious in the split-second precision with which speakers achieve turn taking transitions (Sacks et al., 1974; Local et al., 1985, 1986; Wells & Peppè, 1996; Sidnell, 2001; Szczepek Reed, 2004). The necessity to display whether, and how a turn at talk is situated in relation to a previous turn is particularly pressing in those instances where one participant has issued the First Pair Part (FPP) of a (potential) Adjacency Pair (Schegloff, 2007); for example, a question or a first greeting token. Such turns make the production of a Second Pair Part (SPP), such as an answer or a return greeting, conditionally relevant, and thus create a set of constraints on what can appropriately be said next (Raymond, 2003). At this sequential location, one of the most fundamental decisions participants have to make comes into play: whether to continue the projected action trajectory of previous talk, or whether to start a new one.

The following excerpts from a radio phone-in programme, recorded in the Manchester area of the UK in 1984, present responding participants who take either of the two paths. Excerpt (1) shows a caller issuing a return greeting to the host's first greeting. Excerpt (2) shows a caller initiating a new first greeting instead. In both instances, prosodic matching, or lack of it, plays a decisive role:

(1) Brainteaser: Nigel

1	Host:	next is NIgel HIbbits;
2		who lives in PRESTwich.
3		**<<h> ↑ HI `NI:GE,>**
4	**Caller:**	**<<h> ↑ HI `DA:VE,>**
5	**Host:**	**how ARE ya.**
6	Caller:	.hh nOt too BAD,
7	Host:	GOOD to speak to you agAIn,

(2) Brainteaser: Ann

1	Host:	and we have ANN,
2		who lives in GORton.
3		who's FIRST.=
4		and then of COURSE,
5		After our two callers we do have RACHel back again.
6		**.h ANN.**
7		**HI.**
8	**(0.26)**	
9	**Caller:**	**<<breathy> HELL:ˆO:.>**
10	**Host:**	**<<breathy> HELL:ˆO:.>**
11		**<<h> how ARE you Ann,>**
12	Caller:	I'm FINE,
13		THANKS,
14	Host:	GOOD.

These excerpts demonstrate one of the most central ways in which prosodic matching is used as an interactional resource. Turns that are designed to continue a previous action trajectory typically match the immediately prior prosodic design by a previous speaker. Thus, in (1), the caller's return greeting matches at least four aspects of the prosodic delivery of the host's first greeting (high overall pitch register, a pitch step-up on the first syllable

hi, falling-rising intonation on the monosyllabic names *nige* and *dave*, and lengthening of final vowels). The host responds to the caller's prosodically matching return greeting by initiating a new sequence: *how are ya* (line 5) is the FPP of a new adjacency pair. In contrast, (2) shows a caller who does not imitate the host's prosodic (and lexical) delivery. While the host's first greeting consists of *name + hi*, produced as two separate intonation units, each with low falling intonation and modal voice quality, the caller's next turn is *hello* without an address item, delivered with breathy voice quality, sound lengthening, and rising-falling intonation. The host's reaction at line 10 reveals what this means from a participant perspective: instead of moving on to initiate the next adjacency pair, he issues a return greeting (*hello*), in spite of already having produced a first greeting earlier (lines 6-7). He thus treats the caller's second greeting not as a return greeting, which would allow him to progress to the next sequence, but instead as a new first greeting. The host's third greeting turn is delivered with prosodic matching (voice quality, sound lengthening, intonation), and thus designed as an SPP to the caller's turn. He moves on to a new sequence (*how are you ann*, line 11) immediately afterwards. These examples show that imitating an immediately prior speaker's prosody is a decisive cue for displaying a next turn's sequential status. Even if the lexical item is a candidate for an SPP (*hello* is an appropriate candidate for a return greeting following *hi*), it is not treated as such if it does not display prosodic matching. This presentation explores the interactional role of prosodic matching (Szczepek Reed 2012, 2010a, 2010b, 2009a, 2009b, 2006) and, specifically, how imitation is employed for "doing continuation" in natural talk.

References

Local, J., Wells, B., & Sebba, M. 1985. Phonology for conversation. Phonetic aspects of turn delimitation in London Jamaican. *Journal of Pragmatics* 9, 309–30.

Local, J., Kelly, J., & Wells, B. 1986. Towards a phonology of conversation: Turn-taking in Tyneside English. *Journal of Linguistics* 22, 411–37.

Raymond, G. 2003. Grammar and social organization: Yes/no interrogatives and the structure of responding. *American Sociological Review* 68, 939–967.

Sacks, H., Schegloff, E. A., & Jefferson, G. 1974. A simplest systematics for the organization of turn-taking for conversation. *Language* 50, 696–735.

Schegloff, E. A. 2007. *Sequence Organisation in Interaction. A Primer in Conversation Analysis*. Cambridge: Cambridge University Press.

Sidnell, J. 2001. Conversational turn-taking in a Caribbean English Creole. *Journal of Pragmatics* 33, 1263–1290.

Szczepek Reed, B. 2012. Beyond the particular: Prosody and the coordination of actions. *Language and Speech* 55, 12–33.

Szczepek Reed, B. 2010a. Speech rhythm across turn transitions in cross-cultural talk-in-interaction. *Journal of Pragmatics* 42, 1037–1059.

Szczepek Reed, B. 2010b. Prosody and alignment: A sequential perspective. *Cultural Studies of Science Education* 5, 859–867.

Szczepek Reed, B. 2009a. Prosodic orientation: A practice for sequence organization in broadcast telephone openings. *Journal of Pragmatics* 41, 1223–1247.

Szczepek Reed, B. 2009b. FIRST or SECOND: Establishing sequential roles through prosody. In Barth-Weingarten, D., Wichman, A., & Dehé, N. (eds), *Where Prosody Meets Pragmatics*. Bingley: Emerald, 205–222.

Szczepek Reed, B. 2006. *Prosodic Orientation in English Conversation* Basingstoke: Palgrave MacMillan.

Szczepek Reed, B. 2004. Turn-final intonation in English. In: Couper-Kuhlen, E., & Ford, C. E. (eds), *Sound Patterns in Interaction*. Amsterdam: John Benjamins, 97–119.

Wells, B., & Peppè, S. 1996. Ending up in Ulster: Prosody and turn-taking in English dialects. In: Couper-Kuhlen, E., & Selting, M. (eds), *Prosody in Conversation*. Cambridge: Cambridge University Press, 101–30.

Prosodic structuring imitation in French L1 context – A first step towards correcting phonetic-prosodic features in L2 French

Olivier Nocaudie[1] *& Corine Astésano*[1,2]

[1]Octogone-Lordat, E.A. 4156, Toulouse, France
[2]Laboratoire Parole et Langage, Université d'Aix-Marseille & CNRS, Aix-en-Provence, France
{olivier.nocaudie,corine.astesano}@univ-tlse2.fr

Imitative behaviours play a fundamental role in human communication, and are physiologically determined by the presence of mirror neurons (Studdert-Kennedy, 2002). Biologically, perception is thus moulded by a natural, unconscious urge to adapt to one's behaviour (Chartrand & Bargh, 1999). During childhood, imitation is an innate capacity used for communication with pairs: mimicking games of adults' behaviour (pretend parenting or dining…) serve essential functions for adaptation and child's development, even in developmental pathology contexts (Nadel, 2005). Imitative behaviours persist during adulthood with similar adaptive motivations (adaptation to a new situation, learning how to use new skills or tools, responding to social pressure…) but, along with non-verbal characteristics, they concern finer speech features (Giles et al., 1991; Baudonnière, 1997). Indeed, it has recently been shown that adaptation takes place between speakers during conversational interaction, resulting in variations at a very fine communicative level, namely the phonological and prosodic levels. This "phonetic convergence" phenomenon relies on speakers' abilities to perceive Fine Phonetic Details (FPD; Nguyen et al., 2009) and tends to persist after the actual interaction (Pardo, 2006). However, it still remains to be seen precisely how these imitative characteristics apply to L2 learning, and how they can partake in the construction of the target language's system. More precisely, to what extent can "phonetic talent", as defined by Jilka et al. (2007), impact on L2 learning, and how can these irrepressible imitative processes (phonetic convergence) be used in specific phonetic training activities?

In a first step towards this goal, this study aims at unravelling the imitation processes of prosodic cues by L1 listeners/speakers. Relatively few studies have tackled the issue of prosodic characteristics' imitation, particularly in French (see however Michelas & Nguyen, 2011, for Initial Accent reproduction). Moreover, the tasks proposed usually involve imitation only, but not different degrees of imitation. The main aim of this paper is to describe a scale of imitation from what is done during a non-conscious imitation to what a professional impersonator could do, i.e. from simple repetition to exaggerated mimicry in phonetic/prosodic terms. This scale should allow us to observe more precisely at what degree fine phonetic details are reproduced. This is ultimately intended to inform us on what phonetic characteristics the teacher/"imitee" needs to emphasize in order for the L2 learner to reach the appropriate phonetic goal.

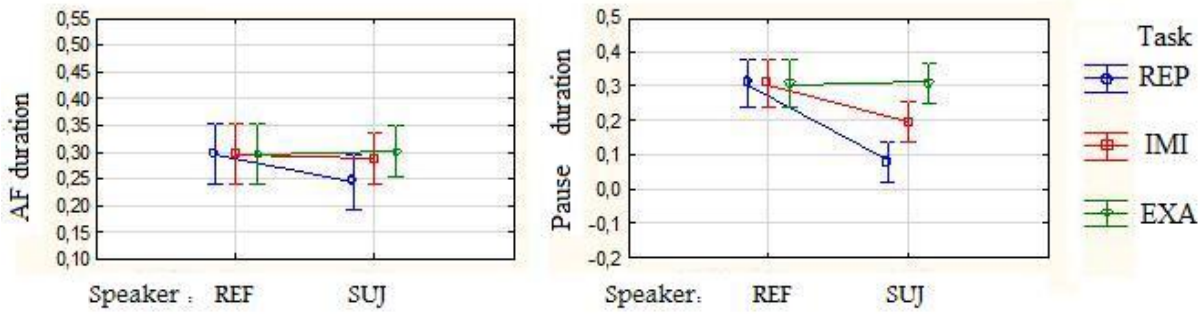
Eight native speakers of French had to perform three tasks: a repetition of prosodic structures, an imitation and finally an exaggerated imitation of these structures. Speakers/imitators performed the three tasks in separate blocks. Stimuli were randomized in each block throughout the speakers. They all began with the "repetition" task (REP), where they were instructed to just "*say the sentence so that the intended structure is preserved*"; they then performed the "imitation" task (IMI), where they had to "*imitate the perceived structure*", and they ended the recording session with the "exaggeration" task (EXA), where they were instructed to "*imitate the perceived structure until exaggeration*". The corpus consists of 8 syntactically ambiguous sentences that can be disambiguated via prosodic cues. Syntactic ambiguity is created by manipulating the adjective scope as in "les gants et les bas lisses", where the adjective (A) "lisses" either qualifies the second noun (N2) "bas" only ([les gants][et les bas lisses], hereafter *C1*, with an intonation phrase (hereafter ip) boundary between N1 and N2, or either the two nouns "gants et bas" ([les gants et les bas][lisses], hereafter *C2*, with an ip boundary between N2 and A. The corpus is read by one female speaker (REF). In *C1*, the ip boundary was produced by REF with a long pause and long final lengthening between N1 and N2. In *C2*, no pause was introduced between N2 and A, and final lengthening was less marked at the ip boundary. We created modified sentences, where the pause was erased in *C1* (hereafter *C1m*) and where a pause was inserted in *C2* (hereafter *C2m*), in order to produce conflicting Final Accent and Pause (hereafter AF and P) acoustic cues as boundary markers, and to measure the impact of FPD perception. We have overall 4 syntactic conditions (2 original *C1o* and *C2o*, and 2 modified *C1m* and *C2m*). The eight native speakers had to perform the three "imitation" tasks on the resulting 32 sentences (8 sentences $\times$ 4 syntactic conditions). Each sentence was repeated three times in randomized order in each block, so that our entire corpus consists of 2304 sentences.

This paper presents the results on a subset of this corpus, on four speakers/imitators only. As a first step, we present results on the reproduction of well-known boundary markers in French, namely AF and P. Anovas were run with speakers (REF vs. the 4 speakers/imitators), tasks (REP, IMI, EXA) and syntactic conditions (*C1o*, *C1m*, *C2o*, *C2m*) as factors, and P occurrence, P duration and AF duration as dependant variables. As an illustration, we present first results in the 2 conditions with a P (*C1o* and *C2m*).

As far as P occurrence is concerned, speakers did not consistently reproduce P in *C1o* in the REP task ($p = .0019$),

and started to come close to REF's realization in IMI and EXA ($p > .05$). When a P was added in *C2m*, the same tendencies were observed (REP: $p < .001$; IMI and EXA: $p > .05$). As far as P duration is concerned, speakers tended to reproduce a shorter P than REF in C1o in REP ($p = .0322$), and were closer to REF's realization in IMI and EXA ($p > .05$). For *C2m* however, P duration was significantly shorter in REP and IMI ($p < .001$ and $p = .0140$ respectively) and reached REF's realization in EXA only ($p > .05$). AF duration in *C1o* is shorter in REP ($p = .0219$) and close to REF in IMI and EXA ($p > .05$), whereas it is close to REF's realization in C2m ($p > .05$ in the three tasks) (see Figure 1 for an illustration on *C2m*).

Figure 1: AF and P duration (sec.) for speaker REF and the speakers/imitators (SUJ) in *C2m* throughout the three tasks.



Altogether, these results indicate that speakers can consistently reproduce Fine Phonetic Details (duration of P and AF) if they are explicitly asked to, and exaggeration does not induce better approximation than mere imitation. However, when conflicting acoustic cues are present (shorter AF in *C2m* where no P was originally there, followed by a long artificial P), speakers can precisely reproduce P duration only in the exaggeration task. In a more natural situation, these results have implication for L2 phonetic correction insofar as the teacher can use exaggeration to facilitate the perception/reproduction of specific prosodic structures.

Further results on the reproduction of fine acoustic cues in the other syntactic conditions are beyond the scope of this abstract but will be discussed in the presentation. The comparison of *C1o* and *C1m* is for example very interesting in order to precisely weigh the impact of AF and P cues in the marking of prosodic structure. Moreover, analyses will be extended to other phonetic cues of prosodic structuring, such as tonal patterns on AF and Initial Accent, which has been described as a consistent boundary marker in French (Astésano et al., 2007). Speech rate, F0 register, spectral characteristics and overall intensity will also be discussed, as they appear to vary throughout the three tasks and clearly approximate REF's characteristics in the exaggeration task. Finally, speakers/imitators individual strategies will be discussed on this REP to EXA scale, since some individuals seem to exhibit "phonetic talent" as soon as the repetition task.

References

Astésano, C., Bard, E. & Turk, A. 2007. Structural influences on Initial Accent placement in French. *Language and Speech* 50, 423–446.

Baudonnière, P. M. 1997. *Le mimétisme et l'imitation*. Paris: Flammarion.

Chartrand, T. L., & Bargh, J. A. 1999. The chameleon effect: the perception behavior link and social interaction. *J. Pers Soc. Psychol.* 76, 893–910.

Giles, H., Coupland, J., & Coupland, N. 1991. Accomodation Theory, communication, context and consequences. In: *Contexts of Accommodation*. Cambridge University Press, 1–61.

Jilka, M., et al. 2007. Assessing individual talent in second language production and perception. *Proceedings of the Fifth International Symposium on the Acquisition of Second Language Speech*, 243–258.

Michelas, A., & Nguyen, N. 2011. Uncovering the effect of imitation on tonal pattern of French Accentual Phrases. *Proceedings of Interspeech 2011*, Florence.

Nadel, J. 2005. Imitation et autisme. In: Berthoz, A., et al. (eds), *Autisme, Cerveau et Développement*. Paris: Odile Jacob, 341–356.

Nguyen, N., Wauquier, S., & Tuller, B. 2009. The dynamical approach to speech perception: from fine phonetic detail to abstract phonological categories. In: Pellegrino, F., Marsico, E., Chitoran, I., & Coupé, C. (eds), *Approaches to Phonological Complexity*. Berlin: Mouton de Gruyter, 193–217.

Pardo, J. S. 2006. On phonetic convergence during conversational interaction. *Journal of the Acoustical Society of America*, 119, 2382–2393.

Studdert-Kennedy, M. 2002. Mirror neurons, vocal imitation and the evolution of particulate speech. In: Stamenov, M. I., & Gallese, V. (eds.), *Mirror Neurons and the Evolution of Brain and Language*. Amsterdam: John Benjamins, 207–227.

Perceptual learning and convergence in sound change

Bridget Smith

Department of Linguistics, The Ohio State University, Columbus, OH, USA
bsmith@ling.ohio-state.edu

In the early days of linguistic study, sound change was a phenomenon that could only be inferred from textual analysis of written language across different time periods or comparative analysis across related languages. As technology has advanced and theories have evolved, we have developed new ways of investigating sound change, using recordings and phonetic analysis to look at synchronic variation within and across groups of speakers, and using perceptual measures to see how this variation may contribute to sound change. However, it has long been thought that the precise moment in which variation turns into sound change could not be witnessed. Synchronic variation could be examined, and sound-change-in-progress, considered to be the expansion phase following a sound change, could be observed, and phonological processes likely to lead to sound change could be probed, but the sound change itself could not be an object of study. That's all changing now. The following experimental breakthroughs are key to replicating sound change in laboratory conditions so that it can be studied: shadowing, perceptual learning, and the perception/production link.

In shadowing tasks, listeners repeat words after hearing them spoken aloud, and it has been found that these listeners tend to imitate aspects of the production they've heard, and that the imitation effect is stronger for less familiar words (Goldinger, 1998). The degree of convergence between talkers in a normal conversational setting is found to vary by sex, with women more likely to converge (Pardo, 2006). On the perceptual front, perceptual adjustment may result from hearing words with pronunciation variants. This effect may be limited to one talker (Eisner & McQueen, 2005), or may become generalized across talkers and grouping factors such as dialect (Clopper & Pisoni, 2004) or foreign accent (Bradlow & Bent, 2008). At the same time, some researchers have begun to put perception and production together, by finding that the areas in listeners' brains that are responsible for planning and executing speech production are also activated while listening to speech (Watkins et al., 2003). All of these findings have exciting implications for studying sound change.

A vocabulary learning experiment, incorporating perceptual learning and shadowing, was designed to induce sound change so that it could be studied in a laboratory setting. In order to examine sound change before it is a change-in-progress, there must be ambient variation that has not yet reached a point in which it is associated with a conditioning or indexical factor. In this experiment, existing variation in pronunciation of the stop+ approximant /tw/ cluster in American English was used as a basis for pushing the variation to the point of sound change. Approximants are known to increase the degree and length of aspiration in preceding stops, which sometimes can lead to the development of affricates. In American English, alveolar stops may become alveo-palatal affricates before /j/. Many American English speakers also palatalize and affricate /t/ before /r/. In /tw/, the lip-rounding that accompanies /w/ may spread to the preceding stop, which, by lengthening the front cavity, may create the percept of a retracted /t/. If the aspiration is strengthened, the resulting sound may be similar to an alveo-palatal affricate. However, the /t/ could be produced with a dental place of articulation, developing more anterior frication, yielding /ts/. Both a front and a retracted variant could then arise from the coarticulation of /t+w/. These two variants were used in separate conditions as the target of sound change. A third control group heard plain alveolar /t+w/.

Productions of words containing the tw- cluster were measured before training, during shadowing, and at the end of the experiment to chart the amount and direction of imitation of the pronunciation variant, and whether the effects would persist in post-training productions, and if subjects would extend the pronunciation variant to untrained words. A lexical decision task and an identification task also measured perceptual learning, and generalization to new talkers, new words, and new phonological environments. Generalization in production occurred for new /tw/ words, but also affected /tr/ words. Most subjects' shadowing productions shifted in the direction of the trainers, according to spectral measurements, while their post-training productions retreated slightly, though usually not to the point of the original productions, but some subjects continued to move in

the direction of the trainers after training was over. Female subjects generally displayed a greater degree of imitation and convergence than males. Interactions with gender motivate a second, ongoing study, in which this interaction is directly explored, using a similar training paradigm.

References

Bradlow, A., & Bent, T. 2008. Perceptual adaptation to non-native speech. *Cognition* 106, 707–729.

Clopper, C. G., & Pisoni, D. B. 2004. Effects of talker variability on perceptual learning of dialects. *Language and Speech* 47, 207–239.

Eisner, F., & McQueen, J. 2005. The specificity of perceptual learning in speech processing. *Perception & Psychophysics* 67, 224–238.

Goldinger, S. 1998. Echoes of echoes? An episodic theory of lexical access. *Psychological Review* 105, 251–279.

