Emotional State Dependence Facilitates Automatic Imitation of Visual Speech

Journal:	*Quarterly Journal of Experimental Psychology*
Manuscript ID	QJE-STD-18-285.R3
Manuscript Type:	Standard Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Virhia, Jasmine; Royal Holloway University of London Kotz, Sonja; Maastricht University, Department of Neuropsychology and Psychopharmacology Adank, Patti; UCL, Chandler House, ;
Keywords:	Imitation, Speech Production, Emotion, Control

RUNNING HEAD: Emotional Speech Imitation

Emotional State Dependence Facilitates Automatic Imitation of Visual Speech

Jasmine Virhia[1], Sonja A. Kotz[2,3], Patti Adank[4]

[1]Department of Psychology, Royal Holloway, University of London, Egham, United Kingdom

[2]Department of Neuropsychology and Psychopharmacology, Maastricht University, Maastricht, The Netherlands

[3]Department of Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

[4]Department of Speech, Hearing & Phonetic Sciences, University College London, UK

Abstract

Observing someone speak automatically triggers cognitive and neural mechanisms required to produce speech, a phenomenon known as *automatic imitation*. Automatic imitation of speech can be measured using the Stimulus Response Compatibility (SRC) paradigm that shows facilitated response times (RTs) when responding to a prompt (e.g., say *aa*) in the presence of a congruent distracter (a video of someone saying *aa*), compared to responding in the presence of an incongruent distracter (a video of someone saying *oo*). Current models of the relation between emotion and cognitive control suggest that automatic imitation can be modulated by varying the stimulus-driven task aspects, i.e., the distracter's emotional valence. It is unclear how the emotional state of the observer affects automatic imitation. The current study explored independent effects of emotional valence of the distracter (Stimulus-driven Dependence) and the observer's emotional state (State Dependence) on automatic imitation of speech. Participants completed an SRC paradigm for visual speech stimuli. They produced a prompt superimposed over a neutral or emotional (happy or angry) distracter video. State Dependence was manipulated by asking participants to speak the prompt in a neutral or emotional (happy or angry) voice. Automatic imitation was facilitated for emotional prompts, but not for emotional distracters, thus implying a facilitating effect of State Dependence. The results are interpreted in the context of theories of automatic imitation and cognitive control, and we suggest that models of automatic imitation are to be modified to accommodate for state dependent and stimulus-driven dependent effects.

Keywords

Imitation, Speech Production, Emotion, Control

Introduction

Humans spontaneously imitate observed actions, including gestures, facial expressions, body posture, and speech (Delvaux & Soquet, 2007; Dimberg, 1982; Goldinger, 1998; LaFrance & Broadbent, 1976; Webb, 1969). This tendency to imitate observed actions is thought to result from activation of mechanisms required to execute this action (Buccino et al., 2004; Fadiga et al., 1998). This phenomenon has been reported for actions of the body as well as facial actions, including speech. For speech, automatic activation production substrates occurs whenever we hear and or/see someone speaking (Nuttall, Kennedy-Higgins, Devlin, & Adank, 2017; Nuttall, Kennedy-Higgins, Hogan, Devlin, & Adank, 2016; Watkins, Strafella, & Paus, 2003; Wilson, Saygin, Sereno, & Iacoboni, 2004). The activation of production mechanisms while observing speech can be measured with functional Magnetic Resonance (fMRI), Transcranial Magnetic Stimulation (TMS), or with behavioural interference paradigms. Using fMRI, it was demonstrated that passively listening to speech engages brain areas linked to speech production, including motor and pre-motor areas (Wilson et al., 2004; Adank & Devlin, 2010). Using TMS, it has been shown that suppressing pre-motor or motor speech areas decreases the ability to categorize speech sounds (Meister, Wilson, Deblieck, Wu, & Iacoboni, 2007; Möttönen & Watkins, 2009). The Stimulus Response Compatibility (SRC) task is an example of a behavioural interference paradigm. SRC tasks were originally used to study the activation of action execution mechanisms during observation of manual actions (e.g., Brass, Wohlsläger, Bekkering, & Prinz, 2000), but have also been employed to study analogous processes in speech (Galantucci, Fowler, & Goldstein, 2009; Jarick & Jones, 2009; Kerzel & Bekkering, 2000; Roon & Gafos, 2015). In a manual SRC task, observers watch videos or images depicting distracter stimuli (e.g., a video of a hand lifting the index or middle finger) and respond to a prompt displayed for a short time on this distracter video. Participants are instructed to ignore the distracter and respond to the prompt only (which cues an action, e.g., lift the index finger). Faster responses can be measured for congruent (e.g., an instruction to lift the index finger together with a video of a hand lifting the index finger) than for incongruent (e.g., the instruction to lift the middle finger together with a video of a hand lifting the index finger) prompt-distracter pairs. For congruent pairs, action observation activates the execution representations associated with executing the prompted action, speeding response times. In contrast, incongruent pairs result in competition between the execution representation activated by the observed action and the association linked to production of the prompt, leading to slower response times. A larger SRC effect, i.e., a larger difference in RTs between incongruent and congruent pairs, indicates that action execution mechanisms were more strongly engaged in response to the distracter stimuli. In the speech version of the SRC task in Kerzel & Bekkering (2000), observers watched distracter videos depicting a person speaking the syllable "ba" or "da" and respond to a written prompt displayed for a short time on top of the distracter video. Participants were instructed to ignore the distracter and respond to the prompt (i.e., say "ba" or "da"). Kerzel & Bekkering (2000) measured faster responses for congruent than for incongruent prompt-distracter pairs, thus showing activation of speech production representations during perception of visible speech.

The activation of action execution mechanisms during action perception as measured using interferences tasks has also been termed *automatic imitation* (Heyes, 2011). Brass & Heyes (2005) and Heyes propose the Association Sequence Learning (ASL) model, which predicts that automatic imitation originates from learned associations between an action's sensory and execution representations, i.e., from the experience of having performed and observed an action. ASL furthermore proposes that automatic imitation is controlled by domain-general executive functions. The current work was conducted within the framework proposed by ASL. See Spengler, von Cramon, and Brass (2009), Wang & Hamilton (2012) and Sowden & Shah (2014) for alternative accounts of automatic imitation.

Much of everyday interaction occurs under emotionally charged conditions, but it is unclear whether and how automatic imitation is influenced by emotional valence. ASL does not make explicit predictions about effects of emotional valence on automatic imitation. Emotional valence could affect automatic imitation either by *observing* an emotional stimulus, or by *producing* an emotional stimulus. Thus far, effects of emotional valence on automatic imitation for speech have not been established, neither for emotional valence of the distracter nor for emotional valence of the prompt.

Yet, emotional valence affects processes linked to automatic imitation, most notably conflict resolution. Using interference tasks related to the SRC task, i.e., Simon, Flanker, or Stroop tasks (Kanske & Kotz, 2010, 2011; Xue et al., 2013; Zinchenko, Kanske, Obermeier, Schröger, & Kotz, 2015), it was shown that conflict resolution for spoken actions is more efficient (implying that automatic imitation is inhibited) for emotional than for neutral distracters. Conflict resolution is a defining feature of cognitive control; it requires the detection and resolving of opposing action tendencies (Kanske, 2012). In a conflict resolution task, participants are presented with a prompt together with distracting information. As in the SRC task, if the distracting information is incongruent, the resulting conflict between goal and distracter needs to be resolved to give a correct response. More efficient conflict resolution implies that participants are able to detect and resolve the conflict more efficiently, leading to smaller interference effects. Conflict resolution and automatic imitation have been suggested to work in tandem: more efficient conflict resolution implies less automatic imitation and vice versa (Cross, Torrisi, Losin, & Iacoboni, 2013). Cross et al. suggest that automatic imitation is governed by an active cognitive control system. Breakdown of this system is linked to loss of control of imitative behaviour, as can be observed in some patients with frontal lobe lesions (De Renzi, Cavalleri, & Facchini, 1996; Lhermitte, Pillon, & Serdaru, 1996).

The current study explored the effects of emotional valence of the distracter and prompt on automatic imitation of visual speech. While various studies have so far examined effects of emotional valance on conflict resolution in the presence of distracting information, only a few studies have so far addressed effects of the emotional valance of the response or prompt. For instance, Hart, Green, Casp, & Belger (2010) found that participants showed less suppression of prepared actions in a Stroop task when the observer was assumed to be in a state of heightened arousal. They conducted a manual conflict resolution study in which participants were primed with emotional images prior to each trial of a Stroop task. They report slower responses for incongruent trials when paired with a preceding aversive image compared to a neutral image. Based on Hart et al.'s results, it appears that producing an emotional prompt reduces the efficiency of conflict resolution, and therefore facilitates, or increases, automatic imitation. Hart et al.'s results support Pessoa's (2008, 2009) model describing the relationship between cognitive and emotional control systems. This model assumes functionally integrated cognitive and emotional systems, where the engagement of cognitive (executive) systems is dependent upon concurrent emotional information being processed. It specifically predicts differential effects of emotion on cognitive control by considering the emotional valence of the distracting stimulus (stimulus-driven effects) and the observer's mental state (state dependence). Pessoa proposes that stimulus-driven dependent effect and state dependent effects can modulate behavioural performance depending on how emotion interacts with control functions, such as conflict resolution. Specifically, emotional stimuli recruit cognitive resources, so that these resources, including cognitive control, are longer be available for other tasks (such as inhibiting distracting information). Pessoa's model does not provide directional predictions of the effects of specific emotional or arousal manipulations on task behaviour. We therefore based our predictions regarding the effect of emotion on automatic imitation of speech on results from studies investigating effects of emotional stimuli on conflict resolution (Hart et al., 2010; Kanske & Kotz, 2010, 2011; Xue et al., 2013; Zinchenko et al., 2015).

