Ipser, A. & Cook, R. Inducing a Concurrent Motor Load Reduces Categorization Precision for Facial Expressions. Journal of Experimental Psychology: Human Perception and Performance, doi: 10.1037/xhp0000177

In press at: *JEP:HPP*
Format: *Research Report*
Running head: *Concurrent motor loads reduce categorization precision*
Word count: *7269*

# Inducing a concurrent motor load reduces categorization precision for facial expressions

Alberta Ipser* and Richard Cook

Department of Psychology, City University London, London, U.K.

*Corresponding author: Alberta.Ipser.1@city.ac.uk
Department of Psychology,
City University London,
Whiskin Street
London, EC1R OJD
Tel: +44 20 7040 8690, Fax: +44 20 7040 8877

**Abstract**

Motor theories of expression perception posit that observers simulate facial expressions within their own motor system, aiding perception and interpretation. Consistent with this view, reports suggest that blocking facial mimicry induces expression labeling errors and alters patterns of ratings. Crucially however, it is unclear whether changes in labeling and rating behavior reflect genuine perceptual phenomena (e.g., greater internal noise associated with expression perception or interpretation) or are products of response bias. In an effort to advance this literature, the present study introduces a new psychophysical paradigm for investigating motor contributions to expression perception that overcomes some of the limitations inherent in simple labeling and rating tasks. Observers were asked to judge whether smiles drawn from a morph continuum were sincere or insincere, in the presence or absence of a motor load induced by the concurrent production of vowel sounds. Having confirmed that smile sincerity judgments depend on cues from both eye and mouth regions (Experiment 1), we demonstrate that vowel production reduces the precision with which smiles are categorized (Experiment 2). In Experiment 3 we replicated this effect when observers were required to produce vowels, but not when they passively listened to the same vowel sounds. In Experiments 4 and 5 we found that gender categorizations, equated for difficulty, were unaffected by vowel production, irrespective of the presence of a smiling expression. These findings greatly advance our understanding of motor contributions to expression perception and represent a timely contribution in light of recent high-profile challenges to the existing evidence base.

**Key words:** Facial expressions, smile sincerity, mirror neurons, simulation, motor theories

## Introduction

The discovery of mirror neurons responsive during the execution and passive observation of facial gestures (Ferrari, Gallese, Rizzolatti, & Fogassi, 2003) prompted considerable interest in motor theories of expression perception. Building on models of 'action understanding' (Gallese & Sinigaglia, 2011; Kilner, 2011; Schütz-Bosbach & Prinz, 2007; Wilson & Knoblich, 2005), motor theories of expression perception posit that observers simulate expressions within their own motor system – a process likened to covert imitation – thereby aiding perception and interpretation (Goldman & Sripada, 2005; Niedenthal, Mermillod, Maringer, & Hess, 2011; Vitale, Williams, Johnston, & Boccignone, 2014). Various authors have linked aberrant simulation to poor expression recognition in Autism Spectrum Disorder (Dapretto et al., 2006), Möbius Syndrome (Bate, Cook, Mole, & Cole, 2013), Parkinson's Disease (Marneweck, Palermo, & Hammond, 2014) and Locked-in Syndrome (Pistoia et al., 2010). However, the clinical significance of putative motor contributions to expression perception remains controversial (Bird & Cook, 2013; Hamilton, 2012; Hickok, 2014; Rives Bogart & Matsumoto, 2009).

### *Existing empirical evidence*

Consistent with motor theories of expression perception, previous reports suggest that blocking facial mimicry impairs performance on expression labeling tasks. For example, observers correctly identified expressions of happiness and disgust less often – their hit rate for these categories was lower on a four alternative-forced-choice (AFC) labeling task – when asked to grip a pen with their teeth, relative to their performance in a free-viewing baseline condition (Oberman, Winkielman, & Ramachandran, 2007). Biting on chopsticks and contracting eyebrow muscles decreased hit-rates for happy and angry expressions respectively, on a 5-AFC labeling task, and both manipulations reduced hit-rates for disgusted expressions, relative to observers in a free-viewing control condition (Ponari, Conson, D'Amico, Grossi, & Trojano, 2012). Recipients of Botox injections – a cosmetic procedure resulting in paralysis of muscles in the forehead – underperformed on the Reading the Mind in the Eyes Test (Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001), in which observers have to label affective and communicative expressions from cues present in the eye-region, compared to individuals given dermal filler – a cosmetic procedure that does not alter muscle function (Neal & Chartrand, 2011). Female participants (but not males) also took longer to respond during an expression labeling task

when asked to clench their teeth, avoid facial movement and attend to feedback from a plaster on their forehead, than when asked to keep their shoulders still (Stel & van Knippenberg, 2008).

Complementary studies have reported motor-induced modulation of expression ratings (Maringer, Krumhuber, Fischer, & Niedenthal, 2011; Rychlowska et al., 2014). Maringer and colleagues asked participants to rate the 'genuineness' of true and false dynamic smiles posed by a computer-generated avatar on a five-point scale (from *not at all genuine* to *very genuine*). Observers in a blocked mimicry condition were informed that more objective judgments are made when facial movement is kept to a minimum, and were required to hold a pen in their mouths. Observers in a free mimicry condition were given no advice and were free to mimic the expressions. Whereas the ratings given to the true and false smiles by participants in the free mimicry condition differed significantly, the ratings given by participants in the blocked mimicry condition did not (Maringer et al., 2011). In a follow-up study, participants were again asked to rate the genuineness of true and false smiles whilst wearing a mouth-guard, squeezing a ball in their hand or under free-viewing conditions (Rychlowska et al., 2014). In contrast to the ratings of the two control groups, the ratings given to the true and false smiles by participants in the mouth-guard condition did not differ significantly.