Pardo, J. 2006. On phonetic convergence during conversational interaction. *J. Acoust. Soc. Am.* 119, 2382–2393.

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.

Unmerging mergers-in-progress through spontaneous phonetic imitation

Molly Babel, Michael McAuliffe, and Graham Haber

Department of Linguistics, University of British Columbia
molly.babel@ubc.ca, mcauliff@interchange.ubc.ca, grahamth@interchange.ubc.ca

In this paper we examine whether mergers-in-progress can be unmerged in a spontaneous phonetic imitation paradigm. Spontaneous phonetic imitation is the unconscious process by which exposure to a speech stimulus causes a listener-turned-talker to display characteristics of the stimulus in their own productions (Babel, 2010, 2012; Goldinger, 1998; Namy et al., 2002; Nielsen, 2011; Shockley et al., 2004). Phonetic imitation has been argued to play a role in new dialect formation (Trudgill, 2004) and sound change (Delvaux & Soquet, 2007; Pardo, 2006), and is indicative of a bridge connecting perception and action in language (Pickering & Garrod, 2004). Mergers are a type of sound change which involve the elimination of contrast between two formerly contrastive phonemic distributions. Previous work has shown that vowels are spontaneous imitated (Babel, 2010, 2012), yet Evans and Iverson (2007) found that when adults with a COULD∼CUD merger were immersed in a new dialect that was not merged, the adults did not unmerge. In this paper we examine whether speakers reduce the degree of a merger when spontaneously imitating an unmerged model talker.

The merger under study involves the diphthongs in the lexical sets NEAR and SQUARE in New Zealand English. This merger-in-progress typically involves raising of the SQUARE diphthong such that in merged speakers, the vowel is realized as [i̟ə] (Hay et al., 2006) and approximates the NEAR vowel. New Zealand English-speaking participants were subjected to an auditory naming task where an Australian male served as the model talker. Australian English is not undergoing this merger, and the model talker's productions of these diphthongs were unmerged: NEAR /iə/ and SQUARE /eə/. Participants produced baseline tokens of 25 monophthongs, 9 NEAR words, and 9 SQUARE words, and then shadowed productions of the words from the AU model talker. The shadowing block was followed by a post-task in which, like the baseline, participants read the words aloud. Participants were equally divided between Positive and Negative Conditions. In the Positive Condition, participants were presented with a text which described the AU talker's positive feelings towards NZ. Those in the Negative Condition were presented with a text which described the AU model talker's negative attitude toward NZ. Participants were presented with their respective texts before the shadowing block thus after the baseline productions. At the end of participation, subjects completed an Implicit Association Task (IAT) to determine their bias towards New Zealand and Australia. Only the diphthongs are analyzed in this paper.

Imitation is measured perceptually through an AXB similarity task (Goldinger, 1998). Due to time limitations, each listener in the AXB task is presented with four shadowers (two from each condition); the task is blocked by shadower. The perceptual task is ongoing and currently 34 listeners have completed the task. In this task listeners are presented with shadowers' baseline tokens (A), the model talker production of the same word (X), and either a shadowed production or a post-task production (B). Each potential trial is presented twice with AXB and BXA orders. Shadowed and Post-Task productions are always compared to the same participant's baseline productions. Listeners' task is to determine whether the A or B token sounds more like X. A repeated-measures ANOVA with condition (Positive or Negative), block (Shadowed Token 1, Shadowed Token 2, or Post-Task), and diphthong category (NEAR or SQUARE) as within-listener variables and IAT as a between-listener variable demonstrated a main effect of diphthong [$F(1,32) = 11.47$, $p < 0.01$] and a three-way interaction of condition $\times$ block $\times$ diphthong [$F(2,66) = 6.06$, $p < 0.01$]. These results are shown in Figure **??**. More imitation was found for the SQUARE words, but this interacted with condition and block; participants in the Positive Condition imitated SQUARE words more in the shadowing task, while those in the Negative Condition exhibited more imitation in the Post-Task.

To assess how advanced the NEAR-SQUARE merger is in the NZ participants, a series of smoothing spline ANOVA (SSANOVA) models were constructed. SSANOVA models are used to assess whether curves, such as formant tracks through a vowel, are significantly different from one another (Davidson, 2006). For each subject, Bark-transformed F1 and F2 splines were constructed for their NEAR and SQUARE productions in each block. The Euclidean distance between the vowels F1 and F2 splines was fed into a between-subjects SSANOVA. In

this analysis condition was not a significant predictor, but IAT was. The results of the SSANOVA, which are assessed through visual inspection of 95% confidence intervals, can be seen in Figure **??**. Overall, speakers with a pro-Australian bias had less of a merger, and their final productions were even less merged. Speakers with a pro-New Zealand bias had more of merger, but their f-numberinal productions were less merged as well, though not to as large a degree. Interestingly, the first shadowed tokens were the opposite of the final productions, with speakers with a pro-Australian bias producing more merged vowels, and speakers with a pro-New Zealand bias producing less merged vowels.

Comparing the holistic AXB imitation measures and the SSANOVA measure of merger we find that condition assignment affects imitation when measured perceptually, while IAT bias interacts with degree of merger throughout the task using acoustic measurements. What this suggests is that listener judgments of imitation may be largely influenced by paralinguistic acoustic information, such as voice quality or timbre, and that these qualities are easily modified through positive or negative situational orientation to a talker. Whether an individual has a pro-NZ or pro-AU bias will affect their baseline degree of merger, and their propensity to become less merged (pro-AU) or more merged (pro-NZ) when exposed to an unmerged model. These results suggest (1) spontaneous phonetic imitation is influenced both by situational factors and pre-existing social preferences, but different aspects of the acoustic signal may be targeted differently in the process, and (2) mergers can become less merged in the context of a spontaneous imitation task.

Figure 1: Imitation as measured in the AXB task. The y-axis is the proportion of shadowed or post-task tokens judged more similar to the model talker's productions. The x-axis reports these results for the first and second shadowed tokens and the post-task productions. The NEAR∼SQUARE sets are plotted separately.

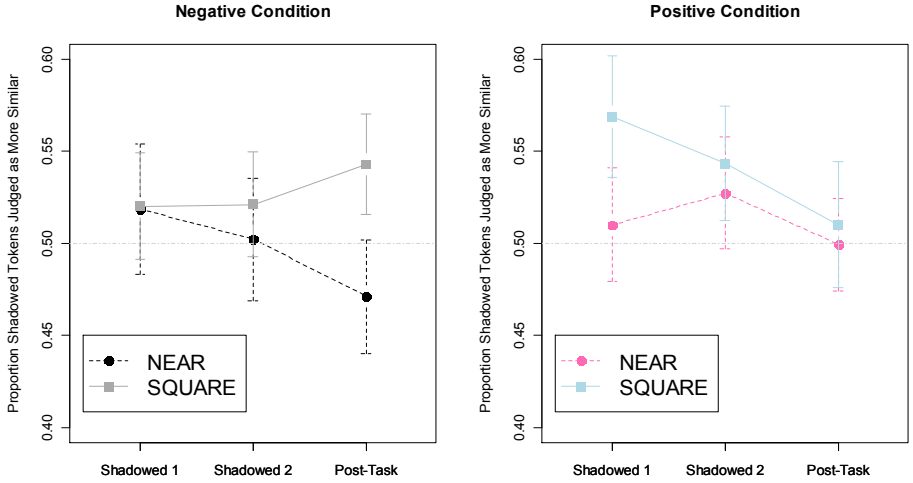
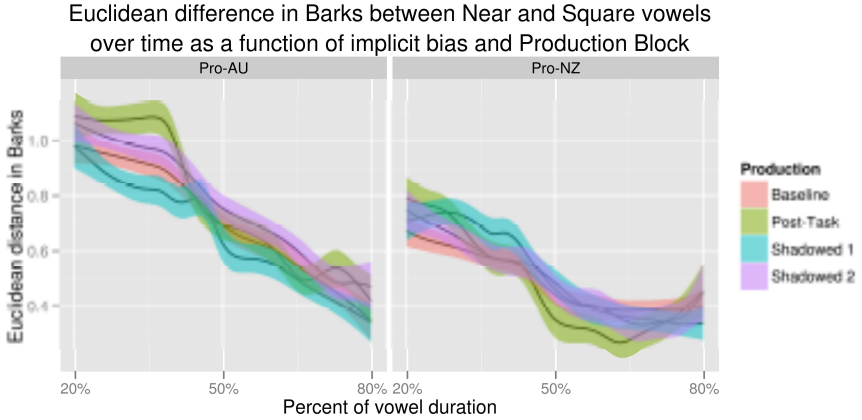


Figure 2: Degree of merger in speakers' vowels. The x-axis is percentage of a vowel's duration and the y-axis is the Euclidean distance in Bark between NEAR∼SQUARE vowels. All productions have a negative slope because both vowels end in schwa. Primary indications of degree of merger are in the first half of the vowel.



Effects of direct dialect imitation on tonal alignment in two Southern varieties of Italian

Mariapaola D'Imperio, Rossana Cavone, and Caterina Petrone

Laboratoire Parole et Langage, Aix-Marseille Université & CNRS
{mariapaola.dimperio,rossana.cavone,caterina.petrone}@lpl-aix.fr

Intonation, or the melody of speech, is a hallmark of language, and it is learnt even before segmental and lexical inventories are acquired. Within intonational features of speech, tonal alignment (cf. D'Imperio, 2006), i.e. the synchronization of tone targets and segments, appears to be systematically modified according to pitch accent category, and appears to be the hardest feature to be acquired/modified by non-native speakers (Mennen, 2004). The aim of this study is to test whether tonal alignment can be rapidly modified in order to imitate the intonational characteristics of a different regional variety of a language, specifically Italian.

Within the imitation paradigm, various studies have established that speakers are capable to modulate phonetic detail of their own speech in order to resemble that of speech to which they have been exposed (Goldinger, 1998; Nielsen, 2011). One hypothesis reflecting this phenomenon is that listeners would update their internal phonetic models in response to utterances heard. The updated phonetic models would be responsible for the observed imitation effect. Specifically, here we test if the alignment features of a Southern variety of Italian (Bari Italian) can be handled "online" and modified to look like those of a speaker of another Southern variety (Neapolitan Italian).

Both Neapolitan and Bari Italian show rising-falling configurations for expressing yes-no questions and narrow focus statements. The primary cue to interrogation in the Southern varieties is in fact a rising LH pitch accent (L+H* in Bari Italian and L*+H in Neapolitan, cf. Grice et al., 2005), immediately followed by a phrasal fall. Despite being very similar, the main difference between the two pitch accents is that the H peak is reached around the middle of the accented syllable in Bari Italian, while it is reached later (at the offset of the nuclear syllable) in Neapolitan (cf. D'Imperio, 2002). A similar alignment difference is found for narrow focus statements, which show a H+L* in Bari Italian and a L+H* in Neapolitan. While the H peak is aligned at the onset of the stressed syllable for Bari Italian, it is aligned around the middle of the stressed syllable in Neapolitan. In other words, while L+H*, with medial H alignment, signals a statement in Neapolitan, it signals a question in Bari.

The main hypothesis tested in this study is that tonal alignment can be rapidly modified by Bari Italian speakers. Specifically, we tested whether Bari speakers would produce later peaks for both H+L* (narrow focus statement) and L+H* (question pitch accent), in the process of imitating Neapolitan L+H* and L*+H. As an alternative hypothesis, we also tested whether Bari Italian speakers would produce higher (instead of later) H peaks for both H+L* (narrow focus statement) and L+H* (question pitch accent), as a substitute feature for peak delay (Gussenhoven, 2002). Moreover, if we also assume that phonetic convergence is selective, at least for segmental phonology (see Nielsen, 2011, for VOT imitation) in that it takes into account possible competition with the inventory of tonal language/dialect of origin, we expect that only the original question pitch accent (L+H*) would be affected in Bari Italian, since it would result in a novel pitch accent category (L*+H). This is because if later narrow focus statement accents are produced as a result of imitation, they would be in conflict with the existing L+H*, which is already assigned to questions.

Two groups of 10 native Bari Italian speakers participated in an imitation task, 10 males and 10 females. The total set of experimental materials consisted of 120 target words (10 low frequency and 10 high frequency - for 2 pitch accent plus 3 repetitions) plus 40 fillers, produced by a native Neapolitan Italian speaker. The recordings were made in a sound-attenuated room at the University of Bari. Subjects were recorded in 2 separate tasks. In the Baseline Task, participants read aloud the words randomly presented on a computer screen. In the Listening Task, participants were told that they would be listening to a recording of a speaker of another variety of Italian

and that they should try to imitate his/her pronunciation while repeating the word they heard (see also German et al., to appear; German, 2012). Explicit imitation instructions were preferred to a simple shadowing task with the aim of maximizing the effects, given that it appears that alignment features cannot be easily modified when learning a second language (Mennen, 2004). Subjects were not given any information about the variety that they were to imitate. In the data analysis phase, we compared measures of H peak alignment and scaling in the Baseline and the Listening tasks in order to test whether imitation had taken place. Since the data analysis is in progress, results will be reported at the workshop.

References

D'Imperio, M. 2006. Current Issues in Tonal Alignment. *Italian Journal of Linguistic* 18.

D'Imperio, M. 2002. Italian intonation: An overview and some questions. *Probus* 14, 37–69.

German, J. 2012. Dialect adaptation and two dimensions of tune. *Proceedings of Speech Prosody 2012*, Shanghai, China.

German, J., Carlson, K. and Pierrehumbert, J. to appear. Reassignment of the flap allophone in rapid dialect adaptation? *Journal of Phonetics*.

Goldinger, S. D. 1998. Echoes of echoes? An episodic theory of lexical access. *Psychological Review* 105, 251–279.

Grice, M., D'Imperio, M., Savino, M., & Avesani, C. 2005. Strategies for intonation labelling across varieties of Italian. In: Sun-Ah Jun (ed), *Prosodic Typology. The Phonology of Intonation and Phrasing*. Oxford: Oxford University Press, 362-389.

Gussenhoven, C. 2002. Intonation and interpretation: Phonetics and phonology. *Proceedings of Speech Prosody*, Aix-en-Provence, 47–57.

Mennen, I. 2004. Bi-directional interference in the intonation of Dutch speakers of Greek. *Journal of Phonetics* 32, 543–563.

Nielsen, K. 2011. Specificity and abstractness of VOT imitation. *Journal of Phonetics* 39, 132–142.

Pierrehumbert, J. B. 2003. Probabilistic phonology: Discrimination and robustness. In: Bod, R., Hay, J., & Jannedy, S. (eds), *Probabilistic Linguistics*. MIT Press, 177–228,.

Effects of imitative training techniques on L2 production and perception

Ewa Wanat, Rachel Smith, and Tamara Rathcke

University of Glasgow
ewa.wanat@hotmail.co.uk, rachel.smith@glasgow.ac.uk, tamara.rathcke@glasgow.ac.uk

Imitation of a native speaker has been claimed to improve both production and comprehension (Adank et al., 2010) of a second language (L2). A recent suggestion (Harrer, 1997) is that imitation in synchrony with a target speaker is particularly beneficial because it simultaneously engages the production and perception systems and provides immediate feedback on performance. We investigated how two types of imitation affected L2 learning of phonological contrasts in production and perception. Polish learners of English were tested on production and perception of English segmental contrasts before and after being exposed to a native speaker's production.

15 Polish learners of English, resident in Glasgow for between 1 to 6 years, were exposed to a set of 18 sentences spoken by a speaker of SSBE containing instances of two English features that are difficult for Polish learners: the contrast between tense /i/ and lax /ɪ/, and voicing of word-final, utterance-final consonants. Each sentence was put into a loop of 8 repetitions. The exposure task differed for the 3 groups of participants (5 per group): one group listened to the loops (listen-only or LO group), the second repeated each sentence after the target speaker (listen-and-repeat or LR group), and the third repeated synchronously with the target speaker (repetitive synchronous imitation or RSI group).

Before and after exposure, subjects did perception tests (a modified AXB task and an intelligibility-in-noise test) and read a set of control sentences containing novel words with the key features. These tasks allowed us to test which method was most successful in improving participants' perception accuracy and L2 pronunciation respectively.

The perception results showed an effect only for the LO group, who significantly improved in the postexposure perception task. This contrasts with Adank et al. (2010)'s finding that vocal imitation improves language comprehension. The post-exposure production test showed that both groups involved in production during exposure were closer to the target in the post-test than the LO group. Further, as far as the vowel duration is concerned, the RSI group accommodated better to the target speaker's production during the exposure task than the LR group, whereas the LR group were better than the RSI group at generalising vowel duration to the new (control) sentences. A significant result was also found for the exposure data, namely that the RSI group diverged less from the target speaker than the LR group on normalised F1 and F2 as well as duration values. The RSI group generalised the formant changes better in the post-test, i.e. showed better vowel quality learning. Taken together, these results suggest that a combination of both perception training through exposing the subjects to a native speaker's voice and production training through synchronising with the target speaker could be a successful method of teaching L2 perception and production.

References

Adank, P., Hagoort, P., & Bekkering, H. 2010. Imitation Improves Language Comprehension. *Psychological Science* 21, 1903–1909.

Harrer, G. 1997. Effectivarespr°akundervisning med nymetod. http://www.diu.se/nr1-97/nr1-97.asp?artikel=rsi (accessed on 14/03/2011).

Discursive convergence in conversation

Mathilde Guardiola and Roxane Bertrand

Laboratoire Parole et Langage, Aix-Marseille Université & CNRS
{mathilde.guardiola,roxane.bertrand}@lpl-aix.fr

Conversation is a joint activity (Clark, 1996). The success of interaction depends on the respect of principle of cooperation by the participants: "Make your contribution such as it is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged" (Grice, 1975). So during an interaction, and even more in a cooperative interaction, interactants' behavior is supposed to be cooperative. Backchannels, highly interactive phenomena, play a crucial role in the establishment of the common ground. They are indeed the explicit mark of the step by step elaboration of mutual knowledge shared among interlocutors. Otherwise, the range of the pragmatic function of backchannels can be wider: understanding, agreement, appreciation, assessment, passive recipiency, incipient speakership, etc. (Allwood et al., 1993).

Storytelling is a very frequent activity in conversation, and back-channel signals appear in some specific places in the narration. The main speaker arranges places for the listener to produce back-channels: they are used to show that the listener ratifies the other one as the narrator, or to show the evolution of shared knowledge (especially with clarification request).

The corpus we study is the Corpus of Interactional Data (Bertrand et al., 2008). We focus on a subset of 6 one-hour-long interactions. These interactions show single-sex pairs of interactants. In each interaction, participants had to tell personal stories: unusual stories for some of the interactions, professional conflict for the other ones. These consigns lead to many narrations in the interaction, interspersed by negotiation sequences. In storytelling, reported speech is a very frequent phenomenon. It particularly appears during orientation and complication phases (Labov, 2007) in the narration. Following Holt (1996), complaints and amusing stories are a privileged place for reported speech. This explains the high number of reported speech found in the corpus (nearly 500 for 12 speakers).

After annotating and analyzing direct reported speech in the corpus, we highlight a specific phenomenon which can be considered as a specific kind of complex back-channel signal (Laforest, 1992). Back-channel signals have been well described in studies about narration, among others. Our study focuses on responses revealing a stronger dimension than cooperation. These productions by the listener include specific and complex responses, such as completions, other-repetitions and comments. Among the completions, a more specific category appears in our corpus: direct reported speech "in echo", produced by the listener. The listener has so well understood the situation described that he is able to produce direct reported speech, whereas he was not present in the situation described. We do not consider these participations as simple back-channel signals. Sometimes they are very loud and long (with two characters' voices). They are designed as complete turns, they constitute a prosodic, syntactic and pragmatic unit. They are a specific response to narration that allows the listener to take punctually the place of main speaker and to produce reported speech instead of him/her. This kind of completion reveals co-narration (Bavelas et al., 2000). The utterance produced is sometimes repeated by the narrator, and so on becomes a part of the narration.

We compare the form and functions of this specific kind of direct sported speech to the classical types of direct reported speech (produced by the narrator). This direct reported speech "in echo" produced by the listener, appears during or after the production of reported speech by the main speaker, in the evaluation phase of the narration. These direct reported speeches "in echo" have an invention function: they clearly cannot have been heard by the speaker who tells them (Vincent & Dubois, 1997). They often do not have an introductory formula, but they can also build on the introductory formula contained in the main speaker's speech, who projected the direct reported speech. This leads to the simultaneous production of direct reported speech by the interactants.

In terms of convergence, this phenomenon leads to the question of adequation and alignment. Direct reported speech is an adequate response: the listener produces direct reported speech in a place where narration requires it. Since it is a phase of narration in which the narrator could (and often does) produce a direct reported speech, the speakers do the same discursive activity at the same time, and this similarity of discursive processes is an alignment at the level of forms. But reported speech "in echo" not only correspond to similarity of discursive processes, they also require a common basis (in terms of shared knowledge) to share the same representations. This sharing allows the emergence of co-development of overbidding humorous sequences, including joint fantasizing (Kotthoff, 2006).