We introduced Stimulus-driven Dependence by manipulating the emotional valence of the distracter stimulus and State Dependence by requiring participants to produce the prompt in a neutral or emotional manner. If presenting observers with emotional distracters affects automatic imitation as it affects conflict resolution, then emotional distracters should *inhibit* automatic imitation compared to neutral distracters. We manipulated state-dependence of the observer by asking participants to produce the prompt in a neutral or emotional voice.

We predicted that producing an emotional prompt would engage a similar mechanism as hypothesised to operate by Hart et al. (2010): increased affective processing would reduce cognitive control and facilitate automatic imitation, compared to producing neutral prompts. Emotional state can be manipulated by interfering with participants' facial configurations (Niedenthal, Brauer, Halberstadt, & Innes-Ker, 2001). In Niedenthal et al., participants identified the point in time when a morphed face changed from happy to sad, and vice versa. Participants were slower to perform the task when their facial muscle movements were restricted by keeping a pen sideways in their mouth. Similar effects of restricting or manipulating participants' facial movements on emotion recognition have been reported (Oberman, Winkielman, &

Ramachandran, 2007; Ponari, Conson, D'Amico, Grossi, & Trojano, 2012). We therefore predicted that stimulus-driven effects on emotional processing would affect cognitive control negatively, which in turn should *facilitate* automatic imitation by reducing the suppression of activated corresponding action execution representations.

We tested both predictions using a within-group speech SRC experiment consisting of eight tasks. These tasks either manipulated the emotional valence of the distracter or the emotion valence of the prompt, or both. These tasks were designed to be combined into a full factorial design intended to tease apart effects of stimulus driven effects and state dependent effects, as well as their interaction. The experiment will therefore examine the role of the observer's affective state independently from the emotional valence of the distracter.

Methods

Participants

Behavioural automatic imitation for speech experiments varied considerably in their sample sizes per experiment: Kerzel & Bekkering (2000) tested eight participants, Jarick & Jones (2009) tested 42 participants, Roon & Gafos (2015) tested 38 and 35 participants. Galantucci, Fowler & Goldstein (2006) tested 24, 42, and 24 participants, respectively. Only Galantucci et al. provide effect sizes for their results, and these sizes range from small to medium-sized for the relevant effects. We therefore decided to select a number toward the higher end of the studies surveyed. We tested 40 participants (18 male, mean 26.1y, SD 6.4y, range 19-45y). One additional participant was replaced for making >20% errors. All participants had minimum high school-level education, with the majority currently studying at University level. Experiments were undertaken with the understanding and written consent of each participant according to University College London Research Ethics Committee (UREC, #0599.001).

Materials

We used isolated vowel stimuli taken from a recent experiment evaluating the effect of emotional distracters on cognitive control (Zinchenko et al., 2015). These stimuli were designed as isolated vowels as this was thought to make it easier for the participants to produce the required emotional valance in the prompt and to facilitate obtaining acoustic measurements of their spoken responses. We used six videos recorded for the female actor from Zinchenko et al. These included two neutral videos in which the actor said *aa* or *oo* with a *neutral* facial expression, plus four emotional videos: one in which she said *oo* with an *angry* facial expression, one in which she said *aa* with an *angry* facial expression, one in which she said *oo* with a *happy* facial expression, and one in which she said *aa* with a *happy* facial expression (Table 1). We refer to Zinchenko et al. for a detailed description of the recording procedure and video specifics. The audio was muted using iMovie running on an iMac.

Incongruent and congruent trials were created using written visual prompts that appeared over the mouth of the actor. These prompts were jpeg images with a resolution of 300dpi, .38x.16cm (45×19 pixels) and consisted of white boldfaced Arial font centred on a black background: *aa*, *oo*, *happy*, *angry*. The neutral videos had a duration of 1240ms each and the emotional videos had a duration of 1480ms each. In the neutral videos, the prompt was displayed at four equally spaced SOAs, Stimulus Onset Asynchrony, (cf. Kerzel & Bekkering, 2000; Adank, Nuttall, Bekkering, & Maegherman, 2018). We included four SOAs and the prompt appeared in the video at 240, 490, 740, 990ms for the neutral videos. For the emotional videos, the prompt appeared at 240, 490, 860, and 1170ms. We decided to keep the videos at their original lengths, as it was not feasible to reduce the emotional videos to the length of the neutral videos without also eliminating observable movement.

Procedure

The experiment was conducted in a sound-attenuated and light-controlled booth. The stimuli appeared on a PC monitor located 50cm away from the participant's face. Stimuli were presented using Presentation (Neurobehavioral Systems). Audio was played through Sennheiser HD25 SP-II headphones. Participants' responses were recorded via voice key in Presentation, using a Røde microphone plugged into a Scarlett pre-amplifier connected to the USB input of the test PC. RTs were measured from prompt onset. If no response was detected after 2500ms from the start of the video, participants received a *no response* warning. Responses were recorded onto the PC's sound card and used to verify that participants had produced the correct response. Moreover, these recordings were acoustically analysed (see details below) to establish whether and how participants changed their speech between neutral and emotional conditions.

We designed a series of eight tasks in all of which the participants produced the vowel *aa* or *oo* in a neutral, happy or angry voice, in the presence of a (silent) distracter video depicting a female actor producing *aa* or *oo* with a neutral, happy or angry facial expression. These tasks were designed so that they could be combined into a full factorial design in which we could probe the individual effect of the two factors Stimulus-driven Dependence (neutral or emotional) and State Dependence (neutral or emotional), and their interaction, on automatic imitation as measured by the facilitation for congruent prompt-distracter stimulus pairs relative to incongruent pairs. Note that automatic imitation is measured not by the speed with which participants respond to the written prompt per trial, but by the difference in response speed between the incongruent and congruent prompt-distracter stimulus pairs. Neutral and emotional prompts and distracters were presented in all combinations required to ensure that data was collected for each level of the two factors Stimulus-driven Dependence and State Dependence. Participants completed eight SRC tasks combined into four conditions: NN (Neutral prompt, Neutral distracter), NE (Neutral prompt, Emotional distracter), EN (Emotional prompt, Emotional distracter, EE (Emotional prompt, Emotional distracter) making up the 2×2 factorial design (Figure 1). By comparing the following contrasts, we were able to assess the individual effect of valence of the prompt, distracter, and their interaction: [NN and EN] vs. [NE and EE] probed the effect of Stimulus-driven Dependence on automatic imitation and [NN and NE] vs. [EN and EE] queried the effect of State Dependence.

-- Insert Figure 1 about here ---

In tasks 1 and 2 (Figure 2), participants produced *aa* and *oo* at a neutral tone of voice, in the presence of two neutral distracter videos of the speaker saying *aa* or *oo*. Tasks 1 and 2 thus represent a standard spoken neutral SRC paradigm (cf. Adank et al., 2018). Task 2 was identical to task 1 and was included to ensure that NN would contain the same number of trials as the other four conditions, which combined the trials of two tasks each. Collecting responses for neutral distracters and neutral prompts in a single task in condition NN only would have resulted in a lower number of trials in this condition than the other three conditions, so the task was repeated to avoid this confound.

Tasks 3 and 4 were combined into condition NE, and participants produced neutral prompts *aa* and *oo* in the presence of emotional distracters. Task 3 used happy *aa* and *oo* distracters and task 4 used angry *aa* and *oo* distracters. Tasks 5 and 6 were combined into condition EN. In task 5, the participants spoke *aa* or *oo* in a happy tone of voice in the presence of neutral *aa* and *oo* distracters. In task 6, the participants spoke *aa* or *oo* in angry tone of voice in the presence of the neutral *aa* and *oo* distracters.

Tasks 7 and 8 were combined into condition EE and participants spoke emotional prompts *aa* and *oo* in the presence of emotional *aa* or *oo* distracters. In task 7, participants spoke the vowel *aa* in a happy or angry tone of voice in the presence of happy or angry *aa* distracters. In task 8, participants produced the vowel *oo* in a happy or angry tone of voice in the presence of happy or angry *oo* distracters.

To enable distinguishing between the emotional valence of the distracter and prompt we combined two types of incongruence - vowel and emotional incongruence. Vowel incongruence occurred in tasks that presented participants with a distracter stimulus that showed a difference in the vowel identity between the distracter and the prompt. For instance, in a trial with vowel incongruence, participants might see a video of the speaker saying *oo*, while they are instructed to speak the prompt *aa*. Moreover, to establish how emotional State Dependence affected automatic imitation, participants were to produce the prompt in an emotional tone of voice, e.g., *happy* or *angry*, in the presence of a congruent, or incongruent emotional distracter. For incongruent emotional distracters, participants, for instance, saw a video of the speaker saying a vowel in a *happy* tone of voice, and were requested to respond by producing the same vowel in an *angry* tone of voice.

We ensured that the two types of incongruence were not combined in any of the conditions, thus avoiding a confound of having to process vowel and emotional incongruence in a single SRC task. Tasks 1-6 contain vowel incongruence, while tasks 7-8 contain emotional incongruence. In task 7, participants only responded with an angry or happy *aa* prompt and in task 8 participants responded with an angry or happy *oo* prompt. The written prompts differed across the eight tasks: in tasks 1-6, the prompt was either *aa* or *oo*, and in tasks 7-8, the prompt was either *angry* or *happy* (cf. Figures 1 and 2). The instructions also varied across the eight tasks: in tasks with vowel incongruence (tasks 1-6) participants produced the vowel indicated on the written prompt in a neutral voice: in tasks 1-4, they produced either *aa* or *oo* in the presence of neutral (tasks 1-2) or emotional (tasks 3-4) distracters, while in tasks 5-6, participants produced either *aa* or *oo* in a happy voice (task 5), or in an angry voice (task 6). In tasks with emotional congruence (7-8), they produced the same vowel in an angry or happy voice (happy or angry *aa* in task 7, happy or angry *oo* in task 8), and the corresponding prompts were *happy* and *angry*.