Considering this, we assume that these reported speeches reveal a strong interactional alignment. These moments are highly convergent places in the interaction.

References

Allwood, J., Nivre, J., & Ahlsen, E. 1993. On the semantics and pragmatics of Linguistic Feedback. *Journal of Semantics* 9, 30.

Bavelas, J. B., Coates, L., & Johnson, T. 2000. Listeners as co-narrators. *Journal of Personality and Social Psychology* 79, 941–952.

Bertrand R., Blache P., Espesser R., Ferré G., Meunier C., Priego-Valverde B., & Rauzy S. 2008. Le CID - Corpus of Interactional Data – Annotation et Exploitation Multimodale de Parole Conversationnelle. *Traitement Automatique des Langues* 49, 105–134.

Clark, H. H. 1996. *Using Language*. Cambridge: Cambridge University Press.

Grice, H. P. 1975. Logic and conversation. In: Cole, P., & Morgan, J. L. (eds), *Syntax and Semantics: Vol.3*. New York: Academic Press.

Holt, E. 1996. Reporting on talk: the use of direct reported speech in conversation. *Research on Language and Social Interaction* 29, 219–245.

Kotthoff, H. 2006. Oral genres of humor: on the dialectic of genre knowledge and creative authoring. *Interaction and Linguistic Structures* 44.

Laforest, M. 1992. *Le back-channel en situation d'entrevue*. Québec: CIRAL, Université Laval.

Labov, W. 2007. Narrative pre-construction. In: Bamberg, M. (ed), *Narrative – State of the Art*. Amsterdam: John Benjamins, 47–56.

Vincent, D., & Dubois, S. 1997. *Le discours rapporté au quotidien*. Québec: Nuit Blanche éditeur.

Accommodation of backchannels in spontaneous speech

Antje Schweitzer and Natalie Lewandowski

Institute for Natural Language Processing, Stuttgart University, Germany
{antje.schweitzer,natalie.lewandowski}@ims.uni-stuttgart.de

We present first results from a project on phonetic convergence in spontaneous speech. Convergence is the process of accommodating one's style of speech to that of an interlocutor. Here, we examine 24 spontaneous conversations between female speakers on topics of their choice. Each dialog lasted approx. 25 minutes. Participants wore head-set microphones and could see each other through a transparent screen. There were 8 speakers, and each talked to 6 different interlocutors.

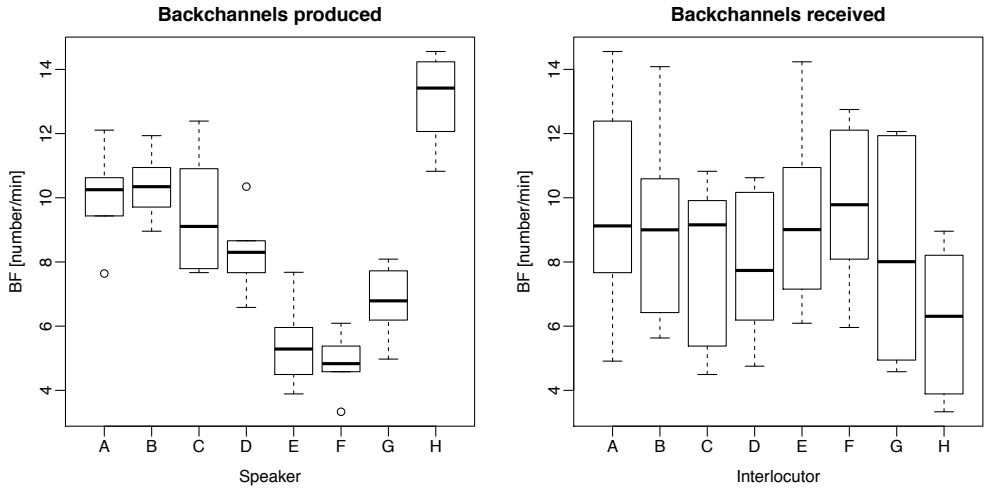
First analyses of turn-taking behavior were carried out using annotations which were generated completely automatically using Praat's (Boersma & Weenink, 2011) silence detection. We automatically identified backchannels (Yngve, 1970), i.e. short utterances produced by the listener which do not serve to interrupt the speaker's turn but serve as feedback for the speaker, such as in English "uh hum", "yeah", "o.k.", etc. We assume that each utterance of a speaker which is shorter than one second and which occurs in between utterances of the other speaker is a backchannel. We calculated backchannel frequency (BF) as the number of backchannels speakers produced in each dialog normalized by interlocutor vocalization duration. All statistical analyses presented here were conducted using R (R Development Core Team, 2011).

We were interested in speaker-specific as well as interlocutor-specific effects. Speaker-specific effects would suggest that speakers differ in their BFs. Interlocutor-specific effects, on the other hand, would indicate that speakers adjust BF depending on their interlocutor, i.e. they are accommodating (either converging to or diverging from) their interlocutor. We verified that BFs were approximately normally distributed and two Levene tests indicated no differences in variances between speakers or between interlocutors. We first ran a between-subjects analysis of variance with two factors without factor interaction (there is only one dialog for each combination of speaker and interlocutor). We found a significant speaker effect ($F(7,33)=38.2$, $p=0.0000$), indicating that BF is indeed highly speaker-specific. The interlocutor effect was also clearly significant ($F(7,33)=3.9$, $p=0.003$). This confirms that speakers differ in their BFs, and that they accommodate their BF depending on interlocutors. However, it does not indicate the direction of the effect—do they adjust their BF towards interlocutors (i.e., do they converge) or away from interlocutors (i.e., do they diverge)?

The left panel of Fig. 1 shows BF by speaker, i.e., each box represents the variability in a speaker's BFs across all her six conversations. The right panel indicates BF by interlocutor, i.e. each box represents the variability in BFs that interlocutors received. It is clearly visible that BF is speaker-specific while the differences between the BFs that interlocutors received are less pronounced: the (interlocutor-specific) ranges in the right-hand graph are less well-separated than the (speaker-specific) ranges in the left-hand graph. Still, there are differences even in the right-hand graph. As for the direction of the accommodation, examine, for instance, speaker H. She produced the highest BFs across her dialogs (Fig. 1, left, speaker=H). Interestingly, she received fairly low BFs from her interlocutors throughout (Fig. 1, right, interlocutor=H), i.e., they diverged. Similarly, speaker F produced the lowest BFs, but received the BFs with the highest median from her interlocutors. For other speakers, BFs produced and BFs received match better.

It is well accepted that the degree of accommodation (and its direction) is related to social factors (e.g. Giles & Smith, 1979; Street, 1984; Pardo et al., 2012). To cater for such social factors in the present database, speakers rated their conversational partner (in terms of likeability, competence, etc.) after each conversation. We can assess the correlation between these mutual ratings and the BFs by fitting a linear model with BF as dependent variable and the mutual ratings as predictors. We found that the more competent or likeable speakers rated their interlocutors, the higher the BFs they produced (competence: $t(46)=4.21$, $p=0.0001$, slope=0.71, adjusted $R2=0.26$; likeability: $t(46)=3.64$, $p=0.0007$, slope=0.61, adjusted $R2=0.21$). Interestingly, the symmetric effect was not present: speakers who produced higher BFs were not rated as more likeable ($t(46)=-0.62$, $p=0.54$) or

Figure 1: Backchannel frequencies (BFs) by speaker (left) and by interlocutor (right).



more competent (t(46)=-1.95, p=0.058, slope=-0.37) by their interlocutors. Quite the contrary, if we count this last effect as marginally significant, it shows the inverse correlation: the slope is negative. Thus, if anything, speakers who produced higher BFs were rated as less competent by their interlocutors. This (marginally significant) second finding is in accordance with results by Jurafsky et al. (2009) and Gravano et al. (2011). The first finding, the positive correlation between how likeable and competent speakers rate their interlocutors and the backchannel frequency that they produce is a new finding, at least to our knowledge. In any case, it clearly corroborates the assumption that social factors contribute to accounting for the degree of accommodation in conversations: competence, for instance, would explain approx. 26% of the variance observed in BFs, as can be inferred from R2 for the first regression model.

In conclusion, this first study shows that there are accommodation effects in this corpus, and that the social ratings collected do serve to capture aspects relevant for accommodation. In the future, we will look at many more parameters, especially more fine-grained ones. Also, we are interested in the dynamic aspects of convergence—we will try to assess how immediately the effects show up in conversations, and their scope.

References

Boersma, P., & Weenink, D. 2011. *Praat: Doing Phonetics by Computer* (version 5.2.26) [computer program]. Retrieved from http://www.praat.org.

Giles, H., & Smith, P. M. 1979. Accommodation theory: Optimal levels of convergence. *Language and Social Psychology*, 45–65.

Gravano, A., Levitan, R., Willson, L., Beňuš, Š., Hirschberg, J., & Nenkova, A. 2011. Acoustic and prosodic correlates of social behavior. *Proceedings of Interspeech 2011*, 97–100.

Jurafsky, D., Ranganath, R., & McFarland, D. 2009. Extracting social meaning: Identifying interactional style in spoken conversation. *Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the ACL*, 638–646.

Pardo, J. S., Gibbons, R., Suppes, A., & Krauss, R. M. 2012. Phonetic convergence in college roommates. *Journal of Phonetics* 40, 190–197.

R Development Core Team, 2011. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

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

Yngve, V. H. 1970. On getting a word in edgewise. *Papers from the Sixth Regional Meeting of the Chicago Linguistic Society*, 567–578.

Effects of language/variety interference and memory on direct imitation of German question intonation

Caterina Petrone, Leonardo Lancia, and Cristel Portes

caterina.petrone@lpl-aix.fr, leonardo_lancia@eva.mpg.de, cristel.portes@lpl-aix.fr

When adult speakers learn a foreign language (L2), the phonological system and the phonetic implementation of their native language (L1) may interfere in the perception and production of L2. At the segmental level, if a new sound is too similar to a sound in the phoneme inventory of L1, it will be perceived as an instance of that category. The acoustic features which may distinguish the new sound from canonical instances of the listeners' original phonological category will be neglected (e.g., Best, 1995; Tuller et al., 2008). Similarly, when L1 and L2 share the same intonational category, its exact phonetic realization in L2 may rely on phonetic implementation rules in L1 (Mennen, 2004).

The imitation paradigm has been used to get an insight into the learning process. Goldinger (1998) suggested that speakers are able to remember phonetic details of speech stimuli they have just heard and to rapidly shift their productions in order to reproduce such details. However, factors such as linguistic knowledge and memory might influence phonetic imitation. For instance, Nielsen (2011) suggested that VOT imitation is selective in that it is constrained by the phonological system of the language/language variety of origin. Moreover, phonetic imitation is less accurate in delayed shadowing (i.e. when speakers hear the stimulus but wait a few seconds before reproducing it) than in fast shadowing (i.e., when speakers repeat the stimulus quickly after its presentation). In fact, with the passing of time, the auditory details of a specific speech sound fade in the working memory (Cowan, 1984) and they are replaced by the properties defining the phonological category instantiated by that sound. Such a replacement can be achieved by matching the auditory stimuli with stored exemplars associated to that stimulus (Goldinger, 1998) or by silent rehearsal of the corresponding articulatory programs (Baddeley, 2002).

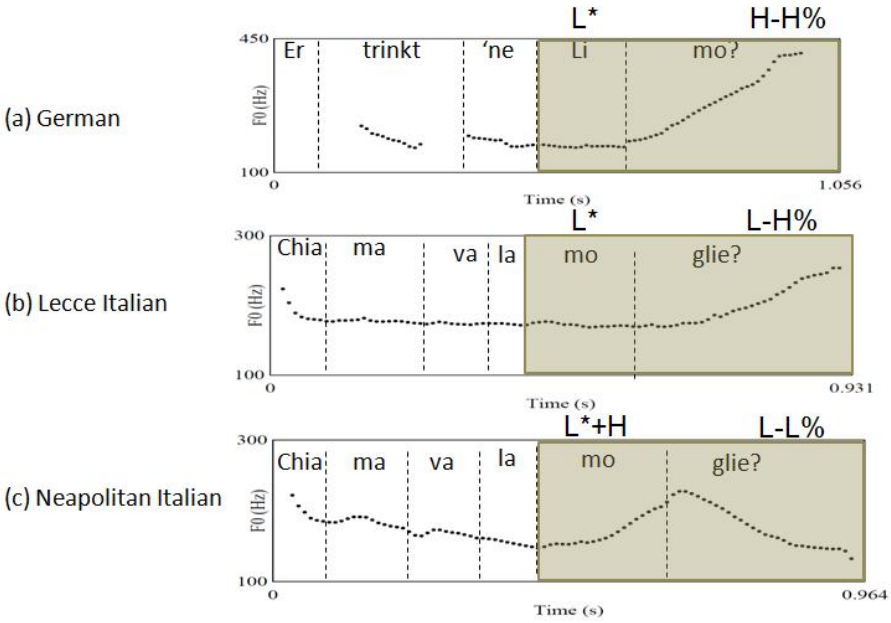
In this study we investigate how the knowledge of the language/variety of origin and memory interact in the imitation of intonation. We focus on the imitation of (Northern Standard) German yes/no questions by speakers of two Southern Italian varieties, Neapolitan and Lecce Italian. In German, questions are often marked by a terminal F0 rising configuration (L* H- H%, cf. Grice & Baumann, 2002), in which a low F0 valley in the accented syllable is immediately followed by a F0 rise (Fig. 1a). Particularly in the absence of syntactic cues (such as the subject-verb inversion), questions requires a high F0 rise which becomes steeper towards the end of the utterance. In Italian, yes/no questions are cued only by intonation but their specific pattern varies across regional varieties (cf. Savino, 2012). In Lecce Italian, they are realized with a low (or falling) accent followed by a terminal rise ((H+)L* L-H%, see Fig. 1b). Though this pattern is similar to the one found in German, the terminal rise in Lecce Italian is characterized by a later onset and a shallower slope. On the contrary, Neapolitan yes/no questions (Fig. 1c) show a rise-fall configuration, with a rising F0 movement starting from the accented syllable and followed by a terminal fall (L*+H L-L%).

In line with the segmental literature, we expect that, if imitation is selective, speakers of Neapolitan and Lecce Italian will take into account the phonological inventory and implementation rules of their own variety when imitating German questions. Specifically, Lecce speakers will be less accurate than Neapolitan speakers, since they will use the phonetic implementation rules for terminal rises of Lecce questions in reproducing German terminal rises. We also hypothesize that the reproduction of the phonetic details of German terminal rises will be worse when imitation is temporally delayed and/or when the possibility of rehearsal strategies is reduced. In fact, in these cases, imitation will rely on the abstract phonological representation which closely matches the terminal F0 rise rather than on the acoustic input itself. Ten Lecce Italian female speakers and ten Neapolitan Italian female speakers with no knowledge of German participated in a shadowing experiment. The lack of knowledge is aimed at maximizing the likelihood that speakers from the same language variety perceive the German stimuli globally without relying on previous imperfect knowledge and reproduce it phonetically. Two tasks were included. In the Baseline task, participants had to listen and reproduce questions uttered by a female speaker from their same variety (i.e. Lecce or Neapolitan Italian). In the Main task, the same subjects were

told that they would be listening to a recording of a speaker producing questions in a foreign language and that should try to imitate her pronunciation as accurately as possible. The instructions were aimed at enhancing the possibility of transfer from L1 to L2 question intonation. The corpus of the Main Task consisted of 8 short questions spoken by a female speaker from Northern German and repeated five times. To check for memory effects, participants had to start to speak only when a GO signal visually appeared on a computer screen. Three conditions were created. In the FAST condition, the GO stimulus appeared immediately after the end of the target question. In the DELAYED condition, the GO signal appeared after a silent pause of 4 s occurring at the end of the question. In the CONTEXT condition, the possibility of rehearsal is further reduced since the question was followed by an answer of 4 s, after which the participant was required to imitate only the given question.

The amount of convergence was assessed by comparing the F0 contours produced during the imitation task to the imitated utterances in a functional mixed models framework (Morris & Carrol, 2006). With this approach, it is possible to estimate the effect of the various experimental factors on the global shape differences between the F0 contours. The work is still in progress and results will be reported at the workshop.

Figure 1: F0 contour and phonological analysis of a yes/no question in German ("Does he drink a lemonade?", a), Lecce ("Did he call the wife?", b) and Neapolitan Italian ("Did he call the wife?", c). The relevant F0 contour at the end of the utterances is highlighted in grey.



References

Baddeley, A. 2002. Is working memory still working? *European Psychologist* 7, 85–97.

Best, C. T. 1995. A direct realist perspective on cross-language speech perception. In: Strange, W. (ed), *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues*. York: Timonium, 167–200.

Cowan, N. 1984. On short and long auditory stores *Psychological Bulletin* 96, 341–370.

Goldinger, S. D. 1998. Echoes of echoes? An episodic theory of lexical access. *Psychological Review* 105, 251–279.

Grice, M., & Baumann, S. 2002. Deutsche Intonation und GToBI. *Linguistische Berichte* 191, 267–298.

Mennen, I. 2004. Bi-directional interference in the intonation of Dutch speakers of Greek. *Journal of Phonetics* 32, 543–563.

Morris, J., & Carrol, R. 2006. Wavelet-based functional mixed models. *Journal of the Royal Statistical Society Series B* 68, 179–199.

Nielsen, K. 2011. Specificity and abstractness of VOT imitation. *Journal of Phonetics* 39, 132–142.

Savino, M. 2012. The intonation of polar questions in Italian: Where is the rise? *JIPA* 42, 23–48.

Tuller, B., Jantzen, M. G., & Jirsa, V. 2008. A dynamical approach to speech categorization: Two routes to learning. *New Ideas in Psychology* 26, 208–226.

Convergence, Complementarity, Co-ordination: Partner-specific effects in the emergence of procedural conventions

Gregory Mills

University of Edinburgh
gmills@staffmail.ed.ac.uk

One of the central findings in dialogue research is that interlocutors rapidly converge in their use of referring expressions. Studies on the emergence of referring conventions have demonstrated that this convergence is driven by the interaction: if interlocutors are able to provide each other with communicative, turn-by-turn feedback, this leads to the quicker development of representations that are more concise (Clark, 1996), more compositional (Garrod et al., 2007), more systematic and more abstract (Healey, 1997), and are also more tailored to specific conversational partners (Healey & Mills, 2006; Brown-Schmidt & Tanenhaus, 2008).
Although these studies underscore the importance of interaction, they differ in their accounts of how convergence is achieved. One of the important distinctions made by these studies is between two potentially different forms of convergence: (1) The global alignment that occurs over the course of the interaction (2) Turn-by-turn, immediate and local alignment, which involves an interlocutor (partially) repeating the communicative behaviour that their partner has just exhibited.
The interactive alignment model (Pickering & Garrod, 2004) proposes that global alignment of representations emerges as a direct and automatic consequence of turn-by-turn repetition (priming) that occurs at all levels of representation, both with- and between- speakers. From this perspective, high levels of local and global convergence are associated with communicative success.
By contrast, other studies argue that local turn-by-turn alignment of representations is best conceived as an interactive resource that is used strategically by interlocutors when encountering or anticipating problematic understanding (Healey, 1997; Saxton, 1997). Here, high levels of local convergence between interlocutors is seen as indicative of lower levels of successful communication.

Complementarity and procedural co-ordination

However, in addition to co-ordinating on the content of referring expressions, interaction in dialogue also requires procedural co-ordination: interlocutors must co-ordinate on the sequential and temporal unfolding of their contributions. Empirical studies of conversational interaction have demonstrated that procedural co-ordination is underpinned by interlocutors' use, not of the same, but of *different* kinds of contribution. For example, questions are usually followed with answers, not with another question, requests are usually followed with compliance, not with counter-requests, praise is usually followed with self-denigration, and offers with acceptance. These adjacency-pairs (Schegloff, 1992) are conventions which operate normatively, and consist of a first-pair part and a second-pair part, performed by different speakers. A central feature is that their successful use typically requires interlocutors to perform different and *complementary* contributions on subsequent turns. However, both conversation analytic and cognitive studies of interaction have treated these adjacency pairs as already shared and known to be shared by interlocutors, and do not study how interlocutors converge on them in the first place. It is also unclear whether convergence is driven primarily by egocentric processes (i.e. relatively low-level routinization), or whether interlocutors readily associate these conventions with specific conversational partners.

Alphabetical sorting task

To address these questions, we report a collaborative 3-participant task which presents participants with recurrent procedural co-ordination problems. Participants communicate via a text-based chat tool (Healey & Mills, 2006). Each participant's computer also displays a task window containing randomly generated words. Solving the task requires participants to combine their lists of words into a single alphabetically ordered list. To select a word, participants type the word preceded with "/". To ensure collaboration, participants can only select words

displayed on the other participant's screen and vice versa. Note that this task is trivial for an individual participant. However, for groups of participants, this task presents the co-ordination problem of interleaving their selections correctly: participants cannot select each other's words, words can't be selected twice, and words need to be selected in the correct order (see Mills, 2011, for a similar task).
To examine whether participants readily associate these conventions with specific conversational partners, the 3 participants were divided into a main dyad and a second side-participant. The task was configured such that at key moments in the development of the conventions, the side-participant is only required to observe the interaction, but does not directly participate in establishing the convention.
To test for partner-specific effects, we drew on the method of Healey & Mills (2006) of using a chat server to intercept and selectively manipulate participants' turns in real-time. This technique is used to generate artificial clarification requests that query the procedural function of participants' turns. The apparent origin of these clarification requests is manipulated to appear as if they originate from either of the 2 other participants (Main Dyad vs. Side participant).
Comparison of the responses to these two types of artificial clarification request allows direct testing of the hypothesis that interlocutors associate the co-ordination they achieve with specific conversational partners.

Results and Discussion

We demonstrate that participants' responses to these clarification requests provide strong evidence of partner-specific effects. Despite the clarification requests having exactly the same surface form (all that differs is the apparent origin), responses to both types of clarification are treated differently: Participants are slower to respond to clarification requests from the side-participants, their responses are also longer, contain more self-corrections, and they also subsequently make more mistakes in the task. We argue that focusing on procedural co-ordination suggests a more nuanced view of convergence in dialogue. The rapid development of conventions which consist of complementary contributions suggests that global convergence that occurs over the course of the interaction involves systematic divergence that occurs at a local turn-by-turn level. Drawing on participants' patterns of interaction in the task, we argue that this differentiation is indicative of a greater "forward momentum" in the interaction, as it indicates that participants have converged on what the next relevant step is in the dialogue. By contrast, high levels of local convergence between turns is indicative of lower levels of communicative success, as this typically indicates that interlocutors have halted the interaction in order to identify and resolve problematic understanding. We also argue that the finding of partner-specific effects also points towards differentiation and divergence occurring at more global levels of interaction – although all the participants are exposed to exactly the same communicative behaviour from each other (they all see the same interaction unfold on the screen), as they become more co-ordinated in the interaction, they systematically adopt different procedural conventions that become progressively complementary as their roles diverge .

References

Brown-Schmidt, S. & Tanenhaus, M. K. 2008. Real-time investigation of referential domains in unscripted conversation: A targeted language game approach. *Cognitive Science* 32, 643–684.

Clark, (1996). *Using Language*. Cambridge University Press.

Garrod, S., Fay, N., Lee, J., Oberlander, J., & MacLeod, T. 2007. Foundations of representation: Where might graphical symbol systems come from? *Cognitive Science* 31, 961–987.

Healey, P. G. T. 1997. Expertise or expert-ese: The emergence of task-oriented sub-languages. *Proceedings of the 19th Annual CogSci Meeting*, Stanford University, CA.

Healey, P. G. T., & Mills, G. 2006. Participation, precedence and co-ordination. *Proceedings of the 28th Conference of the Cognitive Science Society*, Canada.

Mills, G. J. 2011. The emergence of procedural conventions in dialogue. *Proceedings of the 33rd Annual CogSci Meeting*, Boston, USA.

Pickering, M. J., & Garrod, S. 2004. Toward a mechanistic psychology of dialogue. *Behavioural and Brain Sciences* 27, 169–226.

Saxton, M. 1997. The contrast theory of negative input. *Journal of Child Language* 24, 139–161.

Schegloff E. A. 1992. Repair after next turn. *AJS* 97(5).

Imitation, convergence, and phonological learning: A study of bilingual and monolingual patterns in the reproduction of word-final stop realizations in novel accents of English

Laura Spinu[1] *and Jiwon Hwang*[2]

[1]Classics, Modern Languages, and Linguistics - Concordia University Montreal, QC, Canada
[2]Department of English - College of Staten Island/CUNY Staten Island, NY, USA
lspinu@alcor.concordia.ca, Jiwon.Hwang@csi.cuny.edu

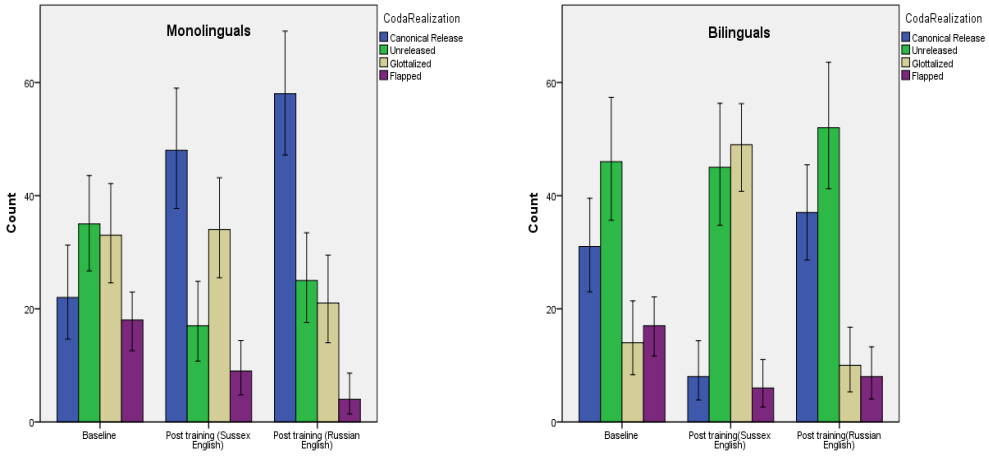
As part of a larger study investigating the acoustic correlates of accentedness and intelligibility in the reproduction of various accents of English by English monolinguals and French-English bilinguals from Canada, we explored speakers' ability to imitate and then spontaneously reproduce patterns of realization of word-final coronal stops in two different accents: SE England (Sussex) and Russian English. Glottalization of word-final coronal stops is a key feature of the Sussex dialect (e.g. the word *beat* is produced as [biʔ]), whereas in Russian English word-final stops are generally produced with an audible release (e.g. *beat* is produced as [bitˈ]). Other factors, among which the prosodic and segmental environment, also have an influence on the exact realization of these word-final phonemes. In the dialect samples to which our subjects were exposed, glottalization of final stops in Sussex English was very consistent (100% of the final coronal stops were realized as such), whereas the pattern of release of final stops in the Russian speaker's production, while still robust, was more variable.

17 monolingual and 12 bilingual speakers took part in the study. They were first recorded reading a set of sentences in their own English accent (baseline), and were then trained and tested on the two novel accents. The training and testing for each accent was done separately, with a short break in between. The training consisted of first listening to short sentences produced by a native speaker of that accent, following which a subset of these sentences were played again, and the subjects were prompted to imitate each sentence right after hearing it. The testing consisted of reading again the sentences that had been previously recorded in the subjects' native accent, this time trying to reproduce the accent they had been trained on to the best of their ability (and in the absence of any audio prompts). The training sentences and the testing sentences contained different target words, which had in common the environment for glottalization and full release of word-final stops, as is typical in Sussex English and Russian English, respectively. For example, subjects were presented with words such as *beat*, *bit*, and *bait* in the training, and words such as *heat*, *hit* and *hate* in the testing phase.

The analysis of the word-final coronal stops consisted of manual inspection and classification of each stop as (1) canonically released, (2) unreleased, (3) glottalized, and (4) flapped (following, in part, Sumner & Samuel, 2005). We hypothesized that, if any learning occurred, subjects' production of final coronal stops after training would be significantly different from their baseline. While our hypothesis was partially confirmed, interesting differences were noted between the monolingual and the bilingual group. For the monolingual group, no increase in glottalization occurred in response to the Sussex accent, however, the rate of release increased significantly in response to both the Sussex and the Russian accent. A decrease in glottalization was also noted with the latter. By contrast, the bilinguals exhibited a different pattern, with the rate of release being reduced and that of glottalization increased after training on the the Sussex accent, while no change as compared to baseline was noted for the Russian accent. The results are summarized in Figure 1.

While our results may appear to suggest that some initial learning of the new patterns has occurred with monolinguals for the Russian accent, but not for the Sussex accent, and vice versa with bilinguals, we note that monolinguals significantly increased their rate of release in reproducing the Russian accent *as well* as the Sussex accent. In light of this fact, we speculate that this is not a case of true learning, but rather one of hyperarticulation – when asked to reproduce different accents, monolinguals produced more careful speech, at least as far as this particular aspect is concerned (i.e. realization of coda coronal stops). By contrast, bilinguals really converged towards the Sussex accent, again, as far as the glottalization patterns are concerned. But if bilinguals are better at accent reproduction, the question arises as to why they did not increase the rate of release with the Russian accent. Going back to our initial observation that release was more variable in the sample of Russian English that was presented to our subjects, we hypothesize that they did not have enough evidence for this pattern, and thus did not modify their production from baseline.

Figure 1: Realization strategies for word-final coronal stops for monolinguals (right) and bilinguals (left) in three conditions: baseline, reproduction of the Sussex accent, and reproduction of the Russian accent.



To summarize, one of our main findings was that only the bilinguals significantly increased their rate of glottalization in response to the Sussex dialect of English, while the monolinguals did not pick up on this pattern. As for the Russian coda stop release pattern, none of the groups was able to learn it, a fact which we attribute to the presence of variability in the input. Instead, the monolinguals increased their release rates with both accents, which we interpret as a sign of hyperarticulation. We conclude that bilinguals are better at convergence as compared to monolinguals, thus adding to the body of work suggesting that bilinguals outperform monolinguals in certain types of linguistic tasks, such as manipulating discrete phonemic units (Bialystok et al., 2005; Bruck & Genesse, 1995) and the acquisition of novel words (Kaushanskaya & Marian, 2009). It remains to be determined whether longer exposure to the novel accents might yield the same results with monolinguals.

Turning to the type of training we employed, specifically the imitation task, our findings suggest that imitation of phonological patterns can facilitate phonological learning even after short-term exposure to a novel accent, at least in the case of bilinguals. Thus, we found that the behavior produced by imitation carried over into the post-imitative tasks, contrary to previous claims (Barry, 1989). We thus add some support to Markham's observation that acquisition itself is an imitative phenomenon (Markham, 1997) and, following Kuhl and Meltzoff, argue against the view that direct imitation bypasses all levels of linguistic processing (Kuhl & Meltzoff, 1995). We also contribute to the recent body of research suggesting that imitation of an action can result in improved understanding of that action (Adank et al., 2010).

References

Adank, P., Hagoort, P., & Bekkering, H. 2010. Imitation Improves Language Comprehension. *Psychological Science* 21, 1903–1909.

Barry, W. 1989. Perception and production of English vowels by German learners: Instrumental-phonetic support. *Phonetica* 46, 155–168.

Bialystok, E., Martin, M. M., & Viswanathan, M. 2005. Bilingualism across the lifespan, the rise and fall of inhibitory control. *International Journal of Bilingualism* 9, 103–119.

Bruck, M., & Genesse, F. 1995. Phonological awareness in young second language learners. *Journal of Child Language* 22, 307-324.

Kaushanskaya, M., & Marian, V. 2009. The bilingual advantage in novel word learning. *Psychonomic Bulletin and Review* 16, 705–710.

Kuhl, P., & Meltzoff, A. 1995. Vocal learning in infants: Development of perceptual-motor links for speech. *Proceedings of ICPhS 1995*, 146–149.

Markham, D. 1997. *Phonetic Imitation, Accent, and the Learner.* PhD Dissertation. Travaux de l'Institut de Linguistique de Lund 33. Lund University Press.

Sumner, M., & Samuel, A. G. 2005. Perception and representation of regular variation: The case of final /t/. *Journal of Memory and Language* 52, 322–338.

Accommodation and sociolinguistic meaning: Phonetic after-effects of *being* and *interacting with* a (dis)engaged interviewer

Kodi Weatherholtz, Abby Walker, and Kathryn Campbell-Kibler

kweatherholtz@ling.ohio-state.edu

Communication Accommodation Theory (CAT) (Giles & Powesland, 1975) explains linguistic convergence and divergence as strategies employed to minimise or maximise social distance. Studies have tried to investigate such socially-motivated shifting by putting naive participants into conditions that might evoke solidarity or distance between the participant and their interlocutor (Bourhis et al., 1979), or the participant and a pre-recorded speaker (Babel, 2010; Abrego-Collier et al., 2011). This paper reports a first analysis from a larger study on the relationship between conversational engagement and accommodation or other linguistic changes, at multiple levels. Data were collected in cross-dialect conversations in two social conditions, in which the interviewer tried to create or minimise distance. In this paper, we examine the impact of engaged and disengaged conditions on the vowel productions of both the interviewer and participants, and to explain the observed shifts we consider the role of convergence or divergence as well as sociolinguistic associations of the vowels themselves.

Participants were invited to participate in an interview-style experiment, during which they were audio and video recorded engaging successively with two different interviewers. The first interviewer— always an American who behaved in a friendly and engaged manner—talked with participants about high school social categories. The second interviewer—a New Zealander who alternated across participants between being a friendly, engaged interviewer and being a bored, disengaged interviewer—talked with participants about the college transition and college identity. When performing the disengaged role, the experimenter was careful not to be mean or nasty, but simply to signal lack of interest. No explicit choices regarding linguistic cues were made, instead the experimenter was free to alter speech, content and body language to display the appropriate stance.

Analysis of the audio and visual interview data is ongoing. Here we present an initial analysis of wordlist data from the New Zealand experimenter and 36 participants (13 M; 23 F). The interviewers and participants read a wordlist before and after each interview; thus, each participant read three lists, and each interviewer read two. All lists contained five low frequency CVC words from each of seven vowel classes, with different words comprising each list. The seven vowel classes were chosen to contain words that are known to differ between New Zealand English and Standard American English (LOT, DRESS, TRAP, BATH, NEAR) and words that are not (FLEECE, GOOSE). Participants did not hear the experimenters read any of their wordlists, but the experimenters did hear participants read a wordlist before they read their second list. However, the list they heard and the list they read usually differed. This critically means that any accommodation observed is unlikely the result of the (non)imitation of a particular lexeme, but rather a more systemic shift to the interlocutor's vowel space.

After the interview, the experimenters and participants rated their interlocutor on 8 dimensions, using a five-point scale. Factor analysis on participants' ratings of the interviewers revealed that these dimensions were highly correlated, with seven of the eight loading on a single factor indicating degree of liking or comfort. This combined measure was significantly greater in the engaged condition than the disengaged condition (4.53/3.73, t=2.62, p = 0.013). Factor analysis of the New Zealander's ratings of the participants motivated a measure of comfort, which correlated with condition (4.19/2.86, t = 6.72, $p < 0.001$) and a measure of participant engagement, which did not (4.66/4.49, t=1.70, p = 0.098).

Recordings were automatically segmented using the Penn Forced Aligner, and hand corrected in Praat. F1 and F2 were extracted from each vowel and Lobanov normalized. These normalized measures were fitted with mixed-effects linear regression models using the R package lme4. Random intercepts for participant and word were included, as well as random slopes of pre/post interacting with condition for word and pre/post for participant. Fixed effects examined were gender, regional accent, and the interaction of pre/post interview lists with condition and residualized measures of liking (for participant vowels) and comfort and participant distraction (for interviewer vowels).

Preliminary results suggest that the interviewer's and participants' vowel productions changed as a result of interaction and that those changes varied based on condition. However, the factors motivating these changes are open to multiple interpretations; for the interviewer, performative factors appear to play the largest role in post-interview speech shifts, while for the participants, complicated interactions on a subset of vowels may signal divergence, but may also reflect the broader socioindexical meaning of the particular variables.

The interviewer's speech shows a consistent pattern of reduced F1 space in non-high vowels: specifically, the vowel classes BATH, TRAP, DRESS and LOT (but not FLEECE and GOOSE) raise after the disengaged condition only. Given the consistency of raising across vowels, accommodation toward or away from the participant's productions is unlikely. Rather, we tentatively suggest that this represents a spillover effect from the effort of behaving in an uncharacteristically disengaged manner.

The participants show more restricted movement. The GOOSE vowel, fronting in many varieties of English, is more advanced among Midland than Northern US speakers, both of whom are represented in our study. The GOOSE of the New Zealand interviewer is even farther fronted than either of these group's mean. While relatively Midland accented speakers did not change in this vowel, we found that Northern male speakers, but not Northern females, showed significant backing of GOOSE after the disengaged condition only. While it is possible to interpret this as divergence from the interviewer's more fronted productions, the regional and gender differentiation suggests a role of local meaning, and/or linguistic experience or ability associated with this class.

Lastly, participants in the disengaged condition show movement in their TRAP vowels, but curiously, the direction of the shift is different for words that contain TRAP class and BATH class vowels in the New Zealander's dialect: American participants are raising TRAP and lowering BATH. The motivation for this split is unclear. An explanation based on distance in F1 relative to the New Zealander's wordlist productions of these vowels would point to divergence on TRAP and convergence on BATH. Alternatively, participants could be converging to the New Zealanders' split BATH-TRAP system. A further possibility is that this shift is not accommodation at all; some participants reported that the disengaged condition seemed very formal, and it is plausible that the observed TRAP-BATH splitting is a function of participants' speaking styles in this condition, particularly given that BATH is a highly marked, prestigious variable to American speakers (Boberg, 2009).

These results support an argument that interactional dynamics, including the degree of engagement of an interlocutor, have an impact on vowel productions in immediately following read speech. However, these shifts are not easily motivated by convergence or divergence strategies, or by automatic alignment due to priming during interaction (cf. Pickering & Garrod, 2004). While further insight will hopefully be gained through the (ongoing) analysis of the interview productions, the current data suggests that an accommodation-based analysis which does not also account for the sociolinguistic variation already present in a speaker's system may miss key factors.

References

Abrego-Collier, C., Grove, J., Sonderegger, M., & Yu, A. 2011. Effects of speaker evaluation on phonetic convergence. *Proceedings of ICPhS XVII*.

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

Boberg, C. 2009. The Emergence of a new phoneme: Foreign (a) in Canadian English. *Language Variation and Change* 21, 355–380.

Bourhis, R. Y., Giles, H., Leyens, J. P., & Tajfel, H. 1979. Psycholinguistic Distinctiveness: Language Divergence in Belgium. In: Giles, H., & St. Clair, R. N. (eds), *Language and Social Psychology*. Oxford: Basil Blackwell, 158–185.

Giles, H., & Powesland, P. F. 1975. *Speech Style and Social Evaluation.* London: Academic Press.

Pickering, M. J., & Garrod, S. 2004. Toward a mechanistic psychology of dialogue. *Behavioural and Brain Sciences* 27, 169–226.

Interpersonal long-term phonetic Accommodation-patterns in close acquaintances

Yshai Kalmanovitch

Department of German Linguistics, University of Zurich, Switzerland
yshai.kalmanovitch@sunrise.ch

Based on the general statement of accommodation theory (Giles et al., 1991) that speakers – under the relevant social and communicational conditions – are motivated to converge (to each other's speech) as well as on evidence from experiments showing speakers' ability and tendency to imitate recently perceived phonetic forms (e.g., Goldinger, 1998; Pardo, 2006), a study was conducted aimed at examining the effect of long-time acquaintance on phonetic accommodation patterns in interpersonal communication.

A group of four female native speakers of German from three different German-speaking regions, who have a longtime and intensive professional and social acquaintance with each other, was researched in a case study. Unlike Pardo et al. (2012), whose work is focused on a time-related intensification process of phonetic convergence between speakers, in the current experiment emphasis was laid on differences in accommodation-patterns between differing types of interpersonal interactions, with different levels of exposure of the observed speakers to each other's speech characteristics.

The four speakers were recorded under three conditions that simulated three types of speech-interaction: 1) two only slightly moderated group-interactions, with a time-interval of about one year between them, simulating long-term accommodation (**GI** and **GII** respectively), 2) a short personal interview – which was more of a friendly talk with the same speaker moderating the two group interactions and recorded at the same time as the second group-interaction, simulating short-term accommodation (**INT**). For conditions 1) and 2) no attempt was made to influence speech-production on the lexical level, and phonetic characteristics were examined. Finally, condition 3) consisted in the individual reading of a printed article on an unfamiliar subject in order to simulate zero-accommodation condition (for this last condition, labeled as **TXT**, an extra session a month after the second group-session took place, as text-reading parts in the first two recording-sessions were found quantitatively and qualitatively insufficient for the analysis).