There were thus three types of prompts: prompts *aa* and *oo* used in tasks 1-4 that were to be produced in a neutral voice, prompts *aa* and *oo* in tasks 5-6 that were to be produced in a happy (task 5) or angry (task 6) voice, and prompts *happy* and *angry* in tasks 7-8, were participants produced aa in a happy or angry voice (task 7), or oo in a happy or angry voice (task 8).

Participants received written task-specific instructions before each task and completed a familiarisation session with five trials. The experimenter stayed in the room to ensure the task was performed as instructed and left after the familiarisation session. For all tasks, participants were instructed to speak the prompt aloud as fast as possible, ignoring the distracter video (Figure 3). For the neutral prompts *aa* and *oo* (tasks 1-4, Figures 1 and 2), participants were instructed to speak the instructed vowel as soon as they had seen the prompt, in a neutral voice. For the emotional versions of the prompts *aa* and *oo* (tasks 5-6) they were instructed at the start of the task about the specific emotional they were supposed to produce both vowels in (*happy* in task 5, *angry* in task 6). For the emotional prompts *angry* and *happy*, used in tasks 7 and 8, participants were instructed to speak the vowel specific to that task as soon as they had seen the prompt, in the depicted emotional voice.

Each task consisted of 80 trials and participants performed all eight tasks, resulting in 640 trials in total. Participants could take a short break every 40 trials. Stimulus lists and task order were randomised per participant. The experiment lasted approximately 40 minutes. Data can be found on the Open Science Framework at https://osf.io/u478f/.

-- Insert Figures 2 and 3 about here ---

Analysis

RTs were measured from prompt onset. Errors were excluded from the RT analysis. Errors were defined as too early or late (i.e., faster than 100ms or slower than 1400ms after the onset of the prompt), absent or partial responses, plus trials in which participants produced an incorrect prompt or multiple prompts. Editing of sound files was conducted by a research assistant blind to the Congruence condition. Errors were double-checked by a phonetically-trained listener. RTs were log-transformed before entered into the statistical analyses (Baayen, 2008). We conducted a 2×2×2 repeated-measures factorial ANOVA on the log-transformed RTs, including factors Congruence (congruent vs. incongruent), Stimulus-driven Dependence (neutral emotional: [NN and EN] vs. emotional [NE and EE]), and State Dependence (neutral [NN and NE] vs. emotional [EN and EE]). Factorial designs offer a flexible and powerful approach allowing for direct scrutiny of main effects and their interactions (Collins, Dziak, Kugler, & Trail, 2014). Note that our study was not designed to directly compare effects of the individual eight tasks on automatic imitation. See Figure 1 for a clarification of how the eight individual tasks were collapsed into the four conditions NN, NE, EN, and

EE. Finally, follow-up tests were performed as needed and controlled for multiple comparisons (Bonferroni).

Second, we measured acoustic characteristics of the speech response to establish whether and how participants changed their utterances between neutral and emotional prompts. Acoustic analysis of recorded responses was conducted in Praat (Boersma & Weenink, 2001). We predicted that duration, intensity, and f_o would vary between neutral and emotional conditions (Scherer, 2003). Duration was measured in milliseconds (ms), intensity was measured in dB SPL (Sound Pressure Level). Fundamental frequency (f_o) was measured in hertz (Hz) using Praat's default procedure (autocorrelation). Any intensity or f_o measurements larger or smaller than three standard deviations per participant were excluded. The three acoustic variables were analysed separately using the same procedure as the RTs. The results were analysed collapsed across response vowel (*aa* or *oo*) and emotion (*angry* or *happy*) as a consequence of the specific design of the eight tasks. It was not feasible to split the data according to vowel or emotion as a consequence of the chosen design, as tasks 7 and 8 required participants to produce a single vowel only, and in tasks 3-6 a single emotion was presented either as distracter or prompt.

Results

Participants made 6.5% errors (2.2% late responses, 1.9% wrong responses, 1.8% missing responses, 0.5% early responses). Errors were excluded from the RT analysis. The repeated-measures ANOVA on log-transformed RTs reported a main Congruence effect, with slower RTs for incongruent than for congruent stimulus-response pairs (Figure 4, Tables 1 and 2). A main effect was found for Stimulus-driven Dependence; slower RTs were measured for emotional (NE and EE) than for neutral distracters (NN and EN). Another main effect was reported for State Dependence, slower RTs were measured for emotional (EN and EE) than for neutral prompts (NN and NE). The interaction between State Dependence and Congruence was not significant, indicating that emotional valence of the distracter did not affect automatic imitation. Importantly, the interaction between State Dependence and Congruence was significant. The interaction between Stimulus-driven Dependence and State Dependence was further investigated using a repeated-measures ANOVA on the automatic imitation effect as indexed by difference scores (Figure 5). We calculated the difference scores by subtracting the log-RTs for the congruent trials from the incongruent trials. Follow-up t-tests showed that the main effect for State Dependence could be traced to larger automatic imitation effects for EN than for NE (p=0.003). A final series of t-tests established that the overall log-RTs for both congruent and incongruent trials were significantly slowed down for condition EE relative to the other three conditions (p<0.001). There were no differences between the respective log-RTs associated with congruent and incongruent trials of the other three conditions (Figure 4).

A follow-up repeated-measures ANOVA on the log-RT values for Condition (NN, NE, EN, EE) combined for congruence level, showed significant differences between the four conditions (F[3,117]=53.39, p<0.001, η_{2part}=0.58). This analysis was conducted to confirm that the enhanced automatic imitation effect was not attributable to overall increased effortful processing associated with producing an emotional prompt. Follow-up t-tests showed that responses for EE were slower than for the other three conditions. Yet, no other comparison was significantly different from one another (p<0.0125). Importantly, these result show that there was no difference in log-RT between conditions EN and NN and EN and NE, meaning that producing an emotional prompt in condition EN did not slow down overall task performance. It seemed thus unlikely that increased automatic imitation for emotional prompts was attributable to increased effort. In summary, the results show facilitated automatic imitation for emotional State Dependence compared to neutral State Dependence, irrespective of the emotional valence of the distracter, while the emotional Stimulus-driven Dependence did not affect automatic imitation relative to neutral Stimulus-driven Dependence.

-- Insert Figures 4 and 5 and Table 1 about here ---

The repeated-measures ANOVA (cf. Table 2) for the Duration measurements showed no Congruence effect. A main effect was found for Stimulus-driven Dependence: participants produced longer prompts for emotional distracters (conditions EN, 397ms, and EE, 474ms) compared to neutral distracters (conditions NN, 365ms, and NE, 463ms); participants therefore produced prompts that were on average 22ms longer. Another main effect was found for State Dependence: participants produced longer responses for emotional prompts (conditions EN, 463ms and EE, 474ms) compared to neutral prompts (conditions NN, 365ms, and NE, 397ms) and participants produced prompts that were on average 88ms longer. The results showed an interaction between Stimulus-driven Dependence, State Dependence, and Congruence. However, the locus of this interaction could not be traced, as a follow-up repeated-measures ANOVA on the Duration values for Condition combined for congruence level, was not significant ($F[3,117]=53.39$, $p=0.064$, $\eta_{2part}=0.06$). A final follow-up t-test showed that duration differences were larger for emotional prompts than emotional distracters ($p=.003$).

Intensity measurements showed a main effect of Stimulus-driven Dependence: participants produced the prompt at higher intensities for emotional distracters (conditions NE 65.9dB and EE, 69.2dB) compared to neutral distracters (conditions NN, 65.6dB, and EN, 68.1, dB), and participants produced prompts with an intensity that was 0.7dB higher. A second main effect was found for State Dependence, participants produced emotional prompts at higher intensities (EN, 68.13 and EE, 68.9dB) compared to neutral prompts (NN, 65.6, and NE, 65.8dB), participants produced prompts that had on average an intensity that was 2.9dB higher. No other effects were found. A follow-up t-test showed that intensity differences were larger for emotional prompts than emotional distracters (p<.001).

The results for the f_0 measurements showed a main effect of Stimulus-driven Dependence, participants produced the prompt with a higher f_0 in the presence of an emotional distracter (NE, 170Hz and EE, 208HzHz) than in the presence of a neutral distracter (NN, 169Hz and EN, 170Hz). A second main effect was found for State Dependence: participants produced the prompt with a higher f_0 in conditions with an emotional prompt (EN, 199Hz, and EE, 204Hz) than in conditions with a neutral prompt (EN, 170Hz and NE, 170Hz); participants produced prompts with on average an f_0 that was 5Hz higher. Third, the significant interaction between Stimulus-driven Dependence and State Dependence indicated that for emotional prompts, a larger difference between emotional and neutral distracters (39Hz) was found than for neutral prompts (1Hz), participants produced prompts that had on average an f_0 that was 34Hz higher. A final t-test showed that f_0 differences were larger for emotional prompts than for emotional distracters (p<.001). In sum, participants produced longer emotional prompts at a higher intensity and a higher f_0, than neutral prompts. In addition, they produced longer prompts, at a higher intensity, and with a higher f_0 when the distracter was emotional, and this difference was larger when the prompt was also emotional, for f_0 only. Participants thus changed the production of the prompt when the distracter and/or the prompt were emotional, compared to neutral prompts, but the acoustic changes were considerably larger for emotional prompts than for neutral distracters.

-- *Insert Table 2 about here* ---

Discussion

This study explored the effect of emotional valence on automatic imitation of visual speech in a factorial design with the two fully crossed factors Stimulus-driven Dependence and State Dependence. We found facilitated automatic imitation for emotional State Dependence, but not for emotional Stimulus-driven Dependence, relative to neutral conditions. Our results for emotional State Dependence extend results reported in Hart et al. (2010) to automatic imitation of visual speech. However, no effect of emotional Stimulus-driven Dependence on automatic imitation of visual speech was detected.