A comprehensive analysis of some phonetic characteristics showed clear differences for each speaker between the three accommodation-conditions, which suggests that accommodation occurred, both in long-term as well as in short-term instances. The two group-interactions showed repeating accommodation patterns leading to objective phonetic convergence – in comparison with the zero-accommodation condition – with regard to intonation-range, voicing of Plosives and fricatives, VOT measures in accented syllables, fricative-fortition-level (tested on the basis of spectral characteristics), and vowel qualities. These observations supply clear evidence for accommodation-patterns that reflect a convergence effort targeted at the objective characteristics of the interacting participants.

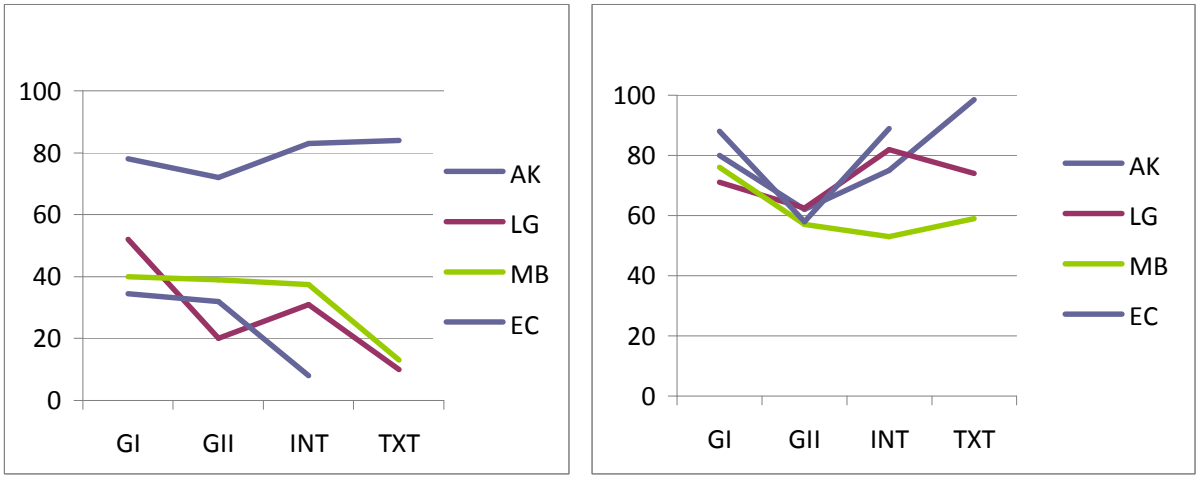
Although in the short-term accommodation condition all speakers were interviewed by the same speaker, no such regular accommodation-patterns as in the long-term accommodation could be observed, and speakers in this condition often showed mixed and contradicting tendencies. Such inconsistency between speakers as well as between observed phonetic features could not be explain by change of style or register alone, but would confirm with Coupland (1984)'s assumption that such patterns reflect the speaker's subjective conception of the hearer's speech-characteristics, rather than any perception of objective characteristics of the hearer (to whom the speaker would try to converge). This does not prevent "occasional success" – i.e., a high level of convergence – as indeed was observed in the current study.

The results confirm qualitative differences in accommodation patterns in the long-term and short-term accommodation of phonetic features in interpersonal communication. They also show that long and continuous exposure of the interacting participants to each-other's phonetic characteristics allow them to converge on the systemic level, and not only by imitating recently perceived phonetic forms.

Interestingly, no clear effect was observed for the time passing between the two recordings of the group-interactions on convergence level. Despite general changes in speech patterns noticed for all members of the group – such as a loss of the degree of voicing – the convergence level between the four speakers remained

more or less the same. This suggests that in long-term accommodation speakers do indeed react to each other's objective phonetic characteristic when converging – not in order to converge as much as possible, but only to a seemingly acceptable level.
The following diagram exemplifies some of the results observed describing the percentage of voiced realizations of standard High-German /z/ (left) and /g/ (EC was no longer available for TXT, hence the lack of data):

Figure 1



Voicing is a distinctive feature between plosives and fricatives in German standard pronunciation, which in most regions – with the exception of northern Germany, where the speaker AK comes from – would be replaced by a feature [±tense]. While features in TXT – especially with regard to the pronunciation of /z/ – confirm with regional varieties of the speakers observed, neither of the features in the other conditions show a general shift in one direction. In GI and GII the speakers converge towards what could be described as the groups "average" pronunciation rather than adopting one of the regional varieties or the standard pronunciation. A loss of voicing for all speakers can be observed in GII in comparison with GI, probably as a result of contact with the surrounding to which the four speakers arrived (Basel, Switzerland), which favors a voiceless realization of those consonants. This seems nevertheless to have no significant effect on the general convergence level between the four speakers in this condition. Similar patterns could be observed also in more purely phonetic features.

References

Coupland, N. 1984. Accommodation at work: Some phonological data and their implications. *International Journal of the Sociology of Language* 46, 49–70.

Giles, H., Coupland, N., & Coupland, J. (eds) 1991. *Contexts of Accommodation: Developments in Applied Sociolinguistics*. Cambridge: Cambridge University Press.

Goldinger, S. D. 1998. Echoes of echoes? An episodic theory of lexical access. *Psychological Review* 105, 251–279.

Pardo, J. S. 2006. On phonetic convergence during conversational interaction. *Journal of the Acoustical Society of America* 119, 2382–2393.

Pardo, J. S., Gibbons, R., Suppes, A., & Krauss, R. M. 2012. Phonetic convergence in college roommates. *Journal of Phonetics* 40, 190–197.

Developing sound categories in adult language learners' imitated and read speech

Terhi Peltola and Pertti Palo

Speech Sciences, Institute of Behavioural Sciences, University of Helsinki, Finland
terhi.peltola@helsinki.fi, pertti.palo@helsinki.fi

In this study we investigate the emergence of new foreign speech sound categories in read and imitated speech produced by language learners. According to some recent studies (Strange, 2011), language learners construct new vowel categories by gathering information in a statistical manner from the environment. Our aim is to investigate how this information emerges in speech.

Several recent studies (Fowler et al., 2003; Honorov et al., 2012) have proposed an important role for the distal events (Best, 1995) in speech perception and production. Rizzolatti & Craighero (2004) propose that the mirror neuron system is a fundamental part of the human capability of learning and understanding others. According to them, learning and understanding are closely linked together, because in order to understand the other we have to first learn the actions they perform by imitating the others. All this brings our attention back to the motor cortex and the motor theory of speech perception by Liberman et al. (1967, 1985).

The standard Finnish vowel inventory differs from the standard Hungarian inventory in several ways; category boundaries, orthographical conventions for some of the vowels and prototype locations have been proposed to differ (Winker et al., 1999). Especially the Finnish mid and open front vowels tend to cause problems to the Hungarian learners. As in Finnish there are three unrounded front vowels /i, e, æ/, all of which have long phonemically differing counterparts, in standard Hungarian there are only two short unrounded front vowels /i, ɛ/, which have long phonemic counterparts /iː, eː/. The Hungarian length opposition is not symmetrical while Finnish does have a symmetrical length opposition. Also the back vowel categories differ, but tend not to cause severe problems for Hungarians learning Finnish, since there are four back vowels in Hungarian and only three in Finnish. The orthographical conventions regarding the lower front vowels Finnish /e/s and Hungarian /ɛ, eː/ differ as well.

The participants for this study were chosen from a group of Hungarian university students who had enrolled in a three month conversation and Finnish phonetics rehearsal course for beginners. Four female students, with a mean age of 19.24, were chosen to participate because of their regular participation in the course. Two of them had some prior Finnish knowledge, but nevertheless they considered themselves beginners. They are all from the Budapest area, and have identical dialect backgrounds.

The course consisted of nine 45 minute lessons: a fifteen minute theoretical part and half-an-hour group rehearsal sessions mediated by native Finnish informants. The lessons were held in Finnish. Recordings took place during three additional single person rehearsal sessions also with a native Finnish informant. The three recordings were conducted before the start of the rehearsals, half way through them as well as after the last rehearsal. In the recordings participants read and imitated the same word list with minimal or subminimal pairs in a randomized order. All of the analyzed vowels were in the first syllables in two syllable words. We compare the quality of the vowels in both speech modalities at three time points. The analysis is based on formant frequencies and the results are illustrated in formant charts.

The results suggest that the new are emerging in spoken language. Qualities of both read and imitated /e, æ/ shifted towards the native Finnish vowel categories' acoustic qualities during the three month rehearsals. The data provides evidence, that the changes in read and imitated speech sounds are not identical.

References

Best, C. T. 1995. A direct realist perspective on cross-language speech perception. In: Strange, W. (ed), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Timonium, MD: York Press, 171–204.

Fowler, C. A., Brown, J. M., Sabadini, L., & Weihing, J. 2003. Rapid access to speech gestures in perception: Evidence from choice and simple response time tasks. *Journal of Memory and Language* 49, 396–413.

Honorov, D. N., Weihing, J., & Fowler, C. A. 2012. Articulatory events are imitated under rapid shadowing. *Journal of Phonetics* 39, 18–38.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. 1967. Perception of the speech code. *Psychological Review* 74, 431–461.

Liberman, A. M., & Mattingly, I. G. 1985. The motor theory of speech perception revised. *Cognition* 21, 1–36.

Rizzolatti, G., & Craighero, L. 2004. The mirror neuron system. *Annual Review of Neuroscience* 27, 169–192.

Strange, W. 2011. Automatic selective perception (ASP) of first and second language speech: A working model. *Journal of Phonetics* 39, 456–466.

Winker, I., Kujala, T., Tiitinen, H., Sivonen, P., Alku, P., Lehtokoski, A., Czigler, I., Csépe, V., Ilmoniemi, R. J., & Näätänen R. 1999. Brain responses reveal the learning of foreign language phonemes. *Psychophysiology* 36, 638–642.

Convergence in talk-in-interaction across languages: Multilingual format tying in peer talk-in-interaction

Gudrun Ziegler, Natalia Durus, & Neiloufar Family

gudrun.ziegler@web.de

In this study, we address format tying as an interactional feature, which occurs within turn-taking in interaction across speakers. This analysis of spontaneous talk in interaction highlights the importance of analyzing peer dialogue in language in general and in language learning in particular.

Studies in language development (Bruner, 1983) as well as in interaction in general (Selting & Couper-Kuhlen, 2000) highlight two main tasks that speakers need to manage when engaging in interaction.

Firstly, speakers need to coordinate their actions in line with prior actions and the joint activity in which they engage. Coordination therefore not only requires semantic or verbal convergence when managing the topic at talk (Hopper, 1987), but speakers also coordinate prosodic features (Wells & Corrin, 2004) and other elements when taking the next turn in talk. Resources for doing such coordination work are multiple and particularly challenging for young speakers and learners of a foreign language.

Secondly, speakers need to acknowledge another speaker's previous turn in order to make sure that the follow-up turn being produced continues the topic and/or action (cf. asking a question) (Auer, 1986). Repetition of previous elements or recycling of elements, including non-verbal elements is a key resource in doing such acknowledging of previous action activities in talk-in-interaction.

Both tasks, coordination of actions and doing acknowledging of a previous action, can be accomplished through various resources. This study focuses on format tying as a resource that has been mainly described in the context of adolescent talk in interaction (Goodwin, 2007; Goodwin & Goodwin, 1990). Goodwin (2007), in discussing format tying in adolescents talk-in-interaction suggests that "children build a new utterance by tying closely to prior talk, maintaining the grammatical structure of a prior sentence while making minimal semantic shifts". The current empirical study takes verbal repetitions, taking into consideration utterances which are modeled on the prior turn but which come either as tying actions (Example 1) or as tying candidate alternatives (Examples 2 and 3). In both cases, the participants in the interaction construct their subsequent turn within an adjacency pair and within the same sequence.

We discuss three occurrences of format tying as convergence in talk-in-interaction. In the following examples, format tying is constructed either as a side sequence (Example 1) or as side-talk (Examples 2 and 3). The orientation to Example 1 is closing the side-sequence by one of the participants, Rio (in line 14), passing to task-talk in Example 2 and both laughter and rejection in Example 3.

The examples were chosen because of their multilingual specificity. We show that in the three examples the use of the resources from various languages is not oriented to as the "trouble". The study aims at describing a collection of multilingual format tyings as these actions show how the interaction converges across the languages used.

The current article uses Conversation Analysis (Sacks et al., 1974) and a multimodal approach (e.g. Goodwin & Goodwin, 1990) to analyze peer classroom talk when students have plurilingual repertoires. The study shows how students draw on their plurilingual repertoires and challenge expected languages practices in the otherwise monolingual English classroom. The analysis is based on data from interE corpus (international English) collected in two settings: an international school in Germany (Example 1) and the European School in Luxembourg (Examples 2 and 3).

1)

001Tea:she always (.) says please.
002 ((3.4 writes on whiteboard))
003 Rio: (<garbled>what=re=you eating?>)
004 ((2.0 writes on whiteboard))
005 Flo: un ´bon=´bon
006Sta: ´mh=´mh
007 Flo: ´si ´si.

```
008Sta:     ´mh=´mh
009    Tea:    oh [dear.
010    Flo        [(bouffelaurent le fait exprès).
011    Rio:    how many [times (says).
012Sn  :             [apple?
013S    :    apple.
014    Rio:    alright yeah.
```

2)

```
001    Chris:    so<<acting voice>hello daddy>
002    Kosta:    (   )
003    Fritz:    <<acting voice>hello papa>
004    Kosta:    hallo papi
005    Bojan:    what do i write
006    Chris:    he[llo daddy
007    Fritz:      [m:::hello dada:
008    Chris:    hall=<<acting voice>hello daddy>
009    Pedro:    no=no=no=no
010    Chris:    yes:::
011    Bojan:    hello daddy
012    Chris:    daddy
013    Bojan:    <<french>common on ecrit da[dy>
014    Fritz:                               [no
015              write no noNO
016           you write hello papa beer
017bear_bear
018    Fritz:    hello papa beer
019              okay very good.
```

3)

```
001    Fritz:    so: is <<all> made of base of >
002    <<german>ma:rMOR (-)>
003              how do you say(.) mare=mar [mor
004    Pedro:                               [´oh(-)
005    i:n(-) e:nglish(-)
006    Chris:    <<pp>i don=t know>
007->Pedro:      <<german>marmor>
008    (---)
009->Bojan:      Ma:?rme:lade;
010    Pedro:    <<laugh>>ehhe[he
011    Bojan:    <<laugh>>ehehe[he
012    Fritz:                  [no <<german>marmor
```

References

Auer, P. 1986. Kontextualisierung. *StudiumLinguistik* 19, 22–47.

Bruner, J. 1983. *Child's Talk. Learning to Use Language*. Oxford: Oxford University Press.

Goodwin, C. 2007. Participation, stance, and affect in the organization of activities. *Discourse and Society* 18, 53–73.

Goodwin, C., & Goodwin. M. H. 1990. *He-said-she-said. Talk as Social Organization among Black Children.* Indiana: Indiana University Press.

Hayashi, M. 1999. Where grammar and interaction meet: A study of co-participant completion in Japanese conversation. *Human Studies* 22, 475–499.

Hopper, P. 1987. Emergent grammar. *Proceedings of the Annual Meeting of the Berkeley Linguistics Society* 13, 139–157.

Sacks, H., Schegloff, E. A., & Jefferson, G. 1974. A simplest systematics for the organization of turn-taking for conversation. *Language* 50, 696–735.

Selting, M., & Couper-Kuhlen, E. 2000. Argumente für die Entwicklungeiner "InteraktionalenLinguistik". *Gesprächsforschung – Online Zeitschriftzurverbalen Interaktion* 1, 76–95.

Wells, B., & Macfarlane, S. 1998. Prosody as an interactional resource: Turnprojection and overlap. *Language and Speech* 41, 265–294.

Wells, B., & Corrin, J. 2004. Prosodic resources, turn-taking and overlap in children's talk-in-interaction. In: Couper-Kuhlen, E., & Ford, C. E. (eds), *Sound Patterns in Interaction*. Amsterdam: Benjamins, 119–146.

Does the interpreter converge with the speaker? Analysing prosodic characteristics of simultaneously interpreted texts

George Christodoulides and Anne Catherine Simon

Université de Louvain, Centre VALIBEL Discours & Variation, Louvain-la-Neuve, Belgium
george@mycontent.gr, anne-catherine.simon@uclouvain.be

Our study focuses on convergence effects in the context of simultaneous conference interpreting. We research whether and to what extent the prosodic characteristics of an interpreter's speech are influenced by those of the speaker. We explore the extent of unidirectional convergence in the phonostyles of a speaker producing a text in the source language and an interpreter producing a text in the target language.

A person's phonostyle is determined both by the situational context and by individual characteristics (Llisteri, 1992; Eskenazi, 1993; Léon, 1993; Simon et al., 2010). The interpreter's phonostyle can be expected to reflect the situational context and the speaker's phonostyle, but also the cognitive process of interpreting. Simultaneous interpreting has been described as a taxing cognitive task, during which the interpreters are working at the limits of their processing capacity (the tightrope hypothesis; Gile, 2009). An interpreter may also deliberately choose to alter some of the local prosodic characteristics of their speech to mimic choices made by the speaker (Couper-Kuhlen, 1996), if they consider it necessary for expressive reasons. The interplay of these factors eventually creates the phonostyle of simultaneous interpreting. Professional interpreters may adopt different strategies in this respect: some will have a more uniform, personal style regardless of the speaker, while others will be more influenced by and converge with the speaker.

The aim of our study is to explore these questions based on a small bilingual spoken corpus. The value of using corpora in cross-linguistic research has been increasingly recognized (Granger, 2010). We have chosen to focus on one situational context, i.e. argumentative political discourse in EU institutions, to avoid possible variations due to different contexts. Our corpus consists of speeches produced in English and their interpreted versions into French. The corpus design ensures that for each speaker there are at least two different interpreters, and for each interpreter there are at least two different speakers.

Our methodology consists of building a parallel corpus and analyzing it using computational linguistics and speech analysis tools. We have developed software to manage the parallel corpus (Christodoulides, 2011) and used several tools for the automatic analysis of prosodic properties (Goldman et al., 2007, 2011). Our corpus is transcribed and aligned to the phone level, allowing us to extract temporal, melodic and accentual features of syllables. Among the prosodic features studied are the following: speech rate (including pauses) and articulation rate (excluding pauses); changes in speech rate (acceleration and deceleration); pause patterns and pause length distribution; prominent syllables, including their patterning and density; mean pitch and pitch range; mean intensity and intensity range; and melodic register, based on a model of fundamental frequency movements (measured in semitones per time unit).

The parallel nature of the corpus allows us to compare the evolution of each of these variables in the original speech and the interpreted version over time. We identify points of convergence or divergence (De Looze & Rauzy, 2011) by applying the Time Aligned Moving Average (TAMA) method (Kousidis et al., 2009). Furthermore, we have performed a bi-text alignment of our parallel corpus, to study the prosodic convergence over segments of equivalent content.

The results are interpreted according to two hypotheses. First, that the prosodic features of the interpreter's speech (e.g. speech rate, melodicity, number of prominent syllables etc.) are influenced by the speaker. And second, that the interpreter's phonostyle is less uniform than those of the speakers, i.e. the prosodic features of the interpreter's speech present a higher variance.

References

Christodoulides, G. 2011. *Praaline: a Tool for Managing Spoken Corpora and Speech Research*. http://www.mycontent.gr/praaline/

Couper-Kuhlen, E. 1996. The prosody of repetition: on quoting and mimicry. In: Couper-Kuhlen, E., & Selting, M. (eds), *Prosody in Conversation, Studies in Interactional Sociolinguistics*. Cambridge University Press, 366–405.

De Looze, C., & Rauzy S. 2011. Measuring speakers' similarity in speech by means of prosodic cues: Methods and potential. *Interspeech 2011*, 1393–1396.

Eskenazi, M. 1993. Trends in Speaking Styles Research. *ISCA*, 501–509.

Gile, D. 2009. *Basic Concepts and Models for Interpreter and Translator Training*. Revised edition. Amsterdam: John Benjamins.

Goldman, J.-Ph., Auchlin, A., Simon, A. C., & Avanzi, M. 2007. Phonostylographe: un outil de description prosodique. Comparaison du style radiophonique et lu. *Nouveaux Cahiers de Linguistique Française* 28, 219–237.

Goldman, J.-Ph., Auchlin, A., & Simon, A. C. 2011. Description prosodique semi-automatique et discrimination des styles de parole. *Actes d'IDP 2009*, 207–221.

Granger, S. 2010. Comparable and translation corpora in cross-linguistic research. Design, analysis and applications. *Journal of Shanghai Jiaotong University*.

Kousidis, S., Dorran, D., Mcdonnell, C., & Coyle, E. 2009. Convergence in human dialogues. Time Series Analysis of Acoustic Features. *Proceedings of SPECOM 2009*, 2.

Léon, P. 1993. *Précis de phonostylistique, Parole et expressivité*. Paris: Nathan.

Llisteri, J. 1992. Speaking Styles in Speech Research. *ELSNET/ESCA/SALT Workshop on Integrating Speech and Natural Language*, Dublin, Ireland, 28.

Pöchhacker, F. 2004. *Introducing Interpreting Studies*. London: Routledge.

Simon, A. C., Auchlin, A., Avanzi, M., & Goldman, J.-Ph. 2010. Les phonostyles: une description prosodique des styles de parole en français In: Abecassis, M., & Ledegen, G. (eds), *Les voix des Français. En parlant, en écrivant*. Berne: Peter Lang, 71–88.

Prosodic convergence and divergence: the building of coherence and shared meaning in conversational dialogues

Li-chiung Yang[1,2] *and Shu-Chuan Tseng*[2]

[1]Tunghai University, Taichung, Taiwan
[2]Institute of Linguistics, Academia Sinica, Taipei, Taiwan
yang_lc@thu.edu.tw, tsengsc@gate.sinica.edu.tw

Recent studies have shown convergent behavior in body movements and gesturing in conversation (Nagaoka et al., 2007; Gill, 2012), and in speech (Campbell & Scherer, 2010; Lelong & Bailly, 2011), and focused on their role in creating harmony and rapport between conversational participants through the use of feedback markers, and through timing and frequency of non-verbal facial and movement gesturing (Gratch et al., 2007).

Human language provides an ideal environment for studying the phenomenon of imitative and convergent behaviors in human communication. Spontaneous conversation is multifunctional in both its goals and processes: the most evident goal of transmitting information simultaneously carries a social goal of building rapport and the sharing of attitudes and emotions towards the information transmitted. In the conversational process, speakers provide propositional and emotional and information through prosody, gesturing, and feedback, and engage in interactional probing to build a shared knowledge state and guide topic in a mutually desired direction. Prosody plays a key role in this process, as it provides a powerful and informative resource to communicate multiple levels of coherence and meaning by providing a direct and immediate link to fundamental expressive states.

The current study presents our results on prosodic convergence and divergence in spoken dialogues, drawing from extended conversational data in Mandarin Chinese. Because of the multidimensional goals at work in language, synchrony is approached as both building social interactional harmony, and also reflecting informational, organizational and expressive processes in conversations. The coherence achieved in a successful dialog is a shared coherence, one that is constructed through interactions of participants to discover and overcome respective inadequacies of information state. Thus, in addition to imitative speech patterns, prosodic convergence and divergence are considered as information-rich patterns that speakers use to monitor comprehension, communicate disinterest or encouragement, and signal different levels of agreement and judgment on topic.

Our data corpora consist of two extended spontaneous conversations in Mandarin Chinese, each approximately one hour in length. The Mandarin data are a subset of Academia Sinica's Mandarin Conversational Dialogue Corpus (MCDC) of natural conversations between strangers. The conversations were segmented to the phrase level, and measures of fundamental frequency (f_0) and amplitude were automatically computed, and normalized to each speaker's pitch mean and range. For each speaker and each phrase, *low*, *average*, and *high* values for both f_0 and amplitude were extracted and calculated as a means to show the participants global pitch movement throughout the course of the conversation.

Our results show that both convergence and divergence in prosody occur at both local inter-phrase level pitch level changes, as well as over dialogue sections extending globally across topics and subtopics. The pattern found for our Mandarin conversational corpora is that prosodic convergence is arrived at gradually, with an initial probing stage where topic is negotiated, followed by mixed convergence and divergence as options are explored or overturned from a one-sided viewpoint, until speakers arrive at a mutually fulfilling topic theme, where convergence is frequent. Near conversation end, participants converge in a descending pitch pattern in a shared recognition of the coming conclusion.

By comparison to talks between friends, conversations between strangers may be more susceptible to lags in convergence, as speakers work to construct a common conversational outlook. The current results indicate that prosodic lags go in both directions, as speaker roles change and new topics are brought up. At the local level,

prosodic synchrony at phrase-to-phrase pitch movement is common: convergence is associated with agreement or encouragement of topic, divergence with disagreement, doubt, or non-interest. Speaker role was found to be important in the incidence and location of feedback tokens with respect to the prosodic patterns. Feedback markers of high interest or surprise such as "*oh*", and encouraging markers such as "*um*" or "*umhum*" occur more frequently in areas of high pitch and convergence, and less frequently in divergent prosodic sections. The marker "*dui*" *right* occurs more frequently in areas of convergence and stretches of extended rise as the hearer provides added encouragement or confirmation respectively. Thus, feedback markers often provide explicit marking of the same underlying relational states that are provided by synchrony phenomena.

Our analysis suggests that prosodic synchrony phenomena occur as a mirror of topically and emotionally synchronized or dis-synchronized participant states and that convergence and divergence phenomena are not only strategies to encourage rapport, but also act as organizational indicators providing key information on the degree of understanding, on emotional synchrony, and on the perceived status of a mutually fulfilling topic flow.

References

Campbell, N., & Scherer, S. 2010. Comparing measures of synchrony and alignment in dialogue speech timing with respect to turn-taking activity. *Proceedings of Interspeech 2010*.

Gill, S. P. 2012. Rhythmic synchrony and mediated interaction: towards a framework of rhythm in embodied interaction. *AI & Soc.* 27, 111–127.

Gratch, J., Wang, N., Gerten, J., Fast, E., & Duffy, R. 2007. Creating rapport with virtual agents. *Proceedings of the 7th International Conference on Intelligent Virtual Agents*. Paris, France: Lecture Notes in Computer Science, Springer, 125–138.

Lelong, A., & Bailly, G. 2011. Study of the phenomenon of phonetic convergence thanks to speech dominoes. In: Esposito, A., Vinciarelli, A., Vicsi, K., Pelachaud, C., & Nijholt, A. (eds), *Analysis of Verbal and Nonverbal Communication and Enactment: The Processing Issue*, 280–293.

Nagaoka, C., Komori, M., & Yoshikawa, S. 2007. Embodied synchrony in conversation. In: Toyoaki N. (ed), *Conversational Informatics: An Engineering Approach*: Wiley Series in Agent Technology, John Wiley & Sons, 331–352.

Mechanisms of speech adaptation

Maëva Garnier

GIPSA-Lab, Département Parole et Cognition, UMR CNRS 5216 & Grenoble Universités, France
maeva.garnier@gipsa-lab.grenoble-inp.fr

Different research areas can be considered around the topic of "speech adaptation".

Numerous studies have been conducted on how speakers react to a modification of their auditory feedback (attenuation, pitch shift, formant shift, delay) and compensate for a perturbation of their articulation ("bite block", lip tube, artificial palate, tongue piercing). Studies on phonetic convergence have provided evidences that speakers tend to unconsciously imitate some features of their interlocutor's speech. Many studies, in line with Lindblom's H&H theory, have also characterized how speakers modify their speech production in perturbed environments (noise, distance) or when they speak to someone with reduced comprehension (child, hard of hearing, non native listener). Other socio-phonetic studies have characterized how speech is adapted to the interlocutor, as a function of the affective or social relationship that we have with him/her/it (life partner, boss, pet, machine).

The mechanisms underlying these different types of adaptation are complex and their understanding still motivates on-going research. In particular, there is still a debate on whether these adaptations are driven by neural reflexes, by low-level mechanisms of regulation, or by higher-level mechanisms involving phonological and social representations. This presentation will summarize and discuss different arguments, from the literature as well as from my own studies, that support or infirm the involvement of these different mechanisms. Arguments come, in particular, from the observation of reaction times, from the neural networks involved, from the possibility or not to inhibit these adaptive reactions, from the existence of these adaptations in babies and animals, and from the persistence of these adaptive behaviours after exposure to the perturbed situation (after-effects). Some other arguments can be found in the inter-speaker variability of adaptation, related to different degrees of empathy or to varying perceptual acuities. Finally, the presentation will focus on communicational observations showing how adaptive behaviors can be influenced by the communicative interaction, by the degree of hierarchy and intimacy between the speech partners, by the phonological system, the acoustic environment and the sensory modalities of this interaction, and showing how these adaptations affect speech intelligibility and social interaction.

Auditory perception bias affects F0 imitation

Marie Postma and Eric Postma

Tilburg center for Cognition and Communication, Tilburg University, The Netherlands
m.nilsenova@uvt.nl, e.o.postma@uvt.nl

Phonetic imitation plays an important role in human interaction in that it reflects the closeness of the social bond between two individuals. Past studies have indicated the importance of the region between 50 Hz to 300 Hz (the fundamental frequency (F0) region) which is the most important source of information regarding emotions, stands and attitudes in the voice (Juslin & Laukka, 2003; Ververidis & Kotropoulos, 2006). The same region also provides acoustic information for imitation exploited in promoting social convergence and status accommodation (Gregory, 1983; Gregory & Hoyt, 1982; Gregory et al., 1993; Gregory & Webster, 1996; Gregory et al., 1997; Gregory & Gallagher, 2002) and expressing ingroup-outgroup bias (Babel, 2009). Interestingly, there appear to be large individual differences in speakers' ability to imitate F0. Using the shadowing task paradigm, originally introduced by Goldinger (1998), a recent study (Babel & Bulatov, 2011) found a considerable amount of variation in F0 accommodation, with some subjects actually diverging from the F0 of the model talker.

We hypothesized that the individual differences in speakers' ability to imitate F0 may at least partly be due to their neurocognitive ability to extract information about pitch from the speech signal. In particular, due to neuroanatomical differences found in the lateral Heschl's gyrus (the "pitch processing center"), some listeners show an auditory perception bias for the sound as a whole (*fundamental* listeners), while others (*spectral* listeners) focus on its harmonic constituents (Rousseau et al., 1996; Schneider et al., 2005a). The auditory perception bias has been almost exclusively analyzed in the context of musical training, but the results of individual studies indicate that it may also affect linguistic performance (Wong & Perrachione, 2007; Wong et al., 2008). This study is the first attempt to explore the role of perception bias in imitation and thus its possible impact on social convergence.

Participants' auditory perception bias was determined with a variation of the psychoacoustic perceptual test described in Smoorenburg (1970), Laguitton et al. (1998) and Schneider et al. (2005b). Participants were asked to categorize 18 perceptually ambiguous stimuli consisting of two complex tones, A and B, that were composed of a number of upper harmonic tones with the same highest harmonic but different levels of virtual fundamental pitch (derived from the harmonics as the best fit) and spectral pitch (based on the lowest harmonic). The other 18 stimuli served as control trials in that their interpretation is unambiguous but helps to determine a participant's level of attention to the task. In order to test the validity of the perceptual test, we repeated the measurement approximately one month later under the same conditions with a subset of the participant set (N=64). The majority of our participants performed as fundamental listeners. A comparison of the first and the second measurement showed that even without feedback, repeated exposure to the ambiguous stimuli results in a shift towards fundamental auditory bias. Interestingly, a similar training-independent increase in the salience of the virtual fundamental pitch has been earlier reported by Seither-Preisler et al. (2009), who ascribed it to learning-induced long-term plasticity reflecting the biological relevance of pitch sensation. In particular, a listener's ability to perceive the missing F0 plays an important role in sound perception in that it helps to track prosodic contours in speech even when they are masked by noise or not transmitted (Seither-Preisler et al., 2007), as in phone speech where the region up to 300 Hz is missing.

We subsequently collected speech data in a shadowing task with two conditions, one with a full speech signal and one with high-pass filtered speech above 300 Hz. The material used during the shadowing task consisted of 16 sentences (8 declaratives and 8 interrogatives) that were presented four times in different orders to each participant. During the first and the fourth presentation, the sentences were shown one-by-one on a computer screen and the participant was instructed to read them in a neutral manner. During the second and the third presentation, the material was played through a Sennheiser HMD26-600 headset and the participant was asked to repeat the sentences as precisely as possible. For model talkers, we used four different Dutch speakers (two male and two female) who were selected from a set of ten candidates on the grounds of speech clarity and lack of regional accent. The model talkers pre-recorded the 16 sentences in a soundproof booth with a Sennheiser HMD26-600 mic headset. Per participant, we used the recordings of a single model talker in order to increase exposure to the model talker's pitch. The participants were randomly divided between two experimental con-

ditions; half of the participants heard full speech recordings while the other half heard recordings that were filtered with a 300-Hz high-pass Butterworth filter implemented in the Signal Processing Toolbox in Matlab. We calculated the "Degree of F0 Imitation" by subtracting the absolute difference between the second and third F0 measurement (where the participant was shadowing) from the absolute difference between the model talker's F0 and the participant's F0 in the first measurement (baseline). Thus, a positive value of the "Degree of F0 Imitation" indicates that the participant adapted to the model talker's F0, a negative value means the participant diverged and 0 represents no measurable change in mean F0. The F0 measurements were analyzed with multiple regression with the between-subject condition F0 Filter (full speech signal vs. signal with frequencies above 300 Hz) and the participant's Perception Bias as predictors and the Degree of F0 Imitation as the dependent variable. The results show that more fundamental listeners are better in F0 imitation than less fundamental (spectral) listeners, especially in conditions where the F0 information is missing and needs to be derived from the speech signal. This suggests advantages for fundamental listeners in communicative situations where F0 imitation is used as a behavioral cue. Future research needs to determine to what extent auditory perception bias may be influenced by training and whether it affects other social processes that rely on parsing of the prosodic information.

References

Babel, M. 2009. *Phonetic and social selectivity in speech accommodation*. Ph.D. diss., Univ. California, Berkeley.

Babel, M., & Bulatov, D. 2011. The role of fundamental frequency in phonetic accommodation. *Lang. Speech* 1–17.

Beerends, J. G. 1989. *Pitches of Simultaneous Complex Tones*. Unpublished Ph.D. diss, University of Eindhoven.

Goldinger, S. D. 1998. Echoes of echoes? An episodic theory of lexical access. *Psychological Review* 105, 251–279.

Gregory, S. W. Jr. 1983. A quantitative analysis of temporal symmetry in microsocial relations. *American Sociological Review* 48, 129–135.

Gregory, S. W. Jr., & Hoyt, B. R. 1982. Conversation partner mutual adaptation as demonstrated by Fourier series analysis. *Journal of Psycholinguistic Research* 11, 35–46.

Gregory, S. W. Jr., Webster, S. W., & Huang, G. 1993. Voice pitch and amplitude convergence as a metric of quality in dyadic interviews. *Language and Communication* 13, 195–217.

Gregory, S. W. Jr., & Webster, S. W. 1996. A nonverbal signal in voices of interview partners effectively predicts communication accommodation and social status perceptions. *J. Personality and Social Psychology* 70, 1231–1240.

Gregory, S. W. Jr., Dagan, K., & Webster, S. W. 1997. Evaluating the relation of vocal accommodation in conversation partners' fundamental frequencies to perceptions of communication quality. *Journal of Nonverbal Behaviour* 21, 23–43.

Gregory, S. W. Jr., & Gallagher, T. J. 2002. Spectral analysis of candidates' nonverbal vocal communication: Predicting U.S. presidential election outcomes. *Social Psychological Quarterly* 65, 298–308.

Houtsma, A. J. M. 1979. Musical pitch of two-tone complexes and predictions by modern pitch theories. *J. Acoust. Soc. Am.* 66, 87–99.

Juslin, P. N., & Laukka, P. 2003. Communication of emotions in vocal expression and musical performance: different channels, same code? *Psychological Bulletin* 129, 770–814.

Laguitton, V., Demany, L., Semal, C., & Liégeois-Chauvel, C. 1998. Pitch perception: A difference between right- and left-handed listeners. *Neuropsychologia* 36, 201–207.

Rousseau, L., et al. 1996. Spectral and virtual pitch perception of complex tones: An opposite hemispheric lateralization? *Brain and Cognition* 30, 303–308.

Schneider, P., et al. 2005a. Structural and functional asymmetry of lateral Heschl's gyrus reflects pitch perception preference. *Nature Neuroscience* 8, 1241–1247.

Schneider, P., et al. 2005b. Structural, functional and perceptual differences in Heschl's gyrus and musical instrument preference. *Ann. N.Y. Acad. Sci.* 1060, 387–394.

Seither-Preisler, A., et al. 2007. Tone sequences with conflicting fundamental pitch and timbre changes are heard differently by musicians and non-musicians. *J. Exp. Psychol.: Hum. Percept. Perform.* 33, 743–751.

Seither-Preisler, A., et al. 2009. The perception of dual-aspect tone sequences changes with stimulus exposure. *Brain Research Journal* 2, 125–148.

Smoorenburg, G. F. 1970. Pitch perception of two-frequency stimuli. *J. Acoust. Soc. Am.* 48, 924–942.

Ververidis, D., & Kotropoulos, C. 2006. Emotional speech recognition: Resources, features, and methods. *Speech Communication* 48, 1162–1181.

Wong, P. C. M., & Perrachione, T. K. 2007. Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics* 28, 565–585.

Wong, P. C. M., et al. 2008. Volume of left Heschl's gyrus and linguistic pitch learning. *Cerebral Cortex* 18, 828–836.

Effects of seeing and hearing vowels on neonatal facial imitation

Marion Coulon, Cherhazad Hemimou, and Arlette Streri

Laboratoire Psychologie de la Perception, UMR 8158, CNRS, Université Paris Descartes, Paris, France
coulon.marion@gmail.com, cherhazad@hotmail.fr, arlette.streri@gmail.com

1. Introduction

For several decades, many authors have claimed the existence early in life of a strong link between speech perceptual and productive systems (e.g., Legerstee, 1990; Meltzoff & Moore, 1983, 1997). However, as the majority of these studies has been conducted when infants already have had some experience with language at both perceptual and motor levels, the question of whether this link is present at birth or acquired by experience remained open. The present study addressed this issue by employing the paradigm of neonatal facial imitation. We compared imitative responses of newborn infants presented either visual-only, audiovisual congruent or audiovisual incongruent models. If newborns' productive responses are modulated according to the model's presentation modality conditions, this would support the hypothesis of an innate link between speech perception and production.

2. Method

The stimuli were color audiovisual sequences of a woman's face. We created video clips that agreed with the classical protocols used by previous studies investigating neonatal imitation (e.g., Meltzoff & Moore, 1977). Newborns were presented two vowels: /a/ (as in "chat" in French) and /i/ (as in "lit" in French). Each 7-minute video clip began by presenting a passive face (i.e., the baseline period) followed by the "presentation block" of the first vowel that alternated between productions sequences and passive face sequences. The second vowel was then presented similarly. The video clip was used as it was for the audiovisual congruent condition. The sound track was edited out for the visual-only condition. For the audiovisual incongruent condition, the video was dubbed: The model's mouth openings were accompanied by the sound /i/ and the model's lip spreadings were accompanied by the sound /a/. In each condition, we counterbalanced the order of vowels in the clips (/a/-/i/ or /i/-/a/).
36 newborns participated in this study (12 per condition). Infants were placed in a seat designed to support a neonate. They sat facing a 19-inch color monitor, placed 35 cm away from their eyes. A camera, fixed above the screen, recorded their behaviors. All the video recordings were coded by two observers, by using "The Observer" software. Both coders, blind to the condition and the presentation order, scored the mouth openings (MO) and the lip spreadings (LS). The infants' gaze toward the screen was also coded.

3. Results

3.1 Visual-only versus audiovisual congruent conditions

According to Meltzoff & Moore (1977), the newborns' responses must fulfill two main conditions to be defined as real imitations: The rate of emission of the observed behavior must be significantly higher during the corresponding modelling interval than (1) the spontaneous rate of emission during the baseline and (2) the rate of this response during the presentation of another modelled behavior. In our experiment, MO and LS imitations were validated for both conditions (two-tailed t tests, $p < .01$). Our results also revealed that, contrary to our expectations, the newborns' imitations were not modulated in terms of frequency according to the modality condition presentation (ANOVA, ns). However, our newborn subjects performed significantly more MO than LS imitations for both conditions (two-tailed t tests, $p < .01$). Finally, both MO and LS reaction times were significantly shorter in the audiovisual congruent condition than in the visual-only one (two-tailed t tests, $p < .05$).

3.2 Audiovisual incongruent condition

First, contrary to the visual-only and the audiovisual congruent conditions, neither MO nor LS responses in the audiovisual incongruent condition could be validated as real imitations (two-tailed t tests, ns). The newborns produced globally less MO and LS in the incongruent condition than in the audiovisual congruent and the visual-only ones (ANOVA, $p < .01$). Moreover, the newborns looked significantly less at the video in the audiovisual incongruent condition than in the visual-only and the audiovisual congruent ones (ANOVA, $p < .01$).