State Dependence

The cognitive mechanisms through which the production of emotional speech leads to facilitation of automatic imitation are unclear. One possibility is that being required to produce an emotional stimulus resulted in a (small and temporary) change in the emotional state of the participants. For facial expressions, this concept is also referred to as the *facial feedback hypothesis* (Strack, Martin, & Stepper, 1988). This hypothesis predicts that our facial expression influences our affective experience. Thus, it predicts, for example, that being required to smile leads to an increase in feeling of happiness. A recent meta-analysis found tentative support for this hypothesis (Coles & Larsen, 2017). Yet, as we did not measure any changes in affective mood associated with producing the emotional prompts, we cannot be sure that participants experienced any chances in mood during vowel production. However, we found that participants changed their speech when producing emotional versus neutral prompts. Emotional prompts were produced with longer durations (+88ms), at a higher intensity (+2.9dB) and with a higher f_0, (+34Hz) than neutral prompts. Emotional distracters were also associated with longer duration, higher intensity, and higher f_0, but these effects were much smaller (+22ms, +0.7dB, and 5Hz, respectively) than for emotional prompts versus neutral prompts. We cannot conclude that the effect of emotional valence of the distracter on prompt duration was purely due to emotional valence, as our design contained a potential confounding effect: the emotional videos were longer (1480ms) than the neutral videos (1240ms). Participants might therefore have adjusted their responses accordingly. Yet, it is interesting that participants changed their prompt production, albeit rather subtly, when presented with a (silent) emotional distracter video. Therefore, even though no effect on automatic imitation was found for the emotional valence distracters, participants' behaviour still showed a tendency to produce speech that can be regarded as converging with emotional speech production. For instance, happy and angry speech utterances both have been linked to increases in intensity and f_0, but with an increase in articulation speed (Banse & Scherer, 1996). The results of the acoustic measurements thus demonstrate, first, that participants changed their speech between neutral and emotional prompts and distracters. Second, they changed their productions towards acoustic profiles associated with happy and angry vocalizations. Based on Banse & Scherer's findings, the duration of the prompt productions should have decreased instead of increased, but this difference between the literature and our results is probably due to the fact that we used isolated (and rather long) vowels instead of longer, more complex, utterances. Also, these acoustic changes were not accompanied by slower response times for the conditions NE (Neutral Prompt, Emotional Distracter), EN (Emotional Prompt, Neutral Distracter), relative to condition NN (Neutral Prompt, Neutral Distracter), indicating that the change in speech production did not necessarily relate to increased attention or effort associated with the emotional status of the stimuli. However, it remains to be elucidated how a hypothesized change in emotional status in conditions EN and EE could have resulted in facilitated automatic imitation. Results from Hart et al. suggest that state-dependent effects may downregulate cognitive control systems. In Hart et al. (2010), the emotional valence of the prime slowed down processing of the incongruent distracter in a Stroop task. We found slower response times associated with both congruent and incongruent distracters for condition EE relative to the other three conditions, so our results do not show clear evidence for downregulation of cognitive control systems. Instead, the facilitation of automatic imitation appeared to be due to larger differences between incongruent and congruent trials in conditions EN and EE relative to NN and NE.

It seems unlikely that facilitation of automatic imitation for emotional State Dependence was due to increased task effort. If this were the case, then overall task performance in condition EN (emotional prompt, neutral distracter) would have been poorer than in the two conditions with neutral prompts (NN and NE), which was not the case. We cannot discard potential effects of attentional differences on automatic imitation; this is a general concern for studies focusing on processing of emotional versus neutral stimuli. Emotional stimuli tend to be more arousing than neutral stimuli, and processing of emotion may be regarded as intrinsically confounded with cognitive or attentional load, with more emotionally arousing stimuli capturing attention to a larger degree (Ohman, Flykt, & Esteves, 2001). Recent results suggest that automatic imitation is largely unaffected by cognitive load (Catmur, 2016; Ramsey, Darda, & Downing, in press). We refer to Ramsey (2018) for detailed discussion of specific and general cognitive mechanisms governing automatic imitation. Nevertheless, the links between attention, emotion, and automatic imitation are poorly understood and need to be elucidated further.

Stimulus-driven Dependence

We found no effect of emotional valence of Stimulus-driven Dependence on automatic imitation of visual speech, while previous studies on the effect of emotional distracters on conflict resolution report inhibited automatic imitation (Kanske & Kotz, 2010, 2011; Xue et al., 2013; Zinchenko et al., 2015). The discrepancy between our results and those of past studies could be due to differences between the SRC task used in our study and the Flanker, Simon, and Stroop tasks used in past studies. The main difference between the SRC task and the other tasks is that participants produce a specific goal-directed action in response to the prompt, instead of a button press. It is not clear to which extent conflict resolution and automatic imitation dissociate, but a meta-analysis of 47 neuroimaging studies showed differences in the neural signature of the SRC task compared to the other tasks (Nee, Wager, & Jonides, 2007). SRC tasks were mostly associated with activation in premotor and cingulate areas, while no premotor activation was reported for the Flanker and Stroop tasks. Execution of goal-directed actions in the SRC task invoked a neural network associated with action execution more than the other tasks. Emotional information could interact with goal-directed actions in the SRC task at a later stage of action planning and execution than in other tasks, but this assertion is to be confirmed experimentally.

Our null result also does not fit with results from two recent manual automatic imitation studies. Rauchbauer, Majdandzi, Hummer, Windischberger, & Lamm, (2015) examined how specific emotional valance (e.g., positive versus negative emotional valence, cf., affects automatic imitation for manual actions. Rauchbauer et al. measured how automatic imitation was influenced by the presence of affective facial stimuli in an fMRI experiment. The affective factor consisted of angry and happy faces accompanying the manual prompts, and the social factor consisted of an in- or out-group manipulation. Rauchbauer et al. report increased automatic imitation for happy accompanying faces and for out-group faces. In Butler, Ward, & Ramsey (2016), participants also performed a manual SRC task while the distracter and prompt stimuli were presented with emotional (smiling or angry) or neutral faces (Experiments 1 and 2). The presence of a smiling face was found to facilitate automatic imitation compared to angry or neutral faces. The results of both studies demonstrate that emotional characteristics of the distracter stimuli in SRC tasks affect automatic imitation, under varying social conditions. Therefore, studies on the effect of emotion on automatic imitation of manual actions report *facilitation* of automatic imitation (Rauchbauer, Majdandzi, Hummer, Windischberger, & Lamm, 2015). Rauchbauer et al. employed a manual SRC task, in which the prompt and distracter were presented together with emotional or neutral face. In contrast, the emotional valence of the distracter was varied by using a video displaying an emotional or neutral face articulating the prompt. The lack of a change on automatic imitation could be due to two factors. First, the fact that the prompt was superimposed on the emotional stimulus could have been less arousing than a stimulus presented next to or below the emotional stimulus. A follow-up experiment could explore the effect of the placement of the emotional stimulus, e.g., by placing an image of a second (emotional or neutral) face next to neutral prompt-distracter stimuli as used in the current experiment, to tease apart the effect of the combining of the vowel identify and emotional valence in a single video of a moving face. Second, in everyday communication, speech is usually audiovisual, but in our experiment we used only visual speech. It was decided to use visual speech only, as past speech SRC experiments have demonstrated no effects of distracter stimulus modality (audio-only, audiovisual, or video-only) on automatic imitation of speech (Jarick & Jones, 2009; Adank et al., 2018). Both Jarick & Jones and Adank et al. used neutral speech stimuli, while the current study used emotional and neutral stimuli. It could be the case that viewing a silent emotional face is less arousing that hearing and seeing an emotional vocalisation. Future studies could therefore explore the effects of distracter stimulus modality on automatic imitation of emotional and neutral speech distracters.

Finally, our results also disagree with those from studies showing that observing an emotional action, such as an emotional facial expression (Dimberg, Thunberg, & Elmehed, 2000), or emotional speech (Neumann & Strack, 2000), leads participants to imitate these actions. Participants in Dimberg et al. activated the congruent facial muscles (as measured using electromyography, EMG) when exposed to smiling or frowning facial expressions, and participants in Neuman & Strack (2000) produced speech with intonation patterns congruent with emotional speech they were exposed to. As we did not measure muscle activity with EMG, we cannot exclude the possibility that observing an emotional distracter resulted in an overt imitative response.

Implications for models of imitation and cognitive control

In conclusion, we report facilitation of automatic imitation when the observer produced emotional vocalisations. The current study explored effects of emotional valence of the distracter and the prompt on automatic imitation of speech. As our study was largely exploratory, we suggest that its results serve to inform hypotheses to be tested in a pre-registered follow-up confirmatory experiment, or series of experiments, that test(s) predictions regarding emotional Stimulus-driven and State Dependent effects of on automatic imitation. Moreover, future experiments could evaluate the time point at which the emotional information distracter is introduced, either coinciding with the trial, or preceding the trial. Second, future experiments could examine the location of the emotional information and compare emotional information superimposed on the distracter with the emotional information placed next to the distracter.

Our study was conducted within the framework of the Association Sequence Learning (ASL) account (Brass & Heyes, 2005; Heyes, 2011). ASL predicts that automatic imitation is controlled by domain-general executive functions (Brass & Heyes, 2005; Heyes, 2011). We formulated our hypotheses with respect to Pessoa's (2008, 2009) model, which describes mechanisms governing the relationship between emotional processing and conflict resolution, which is closely linked to automatic imitation. More efficient conflict resolution implies less automatic imitation and vice versa (Cross et al., 2013). We based our prediction regarding the emotional valance of the distracter (Stimulus-driven Dependence) and the prompt on the results of previous studies examining effects of emotional information on conflict resolution (Kanske & Kotz, 2010, 2011; Xue et al., 2013; Zinchenko et al., 2015). These studies report facilitated conflict resolution in the presence of emotional stimuli. We therefore predicted that emotional Stimulus-driven Dependence affect cognitive control positively, which would in turn inhibit automatic imitation by increasing suppression of activated corresponding action execution representations. With respect to the effect of emotional State Dependence, we based our prediction on the results of Hart et al. (2010), who reported less suppression of prepared actions when participants were presented with arousing stimuli. We found facilitated automatic imitation for emotional State Dependence, but not for emotional Stimulus-driven Dependence.

Pessoa's model (2008, 2009) on the relation between emotion and cognitive control does not apply to automatic imitation. We suggest that this model is modified to include imitative behaviour, to fully account for the role of emotional valence associated with everyday communicative actions. Furthermore, we suggest that ASL (Brass & Heyes, 2005; Heyes, 2011) is extended to explain effects of emotional valance on automatic imitation, as well as effects of general executive (cognitive) factors.

Consequently, we recommend to combine Pessoa's model and principles governing ASL into a comprehensive account of automatic imitation that considers the relationship between cognitive control, state dependence and stimulus driven dependence, with emotional valence. Such an account would have the potential to explain effects of arousal and emotional state in an observing individual independently from social or emotional aspects linked to the observed individual in an interactive situation in neutral and under emotionally charged conditions.

Acknowledgements

This work was supported by the BIAL Foundation under grant number 267/14 to Patti Adank. We thank Deborah Talmi for her comments, and Dan Kennedy-Higgins and Gwijde Maegherman for their assistance in data preparation.

References

Adank, P., & Devlin, J. T. (2010). On-line plasticity in spoken sentence comprehension: Adapting to time-compressed speech. *NeuroImage, 49*(1), 1124-1132. doi:10.1016/j.neuroimage.2009.07.032

Adank, P., Nuttall, H. E., Bekkering, H., & Maegherman, G. (2018). Effects of Stimulus Response Compatibility on Covert Imitation of Vowels. *Attention, Perception & Psychophysics, 80*(5), 1290-1299. doi:10.3758/s13414-018-1501-3

Baayen, R. H. (2008). Data sets and functions with "Analyzing Linguistic Data: A practical introduction to statistics". (Version R package version 0.953.).

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology, 70*, 614-636. doi:dx.doi.org/10.1016/S0167-6393(02)00084-5

Boersma, P., & Weenink, D. (2001). Praat, a system for doing phonetics by computer. *Glot International, 5*(10), 341-345.

Brass, M., & Heyes, C. (2005). Imitation: is cognitive neuroscience solving the correspondence problem? *Trends in Cognitive Sciences, 9*, 489-495. doi:10.1016/j.tics.2005. 08.007

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. doi:10.1006/brcg.2000.1225

Buccino, G., Lui, F., Canessa, N., Patteri, I., Lagravinese, G., Benuzzi, F., . . . Rizzolatti, G. (2004). Neural circuits involved in the recognition of actions performed by nonconspecifics: An fMRI study. *Journal of Cognitive Neuroscience, 16*(1), 114-126. doi:10.1162/089892904322755601

Butler, E. E., Ward, R., & Ramsey, R. (2016). The influence of facial signals on the automatic imitation of hand actions. *Frontiers in Psychology, 71*, 1653. doi:10.3389/fpsyg.2016.01653

Catmur, C. (2016). Automatic imitation? Imitative compatibility affects responses at high perceptual load. *Journal of Experimental Psychology: Human Perception and Performance, 42*(4), 530-539. doi:10.1037/xhp0000166

Coles, N. A., & Larsen, J. T. (2017). A meta-analysis of the facial feedback hypothesis literature. *psyarxiv*. doi:10.31234/osf.io/svjru

Collins, L. M., Dziak, J. J., Kugler, K. C., & Trail, J. B. (2014). Factorial experiments: efficient tools for evaluation of intervention components. *American Journal of Preventive Medicine, 47*(4), 498-504. doi:10.1016/j.amepre.2014.06.021

Cross, K. A., Torrisi, S., Losin, E. A. R., & Iacoboni, M. (2013). Controlling automatic imitative tendencies: Interactions between mirror neuron and cognitive control systems. *NeuroImage, 83*, 493–504. doi:10.1016/j.neuroimage.2013.06.060

De Renzi, E., Cavalleri, F., & Facchini, S. (1996). Imitation and utilisation behaviour. *Journal of Neurology, Neurosurgery & Psychiatry, 61*, 396-400.

Delvaux, V., & Soquet, A. (2007). The influence of ambient speech on adult speech productions through unintentional imitation. *Phonetica, 64*, 145-173. doi:10.1159/000107914

Dimberg, U. (1982). Facial reactions to facial expressions. *Psychophysiology, 19*, 643-647. doi:10.1111/j.1469-8986.1982.tb02516.x

Dimberg, U., Thunberg, M., & Elmehed, K. (2000). Unconscious facial reactions to emotional facial expressions. *Psychological Science, 11*, 86-89. doi:10.1111/1467-9280.00221

Fadiga, L., Buccino, G., Craighero, L., Fogassi, L., Gallese, V., & Pavesi, G. (1998). Corticospinal excitability is specifically modulated by motor imagery: a magnetic stimulation study. *Neuropsychologia, 37*(2), 147-158. doi:10.1016/S0028-3932(98)00089-X

Galantucci, B., Fowler, C. A., & Goldstein, L. (2009). Perceptuomotor compatibility effects in speech. *Attention, Perception & Psychophysics, 71*(5), 1138–1149. doi:10.3758/APP.71.5.1138

Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review, 13*(3), 361-377. doi:10.3758/BF03193857

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105*(2), 251-279. doi:10.1037/0278-7393.22.5.1166

Hart, S. J., Green, S. R., Casp, M., & Belger, A. (2010). Emotional priming effects during Stroop task performance. *NeuroImage, 49*, 2662-2670. doi:10.1016/j.neuroimage.2009.10.076

Heyes, C. (2011). Automatic Imitation. *Psychological Bulletin, 137*(3), 463-483. doi:10.1037/a0022288

Jarick, M., & Jones, J. A. (2009). Effects of seeing and hearing speech on speech production: a response time study. *Experimental Brain Research, 195*, 175-182. doi:10.1007/s00221-009-1765-x

Kanske, P. (2012). On the influence of emotion on conflict processing. *Frontiers in Integrative Neuroscience, 6*(42), 1-4. doi:10.3389/fnint.2012.00042

Kanske, P., & Kotz, S. A. (2010). Modulation of early conflict process- ing: N200 responses to emotional words in a flanker task. *Neuropsychologia, 48*(12), 3661-3664. doi:10.1016/j.neuropsychologia.2010.07.021

Kanske, P., & Kotz, S. A. (2011). Positive emotion speeds up conflict processing: ERP responses in an auditory Simon task. *Biological Psychology, 87*, 122–127. doi:10.1016/j.biopsycho.2011.02.018

Kerzel, D., & Bekkering, H. (2000). Motor activation from visible speech: Evidence from stimulus response compatibility. *Journal of Experimental Psychology: Human Perception and Performance, 26*, 634-647. doi:10.10371/0096-1523.26.2.634

LaFrance, M., & Broadbent, M. (1976). Group rapport: Posture sharing as a nonverbal indicator. *Group and Organization Studies, 1*, 328-333. doi:10.1177/105960117600100307

Lhermitte, F., Pillon, B., & Serdaru, M. D. (1996). Human autonomy and the frontal lobes: I. Imitation and utilization behavior. A neuropsychological study of 75 patients. *Annals of Neurology, 19*, 326-334. doi:10.1002/ana.410190404

Meister, I. G., Wilson, S. M., Deblieck, C., Wu, A. D., & Iacoboni, M. (2007). The essential role of premotor cortex in speech perception. *Current Biology, 17*, 1692-1696. doi:10.1016/j.cub.2007.08.064

Möttönen, R., & Watkins, K. E. (2009). Motor representations of articulators contribute to categorical perception of speech sounds. *Journal of Neuroscience, 5*(29), 9819-9825. doi:10.1523/JNEUROSCI.6018-08.2009

Nee, D. E., Wager, T. D., & Jonides, J. (2007). Interference resolution: Insights from a meta-analysis of neuroimaging tasks. *Cognitive Affective & Behavioral Neuroscience, 7*(1), 1-17. doi:10.3758/CABN.7.1.1

Neumann, R., & Strack, F. (2000). *Journal of Personality and Social Psychology, 79*(2), 211-223. doi:10.1037//0022-3514.79.2.211

Niedenthal, P. M., Brauer, M., Halberstadt, J. B., & Innes-Ker, A. H. (2001). When did her smile drop? Facial mimicry and the influences of emotional state on the detection of change in emotional expression. *Cognition and Emotion, 15*, 853-863. doi:10.1080/02699930143000194

Nuttall, H. E., Kennedy-Higgins, D., Devlin, J. T., & Adank, P. (2017). The role of hearing ability and speech distortion in the facilitation of articulatory motor cortex. *Neuropsychologia, 94*(8), 13-22. doi:10.1016/j.neuropsychologia.2016.11.016

Nuttall, H. E., Kennedy-Higgins, D., Hogan, J., Devlin, J. T., & Adank, P. (2016). The effect of speech distortion on the excitability of articulatory motor cortex *NeuroImage, 128*, 218-226. doi:10.1016/j.neuroimage.2015.12.038

Oberman, L. M., Winkielman, P., & Ramachandran, V. S. (2007). Face to face: Blocking facial mimicry can selectively impair recognition of emotional expressions. *Social Neuroscience, 2*(3-4), 167-178. doi:10.1080/17470910701391943

Ohman, A., Flykt, A., & Esteves, F. (2001). Emotion drives attention: detecting the snake in the grass. *Journal of Experimental Psychology: General, 130*(3), 466-487. doi:10.1037/AXJ96-3445.130.3.466

Pessoa, L. (2008). On the relationship between emotion and cognition. *Nature Reviews Neuroscience, 9*(2), 148-158. doi:10.1038/nrn2317

Pessoa, L. (2009). How do emotion and motivation direct executive control? *Trends in Cognitive Sciences, 13*(4), 160-166. doi:10.1016/j.tics.2009.01.006

Ponari, M., Conson, M., D'Amico, N. P., Grossi, D., & Trojano, L. (2012). Mapping correspondence between facial mimicry and emotion recognition in healthy subjects. *Emotion, 12*, 1398–1403. doi:10.1037/a0028588

Ramsey, R. (2018). What are reaction time indices of automatic imitation measuring? *Consciousness & Cognition, 65*, 240-254. doi:10.1016/j.concog.2018.08.006

Ramsey, R., Darda, K. M., & Downing, P. E. (in press). Automatic imitation remains unaffected under cognitive load. *Journal of Experimental Psychology: Human Perception and Performance*.