4. Discussion

Our study reveals four important findings: First, under some conditions, 2D models seen on a video can elicit real facial imitations by newborns. Our results thus agree with previous reports concerning live models (e.g., Meltzoff & Moore, 1997; Vinter, 1986) and extend them to a new experimental setting. Second, our results document neonatal imitation of lip spreading for the first time. Imitation of this new gesture had been predicted but never tested previously (Meltzoff & Moore, 1997). This study therefore extends the range of observed imitated facial gestures by newborn infants. Third, our results provide evidence that neonatal imitation could be modulated according to the modality condition of the model's presentation: Imitations appeared significantly more quickly when the model was audiovisual congruent than when it was only visual. Moreover, we observed that a mismatch between the auditory stimuli and the mouth movements inhibited the newborns' production of motor matching. To summarize, our findings, by highlighting the influence of speech perception on newborns' imitative responses, evidence the strong link between perceptual and productive systems at birth. They also suggest that newborns already possess some knowledge concerning the auditory-visual-motor correspondences of speech. Several authors suggested previously a link between neonatal imitation and speech development (e.g., Serkhane et al., 2005). Chen et al. (2004) even proposed that this early motor skill could be "a precursor behaviour to more mature forms of vocal imitation and language production in general". According to us, our findings offer strong evidence for such a link between neonatal imitation and language.

References

Chen, X., Striano, T., & Rakoczy, H. 2004. Auditory-oral matching behaviour in newborns. *Developmental Science* 7, 42–47.

Legerstee, M. 1990. Infants use multimodal information to imitate speech sounds. *Infant Behavior and Development* 13, 343–354.

Meltzoff, A. N., & Moore, M. K. 1977. Imitation of facial and manual gestures by newborn infants. *Science* 198, 74.

Meltzoff, A. N., & Moore, M. K. 1983. Newborn infants imitate facial gestures. *Child Development* 54, 702–709.

Meltzoff, A. N., & Moore, M. K. 1997. Explaining facial imitation: A theoretical model. *Infant and Child Development* 6, 179–192.

Vinter, A. 1986. The role of movement in eliciting early imitations. *Child Development* 57, 66–71.

Serkhane, J., Schwartz, J. L., & Bessière, P. 2005. Building a talking baby robot, a contribution to the study of speech acquisition and evolution. *Interactions Studies* 6, 253–286.

Articulations or Preconceptions? An Investigation of Visual Speech Alignment Findings

Kauyumari Sanchez

New Zealand Institute of Language, Brain and Behaviour
mari.sanchez.77@gmail.com

When conversing, people have the tendency to shift their speech toward the speech of their conversational partner. This phenomenon is referred to as speech accommodation, convergence, and also alignment. In addition to natural conversational settings, speech alignment has also been found in socially devoid laboratory settings, often in shadowing contexts (e.g. Goldinger, 1998; Shockley et al., 2004). Shadowing experiments in speech alignment typically occur in the following way. In the baseline phase, participants are recorded reading a list of words from a computer monitor out-loud. In the shadowing phase, participants are recorded as they engage in a shadowing task where they are asked to listen to the speech of a model and say the words they hear out-loud. Finally, a different set of participants rate the perceptual similarity of the baseline and shadowed words to the words uttered by model. More often than not, these raters select the shadowed words as more similar to the model's words, compared to the baseline words, suggesting that the participants shifted their speech toward the model during the shadowing task.

Recently, Miller et al. (2010) found evidence for visual speech alignment in a shadowing experiment. In the shadowing phase, instead of auditory stimuli, participants were presented with a face to lip-read. To facilitate accurate lip-reading, participants were presented with two words (e.g. "Tennis" "Turkey") on the screen prior to viewing the articulating face uttering one of these words (e.g. "Turkey"). Participants were instructed to say the word they lip-read out-loud and were recorded. Here too, raters judged the shadowed words as more similar to the model's words, compared to the baseline words, suggesting that the participants shifted their speech toward the model during the shadowing task.

Thus, speech alignment in the direction of a model, has been found when shadowing words perceived auditorily (e.g. Goldinger, 1998; Shockley et al., 2004) and visually, via lip-reading (e.g. Miller, Sanchez & Rosenblum, 2010). Following Miller, Sanchez & Rosenblum (2010), Sanchez (2011) examined whether multiple shadowers align to a specific model in the same ways or uniquely, and whether the modality perceived affects this similarity. Perceptual raters, performing a matching task, judged the utterances of multiple shadowers of the same model as being more similar than those of shadowers of another model, regardless of whether the model's speech was shadowed auditorily- or visually-only (via lip-reading). These results suggest that shadowers align to similar properties of a specific model's speech even when doing so based on different modalities. However, when reviewing the visual condition, this study surprisingly found that the model these shadowers saw resulted in different rates of perceived similarity between the multiple shadowers of a particular model; whereas, no such difference was observed when reviewing the auditory-only condition. Thus, it seems that the face of the model may have elicited these differences.

It is possible, for visual speech alignment, that people are not necessarily picking up on visual talker-specific characteristics (e.g. articulations) of the model, so much as activating a preconception of how the model *should* sound. Evidence in support of a preconception account can be found in the social priming literature. For example, contextual cues in the environment can lead to shifts in speech perception. It has been found that the same auditory speech can be perceived differently, such as more Kiwi or Australian, if the words "New Zealand" or "Australia" are present on the subject's answer sheet (Hay et al., 2006a). In a related study, it was also found that the same auditory speech can be perceived as Kiwi or Australian speech if items related to these counties (e.g. a stuffed a kiwi bird or koala) are present in the experiment room. In addition, it has been found that various social aspects can shift one's speech perceptions and productions.

Hay et al. (2006b) found that when presenting the same auditory speech accompanied by different photographs of the "talker" (e.g. "older", "younger", "middle class" and "working class" photos), that not only did the

perception of the presented speech shift in line with these social categories, but the subject's speech productions also shifted. Thus, it may be the case that alignment to visual speech may be driven by the activation of socially relevant information which can alter one's speech productions.

The current investigation aims to identify whether alignment to visual speech is based on talker articulations or preconceptions activated upon seeing the model's face. If visual speech alignment is to *preconceptions* rather than articulations, then shadowers should align to a talker whose face they see even when articulations are not perceived (e.g. a still image of a model's face or the top portion, nose and above, of a dynamic face). However, if alignment is to *articulations*, rather than preconceptions, then shadowers should align equally, if not better, when only articulations (e.g. bottom portion of an articulating face) are presented verses when the entire face and accompanying articulations are presented (e.g. full articulating face).

Two sets of subjects were used in this experiment, shadowers and raters. Shadowing subjects were first recorded uttering words in a baseline task. For the shadowing phase, this experiment implements a 4 Presentation Condition (still image, dynamic top face, dynamic bottom face, dynamic full face) $\times$ 2 Model (model 1 or model 2) $\times$ 2 Word Frequency (high or low) $\times$ 2 Syllables (mono- or bi-syllabic) mixed subjects design. Presentation condition and model were between subject factors for the shadowers while word frequency and syllables were within subject factors. Perceptual raters were asked to judge the relative similarity of the baseline and shadowed words to the model's words. Alignment was determined as occurring if the shadowed utterance was selected as more similar to the model's utterance than the baseline utterance.

Initial evidence shows trends for alignment in the bottom articulating face condition and surprisingly, the top articulating condition. Theoretical ramifications relating to speech alignment and social priming will be discussed.

References

Hay, J., & Drager, K. 2010. Stuffed toys and speech perception. *Linguistics* 48, 865–892.

Hay, J., Nolan, A., & Drager, K. 2006a. From fush to feesh: Exemplar priming in speech perception. *The Linguistic Review* 23, 351–379.

Hay, J., Warren, P., & Drager, K. 2006. Factors influencing speech perception in the context of a merger-in-process. *Journal of Phonetics* 34, 458–484.

Goldinger, S. D. 1998. Echoes of echoes? An episodic theory of lexical access. *Psychological Review* 105, 251–79.

Miller, R. M., Sanchez, K., & Rosenblum, L. D. 2010. Alignment to visual speech information. *Attention, Perception, & Psychophysics* 72, 1614–1625.

Sanchez, K. 2011. *Do you Hear What I See? The Voice and Face of a Talker Similarly Influence the Speech of Multiple Listeners*. Doctoral Dissertation, University of California, Riverside.

Shockley, K., Sabadini, L., & Fowler, C. A. 2004. Imitation in shadowing words. *Perception & Psychophysics* 66, 422–429.

Breathing changes during listening and subsequent speech according to the speaker and the loudness level

Amélie Rochet-Capellan[1]*, Susanne Fuchs*[1]*, Leonardo Lancia*[2]*, and Pascal Perrier*[3]

[1]ZAS, Berlin, Germany, [2] MPI, Leipzig, Germany, [3] GIPSA-lab/DPC Grenoble, France
ameliecapellan@free.fr

Introduction

In the context of the interactive alignment model for multi-level adaptations in dialogue situations (Pickering & Garrod, 2004), our work focuses on the adaptation of listener's breathing to a speaker's breathing. Our general goal is to understand if, and how, listeners adapt their breathing to speakers' breathing: what are the information that the listener's breathing system is "catching" from the speakers' behavior to eventually change accordingly? Previous works showed that, in dialogue situation, listeners and speakers tend to synchronize their breathing at the time of turn-taking (Guaïtella, 1993; McFarland, 2001). Moreover, perception studies found that when breathing noise (e.g. when speakers inhale) is added to speech synthesis, the listeners' recall performance increases (Whalen et al., 1995). From the acoustical signal, listeners are also able to discriminate between speech produced starting at high lung volume versus speech started at a low lung volume level (Milstein & Watson, 2004). Breathing profiles during listening are also different than during breathing at rest, and could be an indicator of the perceptual process. In this context, Brown (1962) hypothesised that poor listeners may be less capable to adapt their breathing to the speaker's breathing when compared to better listeners. Recently, Stephens et al. (2010) pre-recorded speakers' acoustical productions and the co-occurring brain activity. Then, they monitored listeners' brain activity during the audio playback of the speakers' productions. They found some speaker-listener neural coupling, even if the speaker was not present. Using a situation analogous to Stephens et al. (2010), the present study investigates if listeners' breathing changes according to the speaker they listen to (male vs. female) and to the loudness level (normal vs. loud) of the speaker's voice. We also evaluate if breathing during speech produced right after listening differs according to the speaker and to the loudness level of the signal heard during the listening task.

Methods

Our protocol was comparable to the one developed by Stephens et al. (2010). We pre-recorded acoustic and breathing movements produced by two *readers*, one male (23 yrs, 1.86 m, 65 kg) and one female (35 yrs, 1.70 m, 58 kg), while they were reading short texts (fables) with a normal and then a loud volume level. We played back these audio recordings to *listeners* (26 females, average age 25$\pm$3 (std) and average Body Mass Index 21.5$\pm$2). The listeners heard either the male or the female reader, and 5 texts in normal speech first and in loud speech second, or in the reverse order. Listeners were instructed to listen attentively to the story and to briefly summarize it afterwards. Readers and listeners were all native speakers of German. Acoustical and breathing signals were recorded for readers and listeners in the same conditions. Breathing movements were recorded for the thorax and the abdomen using Respitrace. We expected different breathing profiles for the two readers, due to their different morphologies, and variations in breathing profiles according to the level of loudness (Binazzi et al., 2006; Huber et al., 2005). These differences between readers should have some echo in listeners' breaths during the listening and the summary task. The effects of readers, loudness and condition order on the amplitude and duration of the breathing cycle were tested using Linear Mixed Model.

Results and conclusion

As expected, the two readers differ in the two conditions of loudness, particularly with respect to breathing frequency (female: normal $<$ loud, male: normal $>$ loud) and with respect to duration of exhalation (longer for loud vs. normal for the male, no diff. for the female). Both readers were generally similar with respect to amplitude of inhalation: loud speech goes hand in hand with deeper inhalation than normal speech. During listening, listeners tended to inhale more frequently and shorter when listening to loud speech as compared to normal speech. This tendency was observed for 18/26 listeners and did neither depend on the reader, nor on the condition order. However, a three level interaction showed that the decrease of cycle duration, when listening to

loud speech as compared to normal speech, could be greater when loud speech was heard before normal speech, especially for the male speaker. The amplitude of the breathing cycle was globally larger in listening to normal as compared to listening to loud speech. The effect of loudness on amplitude was dependent on the reader. It was more prominent for the listeners to the male reader (11/13 subjects) than for the listeners to the female reader (4/13 subjects) who even tended to show the reverse pattern. The effect of the loudness condition on the amplitude of inhalation during listening also showed an effect of condition order and was mainly observed when subjects listened to the loud condition first. This difference in amplitude due to the condition order was observed only for the listeners to the female reader. When listeners spoke to summarize the texts right after the listening task, the asymmetry between inhalation and exhalation strokes of the breathing cycle increased as compared to the shape of inhalation and exhalation during listening. These changes in breathing pattern between listening to speech and speech production are similar to previous observations (McFarland, 2001). However, we did not find any significant effect of the reader, nor of the loudness level on the shape of the listeners' breathing cycle during the summary task. This could be due the fact that a large variability was observed between subjects in the way they achieve the summary task (e.g. number of breathing cycles).

These preliminary results show that listeners' breathing is sensitive to the reader and to the loudness of the reader's speech. This sensitivity could be a physiological reaction, as breathing is closely linked with heartbeats and emotional state. It could also be linked with the fact that the cognitive load could be greater for louder speech as compared to normal speech. Finally, changes in listeners' breathing could result from an adaptation to specific characteristics of the reader's voice and/or rhythms or to a speaker-listener's coupling, as it has been observed for body movements in dialogue or for brain activity during listening (Schmidt et al., 2011; Shockley et al., 2009; Stephens et al., 2010). We are now using spectral methods to evaluate if some synchronization between listener's and reader's breathing could be observed.

References

Binazzi, B., Lanini, B., Bianchi, R., Romagnoli, I., Nerini, M., Gigliotti, F., Duranti, R., Milic-Emili, J., & Scano, G. 2006. Breathing pattern and kinematics in normal subjects during speech, singing and loud whispering. *Acta Physiologica* 186, 233–246.

Brown, C. T. 1962. Introductory study of breathing as an index of listening. *Speech Monographs* 29, 79–83.

Guaïtella, I. 1993. Étude expérimentale de la respiration en dialogue spontané. *Folia Phoniatrica* 45, 273–279.

Huber, J. E., Chandrasekaran, B., & Wolstencroft, J. J. 2005. Changes to respiratory mechanisms during speech as a result of different cues to increase loudness. *Journal of Applied Physiology* 98, 2177–2184.

McFarland, D. H. 2001. Respiratory markers of conversational interaction. *Journal of Speech Language and Hearing Research* 44, 128–143.

Milstein, C. F., & Watson, P. J. 2004. The effects of lung volume initiation on speech: a perceptual study. *Journal of Voice* 18, 38–45.

Pickering, M. J., & Garrod, S. 2004. Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences* 27, 169–190.

Schmidt, R. C., Fitzpatrick, P., Caron, R., & Mergeche, J. 2011. Understanding social motor coordination. *Human Movement Science* 30, 834–845.

Shockley, K., Richardson, D. C., & Dale, R. 2009. Conversation and coordinative structures. *Topics in Cognitive Science* 1, 305–319.

Stephens, G. J., Silbert, L. J., & Hasson, U. 2010. Speaker-listener neural coupling underlies successful communication. *Proceedings of the National Academy of Sciences* 107, 14425–14430.

Whalen, D. H., Hoequist, C. E., & Sheffert, S. M. 1995. The effects of breath sounds on the perception of synthetic speech. *Journal of Acoustical Society of America* 97, 3147–3153.

Mechanisms for interactive alignment during conversation

Simon Garrod[1] *and Martin Pickering*[2]

[1]University of Glasgow, [2]University of Edinburgh
simon@psy.gla.ac.uk

During conversation interlocutors align their linguistic and conceptual representations at various levels. In our original account, Pickering & Garrod (2004) explained this alignment process in terms of cross-modal priming between speakers and listeners. Here we consider an additional alignment mechanism arising from interweaving of language production and comprehension processes within each interlocutor. The argument is based on Pickering and Garrod (in press), who start with the observation that in production, comprehension and dialogue, as in action, action perception and joint action more generally, generative and perceptual processes are intimately interwoven. And that this interweaving of the two supports prediction of what you are about to do or what your partner is about to do. Specifically, we argue that actors construct forward models of their actions before they execute those actions, and that perceivers of others' actions covertly imitate those actions, then construct forward models of those actions. We use these accounts of action, action perception, and joint action to develop accounts of production, comprehension, and interactive language. Importantly, they incorporate well-defined levels of linguistic representation (such as semantics, syntax, and phonology). We show (a) how speakers and comprehenders use covert imitation and forward modeling to make predictions at these representation levels, (b) how they interweave production and comprehension processes, and (c) how they use these predictions to monitor the upcoming utterances. We argue that this account explains a range of behavioral and neuroscientific data and represents a general framework for interactive communication.

References

Pickering, M. J., & Garrod, S. 2004. Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences* 27, 169–190.

Vocal Imitation Positively Affects Language Attitudes

Patti Adank, Andrew J. Stewart, and Louise Connell

School of Psychological Sciences, University of Manchester, Manchester, United Kingdom
Patti.Adank@manchester.ac.uk

Observing an action leads to activation of motor representations required to reproduce that action (Fadiga et al., 2002) as well as an automatic imitative motor response (Brass et al., 2000). Furthermore, observation of the person performing the action leads to stereotype priming (Chen & Bargh, 1997), which may affect the observer's subsequent behaviour (Dijksterhuis, 2005; Iacoboni, 2009). Action perception thus leads to motor imitation and to behaviour congruent with primed stereotypes. But how do both types of imitation affect action perception? Adank et al. (2010) demonstrated that imitating sentences in an unfamiliar accent improves subsequent comprehension of this accent. Motor imitation of perceived actions thus optimises action understanding (Pickering & Garrod, 2007). Nevertheless, it is unclear whether the optimising effect of imitation extends to perceived stereotypes.

We used regionally accented speech to test whether imitation affects stereotype perception associated with speakers of a regional accent. Listening to an accent automatically invokes stereotypes (or attitudes) (Lambert et al., 1960). For instance, speakers of standard accents are perceived as more powerful and competent – but as having less social attractiveness – than speakers of regional accents (Giles & Billings, 2004). Moreover, people commonly imitate each other's regional accent (Delvaux & Soquet, 2007). Participants first listened to sentences spoken in a regional accent of British English, different from their own, namely Glaswegian English (GE). They repeated half the sentences in their own accent, or they imitated the accent of the speaker for the other half. After each repeating/imitating session, participants completed a questionnaire (Bayard et al., 2001) probing perceived power, competence, and pleasantness.

Method

We tested 52 native speakers from England (32 female) who were unfamiliar with Scottish accents. Stimulus materials were 96 sentences spoken by two male GE speakers (cf. Adank et al., 2009, for stimulus details). Participants repeated 48 sentences from GE speaker 1 and subsequently imitated 48 sentences from GE speaker 2. Task Order (repeat or imitate first) and Speaker Imitated (which GE speaker was imitated, 1 or 2) were counterbalanced across participants. After each repeating and imitation session, participants rated 18 personality and voice traits, using a questionnaire (adapted from Bayard et al., 2001, cf. Supplementary Materials. Ratings were as follows: 1: speaker conforms very much, 4: speaker does not conform). Six were classified as Power traits (controlling, authoritative, dominant powerful voice, strong voice, assertive), six as Competence traits (reliable, intelligent, competent, hard working, educated voice, ambitious), and six as Pleasantness traits (Bayard et al.'s Solidarity and Voice factors pooled: cheerful, friendly, warm humorous, attractive voice, pleasant voice).

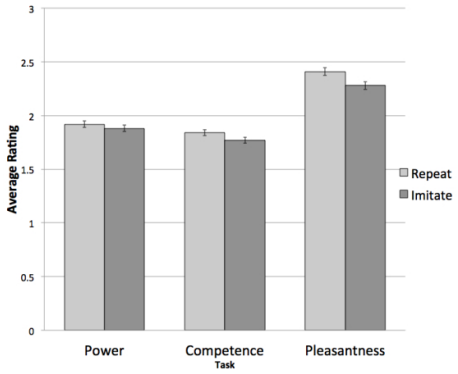
Results

A 2 (Task) $\times$ 3 (Attitude) analysis of variance was conducted on average rating scores with Order (repeat then imitate or imitate then repeat) and Speaker Imitated (imitate speaker 1 or 2) as between-subject factors. Main effects were found for Task and Attitude ($F[1,48]=5.02$, $p=0.03$, partial $\eta^2=0.01$) and $F[1.58, 75.68]= 21.13$, $p<0.001$, Huynh-Feldt-corrected, partial $\eta^2=0.31$) and qualified by a significant interaction, $F(1, 48)=3.55$, $p=0.03$, partial $\eta^2=0.07$). Post-hoc tests (Tukey, $p<.017$) showed that only Pleasantness judgments were more positive after imitation (Figure 1).