Rauchbauer, B., Majdandzi, J., Hummer, A., Windischberger, C., & Lamm, C. (2015). Distinct neural processes are engaged in the modulation of mimicry by social group-membership and emotional expressions. *Cortex, 70*, 49-67. doi:10.1016/j.cortex.2015.03.007

Roon, K. D., & Gafos, A. I. (2015). Perceptuo-motor effects of response-distractor compatibility in speech: beyond phonemic identity. *Psychonomic Bulletin & Review, 22*(1), 242-250. doi:10.3758/s13423-014-0666-6

Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication, 40*, 227-256.

Sowden, S., & Shah, P. (2014). Self-other control: a candidate mechanism for social cognitive function. *Frontiers in Human Neuroscience, 8*(789). doi:10.3389/fnhum.2014.00789

Spengler, S., von Cramon, D. Y., & Brass, M. (2009). Control of shared representations relies on key processes involved in mental state attribution. *Human Brain Mapping, 30*(11), 3704-3718. doi:10.1002/hbm.20800

Strack, F., Martin, L. L., & Stepper, S. (1988). Inhibiting and facilitating conditions of the human smile: a nonobtrusive test of the facial feedback hypothesis. *Journal of Personality and Social Psychology, 54*(5), 768-777. doi:10.1037/0022-3514.54.5.768

Wang, Y., & Hamilton, A. F. (2012). Social top-down response modulation (STORM): a model of the control of mimicry in social interaction. *Frontiers in Human Neuroscience, 6*(153), 1-10. doi:10.3389/fnhum.2012.00153

Watkins, K. E., Strafella, A. P., & Paus, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia, 41*(8), 989-994. doi:10.1016/S0028-3932(02)00316-0

Webb, J. T. (1969). Subject speech rates as a function of interviewer behaviour. *Language and Speech, 12*(1), 54-67. doi:10.1177/002383096901200105

Wilson, S. M., Saygin, A. P., Sereno, M. I., & Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nature Neuroscience, 7*, 701-702. doi:doi:10.1038/nn1263

Xue, S., Cui, X., Wang, K., Zhang, S., Qiu, J., & Luo, Y. (2013). Positive emotion modulates cognitive control: An event-related potentials study. *Scandanavian Journal of Psychology, 54*, 82-88. doi:10.1111/sjop.12031

Zinchenko, A., Kanske, P., Obermeier, C., Schröger, E., & Kotz, S. A. (2015). Emotion and goal-directed behavior: ERP evidence on cognitive and emotional conflict. *Social Cognitive and Affective Neuroscience, 10*(11), 1577-1587. doi:10.1093/scan/nsv050

Tables

Table 1.

Average response times (RT) in milliseconds plus standard deviations (SD) for (Vowel and Emotional) Congruent and Incongruent stimulus pairs, per task, collapsed into four conditions

Condition	Congruence	RT	SD	Task	Congruence	RT	SD
NN	Vowel Congruent	559	154	Task 1	Congruent	560	154
					Incongruent	586	150
	Vowel Incongruent	586	147	Task 2	Congruent	558	154
					Incongruent	586	144
NE	Vowel Congruent	578	163	Task 3	Congruent	572	162
					Incongruent	600	152
	Vowel Incongruent	603	160	Task 4	Congruent	584	165
					Incongruent	607	168
EN	Vowel Congruent	569	165	Task 5	Congruent	569	164
					Incongruent	610	170
	Vowel Incongruent	607	168	Task 6	Congruent	569	166
					Incongruent	605	167
EE	Emotional Congruent	665	168	Task 7	Congruent	656	185
					Incongruent	708	186
	Emotional Incongruent	708	188	Task 8	Congruent	673	189
					Incongruent	709	189

Table 2.

Mean difference (M), 95% confidence intervals (CI), and statistics: F-value (F), degrees of freedom (df), p-value (p), and effect sizes (η_{2par}) for the ANOVAs on log-transformed response times (RT), Duration, Intensity, and f_o measurements. SDD: Stimulus-Driven Dependence, SD: State Dependence. Significant results indicated in boldface.

Effect	M	CI	F	*df*	*p*	η_{2par}
			RT			
Congruence	**0.260**	**[0.022 0.031]**	**170.7**	**1, 39**	**<0.001**	**0.81**
Stimulus-driven Dependence	**0.390**	**[0.031 0.048]**	**89.7**	**1, 39**	**<0.001**	**0.70**
State Dependence	**0.380**	**[0.026 0.050]**	**42.2**	**1, 39**	**<0.001**	**0.52**
Stimulus-driven Dependence × Congruence	0.027 0.025	[0.023 0.033] [0.021 0.029]	2.0	1, 39	0.169	0.05
State Dependence × Congruence	**0.023** **0.030**	**[0.019 0.027]** **[0.025 0.035]**	**9.7**	**1, 39**	**0.003**	**0.20**
Stimulus-driven Dependence × State Dependence	**0.070** **0.010**	**[0.051 0.081]**	**34.9**	**1, 39**	**<0.001**	**0.47**

		[-0.002 0.023]				
Congruence × Stimulus-driven Dependence × State Dependence	0.023 0.068 0.010 0.070	[0.001 0.025] [0.053 0.083] [0.000 0.019] [0.051 0.081]	0.04	1, 39	0.821	0
Duration						
Congruence	**0.001**	**[-0.001 0.003]**	**5.3**	**1, 39**	**0.027**	**0.12**
Stimulus-driven Dependence	0.022	[0.003 0.042]	1.6	1, 39	0.220	0.04
State Dependence	**0.088**	**[0.050 0.126]**	**22.4**	**1, 39**	**<0.001**	**0.37**
Stimulus-driven Dependence × Congruence	-0.002 -0.001	[-0.004 0.001] [-0.004 0.002]	0.2	1, 39	0.646	0.01
State Dependence × Congruence	-0.001	[-0.003 0.000]	0.1	1, 39	0.801	0
Stimulus-driven Dependence × State Dependence	-0.001 0.100	[-0.004 0.002]	1.3	1, 39	0.254	0.03

		[0.064 0.135]				
Congruence × Stimulus-driven Dependence × State Dependence	**0.035 0.010 0.032 0.009**	**[0.015 0.055] [0.015 0.055] [0.013 0.51] [-0.026 0.044]**	**7.6**	**1, 39**	**0.009**	**0.16**
Intensity						
Congruence	0.740	[-0.001 0.149]	4.0	1, 39	0.052	0.09
Stimulus-driven Dependence	**0.493**	**[0.108 8.877]**	**6.7**	**1, 39**	**0.013**	**0.15**
State Dependence	**3.173**	**[2.263 4.083]**	**49.8**	**1, 39**	**0.000**	**0.56**
Stimulus-driven Dependence × Congruence	0.120 0.029	[0.0146 0.225] [-0.064 0.121]	2.0	1, 39	0.164	0.05
State Dependence × Congruence	0.102 0.047	[0.038 0.025] [-0.067 0.160]	0.8	1, 39	0.367	0.02

Stimulus-driven Dependence × State Dependence	3.03 0.632	[2.002 4.064] [-0.032 1.298]	0.5	1, 39	0.476	0.01
Congruence × Stimulus-driven Dependence × State Dependence	0.395 0.682 0.311 0.583	[-0.34 0.823] [0.023 1.341] [-0.870 0.708] [-0.102 1.269]	0	1, 39	0.905	0
f_o						
Congruence	0.173	[-0.289 0.635]	0.6	1, 39	0.453	0.01
Stimulus-driven Dependence	**4.891**	**[0.860 8.921]**	**6.0**	**1, 39**	**0.019**	**0.13**
State Dependence	**34.009**	**[25.562 42.456]**	**66.3**	**1, 39**	**<0.001**	**0.63**
Stimulus-driven Dependence × Congruence	0.170 0.176	[-0.538 0.878] [-0.525 0.877]	0	1, 39	0.991	0
State Dependence × Congruence	-0.109 0.455	[-0.587 0.360]	1.4	1, 39	0.247	0.03

		[-0.369 1.279]				
Stimulus-driven Dependence × State Dependence	**29.762 9.137**	**[19.729 39.796] [2.560 15.714]**	**5.8**	**1, 39**	**0.021**	**0.13**
Congruence × Stimulus-driven Dependence × State Dependence	0.531 9.244 0.758 9.029	[-3.292 4.4.354] [2.516 15.972] [-3.125 4.642] [2.482 15.576]	0.2	1, 39	0.669	0.01

Figure captions

Figure 1. Distracters and prompts in the eight tasks, composition of the conditions, and factorial design, ‘-’: *neutral*, ‘+’: *emotional*.

Figure 2. Examples of vowel and emotional congruence/incongruence used in the eight tasks.

Figure 3. Example of the timeline of a vowel incongruent stimulus pair used in tasks 1 and 2 with neutral *oo* prompt and neutral *aa* distracter.

Figure 4. Mean response times per condition split by congruence, error bars: 1SE. NN: Neutral prompt, Neutral distracter, NE: Neutral prompt, Emotional distracter, EN: Emotional prompt, Neutral distracter, EN: Emotional prompt, Emotional distracter.

Figure 5. Mean automatic imitation effects for the response times in milliseconds (ms) per condition; measured as difference between incongruent and congruent trials, error bars: 1SE,*= $p<0.05$, **= $p<0.005$. NN: Neutral prompt, Neutral distracter, NE: Neutral prompt, Emotional distracter, EN: Emotional prompt, Neutral distracter, EN: Emotional prompt, Emotional distracter.

Figure 6. Mean automatic imitation effects for the acoustic measurements Duration in milliseconds (ms), top panel, Intensity in decibels (dB) middle panel, and f_0 in hertz (Hz), bottom panel, per condition, split by congruence. NN: Neutral prompt, Neutral

distracter, NE: Neutral prompt, Emotional distracter, EN: Emotional prompt, Neutral distracter, EN: Emotional prompt, Emotional distracter.

Table 1. Average response times (RT) in milliseconds plus standard deviations (SD) for (Vowel and Emotional) Congruent and Incongruent stimulus pairs, per task, collapsed into four conditions.

Condition	Congruence	RT	SD	Task	Congruence	RT	SD
NN	Vowel Congruent	559	154	Task 1	Congruent	560	154
					Incongruent	586	150
	Vowel Incongruent	586	147	Task 2	Congruent	558	154
					Incongruent	586	144
NE	Vowel Congruent	578	163	Task 3	Congruent	572	162
					Incongruent	600	152
	Vowel Incongruent	603	160	Task 4	Congruent	584	165
					Incongruent	607	168
EN	Vowel Congruent	569	165	Task 5	Congruent	569	164
					Incongruent	610	170
	Vowel Incongruent	607	168	Task 6	Congruent	569	166
					Incongruent	605	167
EE	Emotional Congruent	665	168	Task 7	Congruent	656	185
					Incongruent	708	186
	Emotional Incongruent	708	188	Task 8	Congruent	673	189
					Incongruent	709	189

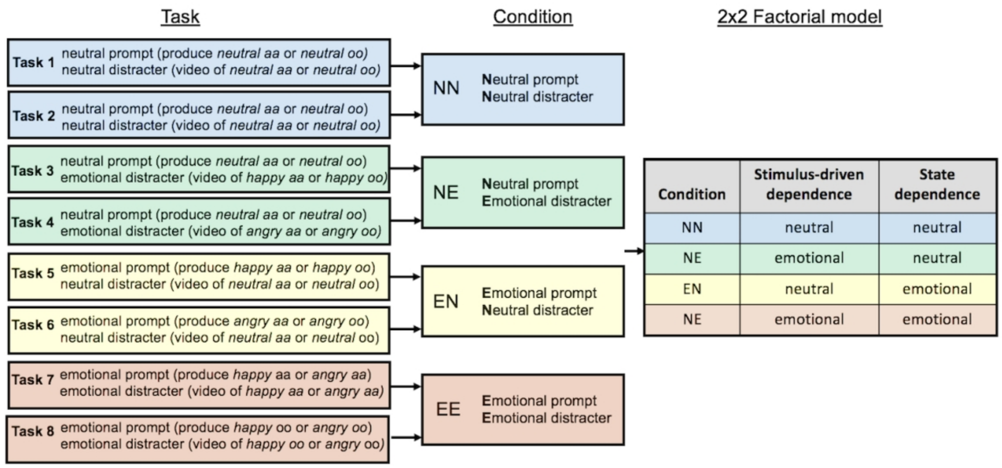
Table 2. Mean difference (M), 95% confidence intervals (CI), and statistics: F-value (F), degrees of freedom (df), p-value (p), and effect sizes (η_{2par}) for the ANOVAs on log-transformed response times (RT), Duration, Intensity, and f_o measurements. SDD: Stimulus-Driven Dependence, SD: State Dependence. Significant results indicated in boldface.

Effect	M	CI	F	*df*	*p*	η_{2par}
RT						
Congruence	**0.260**	**[0.022 0.031]**	**170.7**	**1, 39**	**<0.001**	**0.81**
Stimulus-driven Dependence	**0.390**	**[0.031 0.048]**	**89.7**	**1, 39**	**<0.001**	**0.70**
State Dependence	**0.380**	**[0.026 0.050]**	**42.2**	**1, 39**	**<0.001**	**0.52**
Stimulus-driven Dependence × Congruence	0.027 0.025	[0.023 0.033] [0.021 0.029]	2.0	1, 39	0.169	0.05
State Dependence × Congruence	**0.023** **0.030**	**[0.019 0.027]** **[0.025 0.035]**	**9.7**	**1, 39**	**0.003**	**0.20**
Stimulus-driven Dependence × State Dependence	**0.070** **0.010**	**[0.051 0.081]** **[-0.002 0.023]**	**34.9**	**1, 39**	**<0.001**	**0.47**
Congruence × Stimulus-driven Dependence × State Dependence	0.023 0.068 0.010	[0.001 0.025] [0.053 0.083] [0.000 0.019]	0.04	1, 39	0.821	0

	0.070	[0.051 0.081]				
Duration						
Congruence	**0.001**	**[-0.001 0.003]**	**5.3**	**1, 39**	**0.027**	**0.12**
Stimulus-driven Dependence	0.022	[0.003 0.042]	1.6	1, 39	0.220	0.04
State Dependence	**0.088**	**[0.050 0.126]**	**22.4**	**1, 39**	**<0.001**	**0.37**
Stimulus-driven Dependence × Congruence	-0.002 -0.001	[-0.004 0.001] [-0.004 0.002]	0.2	1, 39	0.646	0.01
State Dependence × Congruence	-0.001	[-0.003 0.000]	0.1	1, 39	0.801	0
Stimulus-driven Dependence × State Dependence	-0.001 0.100	[-0.004 0.002] [0.064 0.135]	1.3	1, 39	0.254	0.03
Congruence × Stimulus-driven Dependence × State Dependence	**0.035** **0.010** **0.032** **0.009**	**[0.015 0.055]** **[0.015 0.055]** **[0.013 0.51]** **[-0.026 0.044]**	**7.6**	**1, 39**	**0.009**	**0.16**
Intensity						
Congruence	0.740	[-0.001 0.149]	4.0	1, 39	0.052	0.09
Stimulus-driven Dependence	**0.493**	**[0.108 8.877]**	**6.7**	**1, 39**	**0.013**	0.15

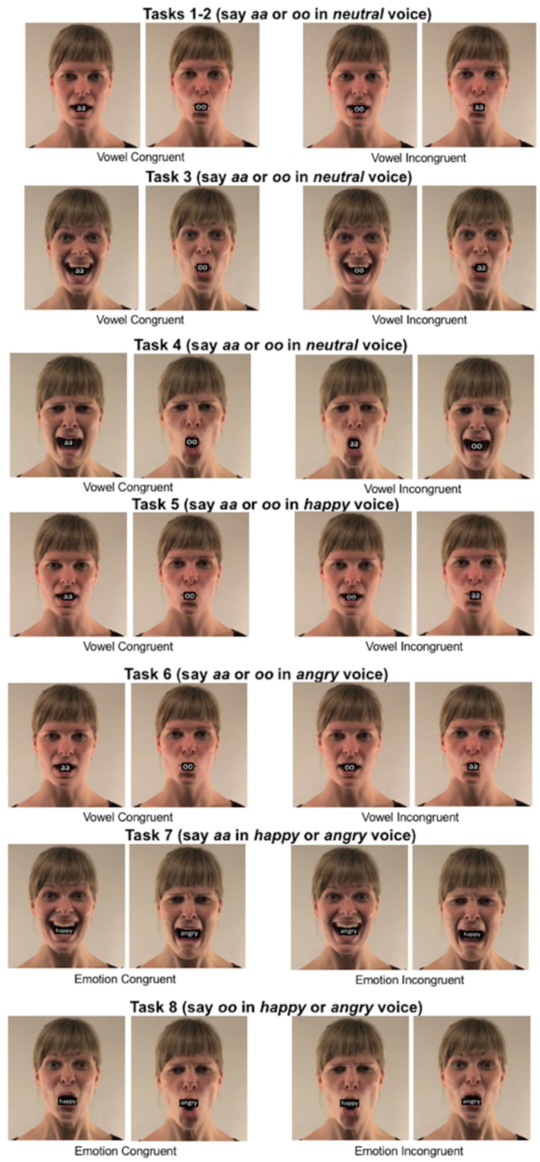
State Dependence	**3.173**	**[2.263 4.083]**	**49.8**	**1, 39**	**0.000**	0.56
Stimulus-driven Dependence × Congruence	0.120 0.029	[0.0146 0.225] [-0.064 0.121]	2.0	1, 39	0.164	0.05
State Dependence × Congruence	0.102 0.047	[0.038 0.025] [-0.067 0.160]	0.8	1, 39	0.367	0.02
Stimulus-driven Dependence × State Dependence	3.03 0.632	[2.002 4.064] [-0.032 1.298]	0.5	1, 39	0.476	0.01
Congruence × Stimulus-driven Dependence × State Dependence	0.395 0.682 0.311 0.583	[-0.34 0.823] [0.023 1.341] [-0.870 0.708] [-0.102 1.269]	0	1, 39	0.905	0
f_o						
Congruence	0.173	[-0.289 0.635]	0.6	1, 39	0.453	0.01
Stimulus-driven Dependence	**4.891**	**[0.860 8.921]**	**6.0**	**1, 39**	**0.019**	0.13
State Dependence	**34.009**	**[25.562 42.456]**	**66.3**	**1, 39**	**<0.001**	0.63
Stimulus-driven Dependence × Congruence	0.170	[-0.538 0.878]	0	1, 39	0.991	0