Discussion

Imitating a regional accent positively influences stereotypes associated with its speakers. Motor imitation possibly selectively affects attitudes reflecting affiliation and bonding between interaction partners – such as pleasantness –, but not attitudes related to other group characteristics. Previous research has found a positive effect of imitation on affiliation for the interaction partner being imitated (LaFrance & Broadbent, 1976), as well as for the individual imitating his or her interaction partner (Stel & Vonk, 2010). Our research demonstrates

Figure 1: Average rating scores for task and attitude (Error bars: 1 SEM).



that vocal imitating of speech positively alters attitudes about the speaker's perceived pleasantness. Earlier research has shown that vocal imitation enhances action perception under noisy listening conditions (Adank et al., 2010), or that vocal imitation improves understanding of the speaker's message. Our results thus indicate that imitation effects also extend to evaluation of the speaker's characteristics.

References

Adank, P., Evans, B. G., Stuart-Smith, J., & Scott, S. K. 2009. Comprehension of familiar and unfamiliar native accents under adverse listening conditions. *Journal of Experimental Psychology: Human Perception and Performance* 35, 520–529.

Adank, P., Hagoort, P., & Bekkering, H. 2010. Imitation Improves Language Comprehension. *Psychological Science* 21, 1903–1909.

Bayard, D., Weatherall, A., Gallois, C., & Pittam, J. 2001. Pax Americana? Accent attitudinal evaluations in New Zealand, Australia and America. *Journal of Sociolinguistics* 5, 22–49.

Brass, M., Wohlsläger, A., Bekkering, H., & Prinz, W. 2000. Compatibility between observed and executed finger movements: comparing symbolic, spatial and imitative cues. *Brain and Cognition* 44, 124–143.

Chen, M., & Bargh, J. A. 1997. Nonconscious behavioral confirmation processes: The self-fulfilling consequences of automatic stereotype activation. *Journal of Experimental Social Psychology* 33, 541–560.

Delvaux, V., & Soquet, A. 2007. The influence of ambient speech on adult speech productions through unintentional imitation. *Phonetica* 64, 145–173.

Dijksterhuis, A. 2005. Why we are social animals: The high road to imitation as social glue. In: Hurley, S., & Chater, N. (eds), *Perspectives on Imitation: From Neuroscience to Social Science*. Cambridge: MIT Press, Vol. II, 207–220.

Fadiga, L., Craighero, L., Buccino, G., & Rizzolatti, G. 2002. Speech listening specifically modulates the excitability of tongue muscles: a TMS study. *European Journal of Neuroscience* 15, 399–402.

Giles, H., & Billings, A. C. 2004. Assessing Language Attitudes: Speaker Evaluation Studies. In: Davis, A., & Elder, C. (eds), *The Handbook of Applied Linguistics*. Blackwell, 187–209.

Iacoboni, M. 2009. Imitation, empathy, and mirror neurons. *Annual Review of Psychology* 60, 653–670.

LaFrance, M., & Broadbent, M. 1976. Group rapport: Posture sharing as a nonverbal indicator. *Group and Organization Studies* 1, 328–333.

Lambert, W. E., Hodgson, R. C., Gardner, R. C., & Fillenbaum, S. 1960. Evaluational reactions to spoken languages. *Journal of Abnormal and Social Psychology* 60, 44–51.

Pickering, M. J., & Garrod, S. 2007. Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences* 11, 105–110.

Stel, M., & Vonk, R. 2010. Mimicry in social interaction: Benefits for mimickers, mimickees, and their interaction. *British Journal of Psychology* 101, 311–323.

Other-repetition: displaying others' lexical choices as "commentable"

Mathilde Guardiola[1]*, Roxane Bertrand*[1]*, Sylvie Bruxelles*[2]*, Carole Etienne*[2]*, Emilie Jouin-Chardon*[2]*, Florence Oloff*[3]*, Béatrice Priego-Valverde*[1]*, & Véronique Traverso*[2]

[1]Laboratoire Parole et Langage, Aix-Marseille Université & CNRS
[2]Laboratoire ICAR, Univ. Lyon 2 & CNRS
[3]Université de Bâle
mathilde.guardiola@lpl-aix.fr

Lexical other-repetition is a process that consists of repeating words that have been previously produced by another interactant. This leads to a lexical similarity of the participants' discourse. According to Tannen (2007), participants use lexical other-repetition to show their involvement in the interaction. She argues that repetition is useful at several levels: production (repetition facilitates encoding), understanding (repetition facilitates decoding), connection (it maintains cohesion in discourse), and interaction (repetition maintains the link between participants). We focus here on a subcategory of lexical other-repetitions: the phenomenon that we call "pinning" is a form of repair (Schegloff, 2007, 100-101), in so far as one element that has been uttered by one of the participants is afterward treated as a "trouble source". It develops like a repair, and specifically like an other-repair:

A turn that contains an element that will be taken as the source, "the repairable"
B picks up this element in A's turn, repeats it and comments on it
A/B various ways of responding to B's repair

Two features characterize "repair": a/ it consists in a process, and b/ it is not inherently related to a real, an actual or an obvious problem. In the following definition, Schegloff insists on this last aspect: "*Not only are 'obvious' problems unaddressed; anything in the talk may be treated as in need of repair. Everything is, in that sense, a possible repairable or a possible trouble-source. It is overt efforts to deal with trouble-sources or repairables – marked off as distinct within the ongoing talk – that we are terming repair*" (Schegloff, 2007, 100-101). For the cases we are dealing with, the repeat in the second position is neither uttered in order to clarify what has been said, nor in order to ask for an explanation, and even if it does, it is above all oriented towards underlining a more or less strange, unexpected or surprising feature of A's turn.

Example 1: CLAPI. Corpus Grillage (Bert et al., 2010)

```
A  euh::::  un truc euh:  bon y a personnage/ enfin tu vois y a trois:  (0.6) trois
   trois symboles (.)  et::  tu vois:  (0.8) et puis sinon/ ben y a l' réglage des a-
   des asas euh:
P  des a[sas/ ((rire))
L       [ouais
A  des asas et puis voilà
```

In this case, P understands the element pronounced by A. Nevertheless he repeats it, with the same form as in a repair, but orienting it toward mockery, with a "savoring" function (Tannen, 2007). "Des asas" is extracted from the current discourse because it is "worthy" to be commented on. We could say that the repeated element is treated as a "commentable" (cf. "repairable") element and insofar initiates a specific form of repair. We propose to investigate whether the formal alignment (lexical similarity) due to repetition, and possibly leading to a humorous sequence, reveals convergence in interaction. During a previous collaboration in the SPIM project, we gathered a large collection of examples of other-repetition, from a set of corpora of semi-spontaneous and spontaneous interactions, with various contexts and activities (conversation, storytelling; Bertrand et al., 2008; Bert et al., 2010). Various aspects of "pinning" have been studied:

- the process of extracting an element from the previous turn
- the devices used to show that the repair is oriented towards mockery or humor, including the link between the first occurrence and the repeat, on the syntactic, phonetic, and interactional level
- the different types of following turns (i.e., turns responding to the repeat)
- the consequences for the degree of convergence within the sequence.

In the cases we work on, a participant produces the first occurrence of a word or expression in the speech stream, and the other participant notes its incongruity by repeating it, eventually in a humorous perspective. The classical humorous schema consists of the presence of a connector (Greimas, 1966) and a disjunctor (Morin, 1966).

In a "pinning", A produces the connector, which allows two different interpretations (a logical one and an unexpected one) and B repeats it, adding a disjunctor, that actualizes the absurd interpretation. The gap between what A said and what B interprets causes the incongruity, thereby creating humor.

Example 2: Corpus of Interactional Data (Bertrand et al., 2008)

```
LJ  j'ai senti qu(e) ça s'adou-#cissait et bon après on a eu des rapports normaux
AP                                      mh mh
LJ  bon euh
LJ  euh mais au début putain j'é- j'étais mal quoi je euh
AP          mh mh
AP                                      normaux c'(es)t-à-dire euh hum
LJ  ((rires))
AP  avec préservatif ou sans euh
LJ  oh putain ((rires)) ça y est t'es dedans là ((rires))
```

In example 2, LJ describes professional relations, and AP produces back-channel signals. Then AP repeats one word "normaux" (normal), treating it as ambiguous, which appears as a repair sequence. "Rapports" constitutes the connector, since it may be interpreted in different senses, and the repetition is already an attempt of humor. This makes LJ laugh. AP then explicitly actualizes a humorous sexual meaning, for which "préservatif" (condom) constitutes the disjunctor. In this case, there is no modification of the repeated element. In other cases from the corpora, we observe a slight phonetic modification, a syntactic restructuring, or a specific prosodic device, that we precisely analyze, in order to highlight the cues given by the "repeater" to show the humorous or mockery dimension of the repetition. If we consider our data, the process of "pinning" with a humorous purpose leads to three possible situations:

1/ Basic ratification: the humorous repetition receives indeed a basic ratification from A (laughter, "yes"...) but A quickly goes back to his narration. This humorous intervention of B constitutes a short digression from the current narration.

2/ Failure or non-continuation of the humorous mode of communication: the author of the first occurrence (the serious one) refuses to switch into a humorous theme, and continues his activity, whereas B develops, alone, a humorous continuation. The interactants develop parallel sequences, a humorous one and a serious one. They diverge on the kind of their activity at this point of the interaction, until B finally concedes and switches back to the serious modality.

3/ Joint fantasizing (Kotthoff, 2006): contrary to the previous situation, the humorous repetition is used as a starting point for a co-elaborated humorous sequence. In this case, A produces the first occurrence, and B repeats it ("pinning") in order to switch to humor, eventually with slight phonetic modifications, or conversely with a prosodic matching (a type of prosodic orientation defined by Szczepek Reed, 2006). Each participant overbids, and they co-elaborate a humoristic sequence (Bertrand & Priego-Valverde, 2011). We consider these sequences as highly convergent moments in the interaction. Therefore, the same pattern: $<$ S1 produces "serious" discourse, S2 produces humorous one with "pinning" of a part of the discourse $>$ can be the starting point of a highly convergent sequence, or conversely, result in an interactional divergence. This is an evidence of the lack of equivalence between formal similarity (lexical repetition) on the one hand, and interactional convergence on the other.

References

Bert, M., Bruxelles S., Etienne C., Jouin-Chardon E., Lascar J., Mondada L., Teston S., & Traverso V. 2010. Grands corpus et linguistique outillée pour l'étude du français en interaction (plateforme CLAPI et corpus CIEL). *Pratiques* 147-148, 17–35.

Bertrand R., Blache P., Espesser R., Ferré G., Meunier C., Priego-Valverde B., & Rauzy S. 2008. Le CID – Corpus of Interactional Data. *Traitement Automatique des Langues* 49, 105–134.

Bertrand, R., & Priego-Valverde, B. 2011. Does prosody play a specific role in conversational humor? *Pragmatics and Cognition* 19, 333–356.

Greimas, A. J. 1966. *Sémantique structurale*. Paris: PUF.

Kotthoff, H. 2006. Oral genres of humor: on the dialectic of genre knowledge and creative authoring. *Interaction and Linguistic Structures* 44.

Morin, V. 1966. L'histoire drôle. *Communications* 8, 102–119.

Schegloff, E. 2007. *Sequence Organization in Interaction*. Cambridge: CUP.

Szczepek Reed, B. 2006. *Prosodic orientation in English conversation*. Palgrave Macmillan.

Tannen, D. 2007. *Talking Voices: Repetition, Dialogue, and Imagery in Conversational Discourse*. Cambridge: Cambridge University Press.

The temporal dynamics of alignment in multimodal interaction

Bert Oben, Geert Brône, and Kurt Feyaerts

bert.oben@arts.kuleuven.be

The tendency towards convergence between speakers in interactional discourse is a widely studied phenomenon in different disciplines and for different semiotic modalities, including speech, gesture and posture. Recent work in (psycho)linguistics and cognitive psychology has focused on the role of imitative behaviour – alternatively referred to as alignment (Pickering & Garrod, 2004), resonance (Du Bois, 2011) or conceptual pacts (Brennan & Clark, 1996) – in establishing successful communication. Interactional discourse requires speakers and their utterances to be geared to one another in multiple ways so as to facilitate meaning negotiation, and this process requires **alignment at different levels of (linguistic) representation**.

In the majority of cognitive studies dealing with imitative behaviour, the perspective has been largely monodimensional and restricted to minimal contexts. The focus is generally on one semiotic channel or on one linguistic level rather than on alignment as a clustered phenomenon with features co-occurring simultaneously on different levels (e.g. Branigan et al., 2007, on syntactic alignment, Kimbara, 2006, on gestural mimicry, De Fornel, 1992, on postural echoing, and many others). The restriction to minimal contexts, such as pairs of utterances rather than longer sequences of discourse, has led to a relative disregard for the discursive **emergence and persistence** of convergence across speakers in interaction. In this paper, we zoom in on the discursive development of interactive synchronisation as an online and gradual process with multimodal alignment sequences that emerge, persist and die out in the interaction. In order to arrive at such a fine-grained multimodal picture, we conducted a corpus study using the **InSight Interaction Corpus** (Brône & Oben, 2012). This corpus consists of video recordings of both targeted and free-range dyadic interactions, with multiple camera perspectives providing a full view of the dialogue partners' nonverbal behaviour, including hand gestures, facial expressions and body posture (see Figure 1 for a screenshot of the recording set-up). The use of head-mounted scene cameras and eye-trackers provides a unique "speaker-internal" perspective on the conversation, with detailed production information (scene camera and sound) and indices of cognitive processing (eye movements for gaze analysis) for both participants.

In order to arrive at a full-fledged account of alignment sequences in ongoing interaction, we first looked at distributional patterns within each of the semiotic channels, and the role of different levels of representation. We address questions such as "does inter-speaker alignment on one level increase in the course of an ongoing interaction?" and "does alignment persist across the boundaries of individual sequences (e.g. specific thematic units) as part of a larger interaction?". For the first question, we singled out lexical and gestural means of object representation in a set of collaborative tasks in the corpus (where subjects were asked to describe spatial scenes projected on a screen). The data reveal a steady increase in cross-speaker convergence (or a decrease of variation) as interactions unfold, both in lexical choice and gestural means of representation. This effect persists across the boundaries of individual tasks in the interactions of the corpus.

A second step in the analysis deals with potential differences in the distribution and temporal build-up **between different semiotic channels**. To what extent does the emergence of strong cross-speaker alignment follow a similar or different temporal path in different modalities? In order to answer this question on the temporal dynamics of multimodal alignment, we compared the time course of establishing the lexical and gestural routines described in the first step above, and plotted their co-occurrence in time. More specifically, by time-aligning the instances of lexical and gestural cross-speaker convergence, we obtained a fine-grained picture of their multimodal discursive development. The results of this analysis reveal both high degrees of interaction/overlap between channels (what we label "clustered alignment") and some significant differences in durability or persistence.

Figure 1: Recording configuration of the InSight Interaction Corpus.

References

Branigan, H., Pickering, M., McLean, J., & Cleland, A. 2007. Syntactic alignment and participant role in dialogue. *Cognition* 104, 163–197.

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

Brône, G., & Oben, B. 2012. *InSight Interaction. A Multimodal and Multifocal Dialogue Corpus*. Antwerpen/Leuven: University of Leuven.

De Fornel, M. 1992. The return gesture: Some remarks on context, inference and iconic gesture. In: Auer, P., & di Luzio, A. (eds), *The Contextualization of Language*. Amsterdam: John Benjamins, 159–176.

Du Bois, J. 2011. Towards a dialogic syntax. Manuscript in press for *Cognitive Linguistics*.

Kimbara, I. 2006. On gestural mimicry. *Gesture* 6, 39–61.

Pickering, M. J., & Garrod, S. 2004. Toward a mechanistic psychology of dialogue. *Behavioural and Brain Sciences* 27, 169–226.

The influence of gender stereotype threat on speech accommodation in same & mixed-gender negotiations

Lauren Aguilar

Department of Psychology, Stanford University
lauren.aguilar@stanford.edu

The proposed paper brings together theoretical perspectives and methodological practices from social psychology and psycholinguistics in an effort to investigate how social stereotypes affect dyadic speech accommodation processes. The research takes at its foundation the notion that individuals tend to imitate the speech of their interlocutors (e.g. Giles & Coupland, 1991; Pardo, 2006; Pickering & Garrod, 2004), and that such imitation facilitates relationship formation, liking, and connection (Aguilar et al., 2012). The current research will demonstrate that negative stereotypes disrupt speech accommodation processes, which then erode the relational connection between interactants. These processes are examined in the context of negative gender stereotypes about women in negotiation.

Negative stereotypes about women convey that masculine qualities are more beneficial than feminine qualities in negotiation (Kray & Thompson, 2005). When stereotypes about individuals are made salient in evaluative contexts this can induce stereotype threat—the apprehension about being judged on the basis of stereotypes—which undercuts performance and achievement (Steele & Aronson, 1995). Based on the theory of stereotype threat, the prediction was made that fear of confirming negative gender stereotypes may ironically lead women to display socially maladaptive accommodation behavior in negotiations. In social psychology, little research has examined how stereotype threat affects women's social interactions and relationships in performance settings. In communication and linguistics, little research has examined how social psychological factors affect speech accommodation in dyads. In an effort to bridge these fields, the research investigated how stereotype threat affects speech accommodation, relational connection, and instrumental outcomes in dyadic negotiations.

In particular, two studies examined how women and men use phonetic accommodation in dyadic negotiations when gender stereotypes are made salient or when assured that gender stereotypes do not apply to the negotiation. The methodology of both studies consisted of two phases: (1) a recording phase whereby two naïve participants competed in a negotiation while their speech was recorded and (2) a listening phase whereby new listeners judged speech samples from the recording phase to derive measures of phonetic accommodation using the AXB technique (Goldinger, 1998). Both studies found that heightened gender-based stereotype threat affected speech accommodation behaviors and outcomes in dyadic negotiations.

Within same-gender dyads (Study 1) dispositional sensitivity to gender-based rejection in traditionally male settings (RS-gender) affected speech accommodation under stereotype threat (London et al., 2011). Women higher in RS-gender, who are concerned about being judged on the basis of gender in social-evaluative situations, were in a heightened threat state when faced with an explicit reminder about gender stereotypes in negotiation. Result show that when gender stereotype threat was explicitly neutralized, there were no differences in actors' or partners' speech accommodation based on RS-gender. However, when explicitly exposed to gender stereotype threat, women higher in RS-gender in showed less speech accommodation, while women lower in RS-gender use more speech accommodation. Also, partners of women higher in RS-gender exhibited more speech accommodation than partners of women lower in RS-gender under threat.

Within mixed-gender dyads (Study 2) when gender stereotype threat was explicitly neutralized, men accommodated marginally less than women; however when women were exposed explicitly to gender-based identity threat, males increased speech accommodation to female negotiation partners. Females did not show differential speech accommodation between the threat and no threat conditions, and specifically, did not reciprocate male partners' increased accommodation while under stereotype threat.

Across the studies, higher levels of speech accommodation were paralleled by higher levels of partner perceived social connection and liking. Stereotype threat also influenced interpersonal impressions and undercut women's instrumental negotiation outcomes. The results implicate that social stereotypes bear influence on dyadic speech accommodation processes. Furthermore, stereotype threat can affect communication processes in ways that go unnoticed and may affect women's advancement in traditionally male domains such as negotiation.

References

Aguilar, L., Downey, G., Krauss, R., Pardo, J., & Bolger. (Under review). Too Much Too Soon: Rejection sensitivity and speech accommodation in dyadic interaction.

London, B., Downey, G., Romero-Canyas, R., Rattan, A., & Tyson, D. 2011. Gender Rejection Sensitivity and the Academic Self Silencing of Women. *Journal of Personality and Social Psychology* 102, 961–979.

Giles, H., & Coupland, N. 1991. *Language: Contexts and Consequences. Mapping Social Psychology*. Belmont, CA: Thomas Brooks/Cole Publishing Co.

Goldinger, S. D. 1998. Echoes of echoes? An episodic theory of lexical access. *Psychological Review* 105, 251–279.

Kray, L. J., & Thompson, L. 2005. Gender Stereotypes: An examination of theory and research. *Research in Organizational Behavior* 26, 103–182.

Pardo, J. S. 2006. On phonetic convergence during conversational interaction. *Journal of the Acoustical Society of America* 119, 2382–2393.

Pickering, M. J., & Garrod, S. 2004. Toward a mechanistic psychology of dialogue. *Behavioural and Brain Sciences* 27, 169–226.

Steele, C. M., & Aronson, J. A. 1995. Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology* 69, 797–811.

UMR 6057 CNRS
PAROLE ET
LANGAGE