	0.176	[-0.525 0.877]				
State Dependence × Congruence	-0.109	[-0.587 0.360]	1.4	1, 39	0.247	0.03
	0.455	[-0.369 1.279]				
Stimulus-driven Dependence × State Dependence	**29.762**	**[19.729 39.796]**	**5.8**	**1, 39**	**0.021**	0.13
	9.137	**[2.560 15.714]**				
Congruence × Stimulus-driven Dependence × State Dependence	0.531	[-3.292 4.4.354]	0.2	1, 39	0.669	0.01
	9.244	[2.516 15.972]				
	0.758	[-3.125 4.642]				
	9.029	[2.482 15.576]				



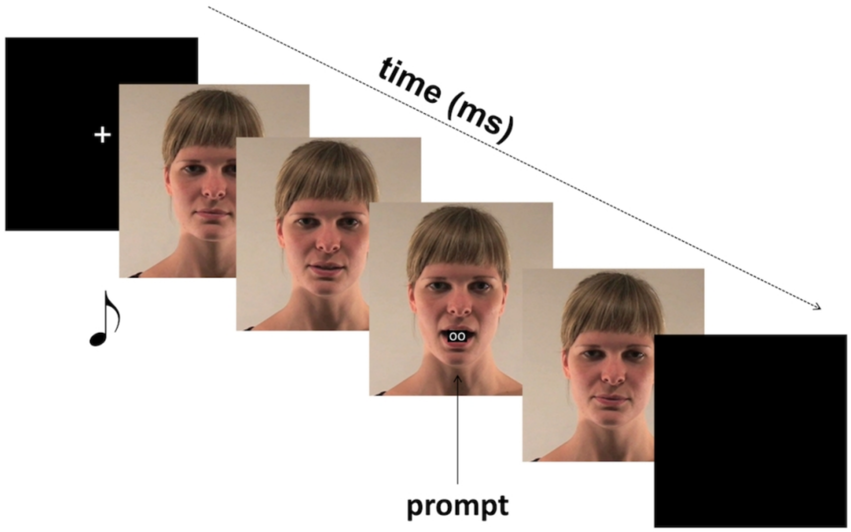
Distracters and prompts in the eight tasks, composition of the conditions, and factorial design.

119x56mm (300 x 300 DPI)



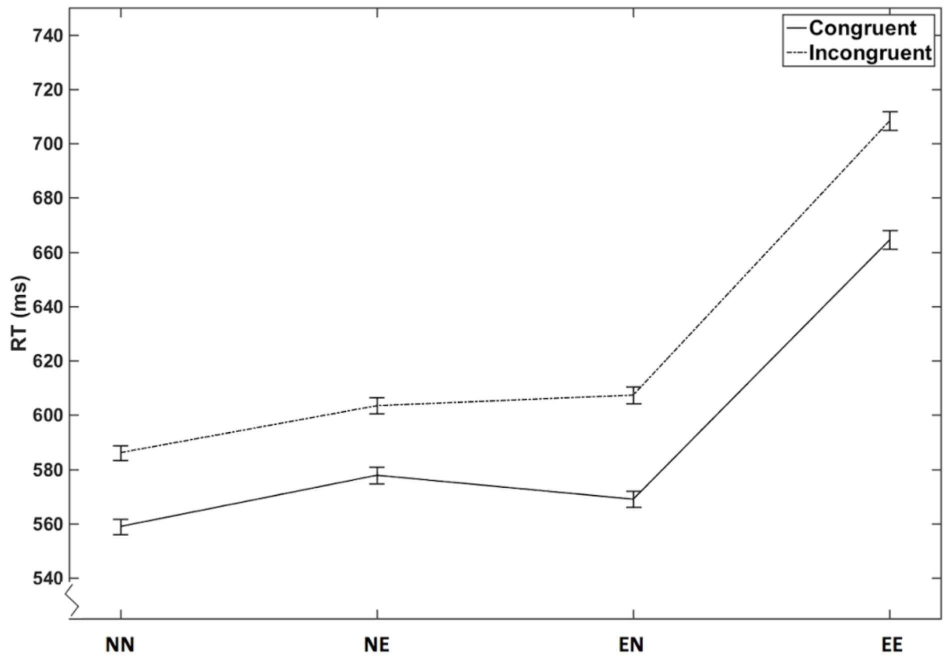
Examples of vowel and emotional congruence/incongruence used in the eight tasks.

119x213mm (300 x 300 DPI)



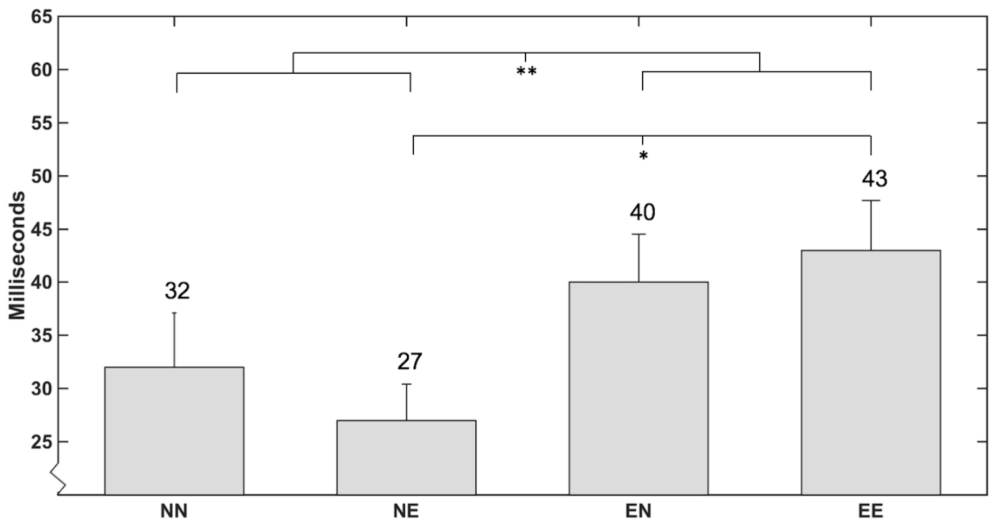
Example of the timeline of a vowel incongruent stimulus pair used in tasks 1 and 2 with neutral oo prompt and neutral aa distracter.

119x67mm (300 x 300 DPI)



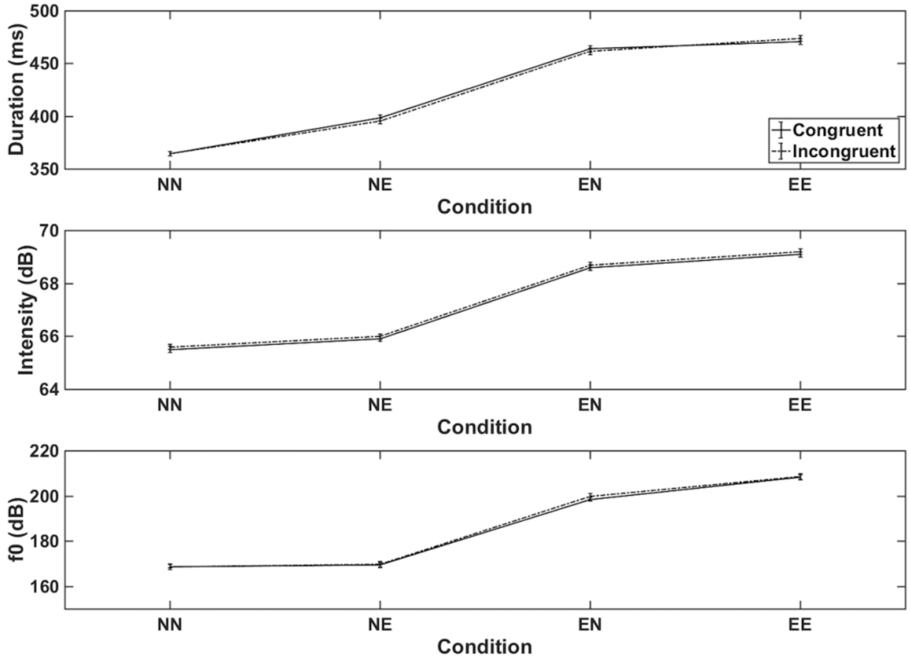
Mean response times per condition split by congruence, error bars: 1SE. NN: Neutral prompt, Neutral distracter, NE: Neutral prompt, Emotional distracter, EN: Emotional prompt, Neutral distracter, EN: Emotional prompt, Emotional distracter.

119x85mm (300 x 300 DPI)



Mean automatic imitation effects for the response times in milliseconds (ms) per condition; measured as difference between incongruent and congruent trials, error bars: 1SE,*= p<0.05, **= p<0.005. NN: Neutral prompt, Neutral distracter, NE: Neutral prompt, Emotional distracter, EN: Emotional prompt, Neutral distracter, EN: Emotional prompt, Emotional distracter.

240x128mm (300 x 300 DPI)



Mean automatic imitation effects for the acoustic measurements Duration in milliseconds (ms), top panel, Intensity in decibels (dB) middle panel, and fo in hertz (Hz), bottom panel, per condition, split by congruence. NN: Neutral prompt, Neutral distracter, NE: Neutral prompt, Emotional distracter, EN: Emotional prompt, Neutral distracter, EN: Emotional prompt, Emotional distracter.