*Internal noise and bias*

Extant studies have described how labeling errors and genuineness ratings vary as a function of motor load. Strikingly, however, no attempt has been made to determine how the precision with which expressions are categorized is affected by a concurrent motor load. Consequently, it is not clear whether performance decrements induced by the foregoing motor manipulations reflect increases in internal noise or are products of bias. Bias errors are made when observers are prone towards a particular percept (perceptual bias), or tend to make or avoid certain responses (response bias). For example, observers judging smile sincerity might incorrectly label a false smile or award it a high genuineness rating, because they tend to *see* smiles as true, or, because they *choose* to respond 'genuine' whenever they experience a degree of indecision. In contrast, precision errors reflect unreliable responses, which fail to vary as a function of smile sincerity. For

example, precision errors may arise where internal noise detracts from the description and interpretation of the physical differences between stimuli.

The distinction between performance decrements attributable to bias and internal noise is not trivial. Existing motor theories of expression perception (Goldman & Sripada, 2005; Niedenthal et al., 2011; Vitale et al., 2014) appeal to the idea that the motor system conveys a top-down signal that can be used to disambiguate ambiguous sensory descriptions (e.g., Gilbert & Li, 2013). Manipulations that corrupt or block the top-down signal should therefore induce noisier, but unbiased responding[1]. Moreover, performance decrements attributable to bias errors can reflect changes in response bias. For example, discomfort experienced by participants asked to grip items between their teeth for prolonged periods may encourage the use of 'sad' responses – one of the most frequently encountered emotional labels outside of the lab – and thereby reduce hit-rates for happy, disgusted and angry expressions (Oberman et al., 2007; Ponari et al., 2012). Should motor manipulations modulate labeling or rating performance by altering response strategies, putative motor effects on expression perception may not index genuine *perceptual* effects. Importantly, studies that manipulate motor load between-subjects (Maringer et al., 2011; Neal & Chartrand, 2011; Niedenthal, Brauer, Halberstadt, & Innes-Ker, 2001; Ponari et al., 2012; Rychlowska et al., 2014) are particularly vulnerable to differences in response strategy, especially where motor manipulations are confounded with guidance about the effects of mimicry (Maringer et al., 2011; Rychlowska et al., 2014).

*Inconsistent findings*

Not only does the existing evidence base fail to distinguish a loss of categorization precision from bias, it also contains a great deal of inconsistency. First, several motor manipulations have failed to impair expression labeling. Observers asked to chew gum or grip a pen with their lips showed no sign of diminished expression recognition (Oberman et al., 2007). Nevertheless, i) both manipulations inhibit overt facial mimicry and disrupt somatosensory feedback, and ii) electromyographic (EMG) recordings revealed that chewing gum induced strong activation of multiple facial muscles (Oberman et al., 2007). Similarly, observers asked to clench their teeth, avoid facial movement and attend to feedback from a plaster on their forehead, exhibited expression labeling accuracy comparable with controls who were simply asked to keep their shoulders still (Stel & van

Knippenberg, 2008). Moreover, the application of restrictive gel to the faces of observers *improved* performance in an expression labeling task, relative to a second group who had gel applied to their inner arm (Neal & Chartrand, 2011). Each of these manipulations (chewing gum, gripping a pen with the lips, teeth clenching, avoiding facial movement, attending to feedback from a plaster placed on the forehead, the application of restrictive gel to the face) would be expected to interfere with a top-down contribution to expression perception derived from motor simulation.

Second, where observed, motor-induced changes in expression labeling are reported for some expressions, but not for others. For example, in the study reported by Oberman and colleagues, gripping a pen between the teeth was associated with reduced hit-rates when labeling happy, and to some degree disgusted expressions, but had no effect on hit-rates for fear and sadness (Oberman et al., 2007). Ponari and colleagues found that biting on chopsticks reduced hit-rates for disgust, happiness and fear, but did not significantly alter hit-rates for anger, surprise or sadness. Similarly, eyebrow contraction reduced the hit-rate for anger and fear, but labeling of disgust, happiness and surprise was unaffected (Ponari et al., 2012). Facial expressions comprise highly correlated changes from across eye and mouth regions (Jack, Garrod, & Schyns, 2014) and are thought to recruit holistic visual processing whereby eye and mouth variation is integrated into a single perceptual representation (Calder, Young, Keane, & Dean, 2000). Manipulations that successfully disrupt movements made with either the upper (e.g., eye-brow contraction) or lower regions (e.g., gripping a pen between teeth) of the face, and thereby block a top-down signal to the visual system, should affect perception of a wide range of facial expressions and gestures. It is therefore unclear why the perception of some expressions, such as sadness (Oberman et al., 2007; Ponari et al., 2012), should be unaffected.

Finally, many individuals with Möbius syndrome, a disorder associated with partial or complete paralysis of the facial muscles, show unimpaired facial expression recognition (Calder, Keane, Cole, Campbell, & Young, 2000; Rives Bogart & Matsumoto, 2009), or evidence of non-specific visual deficits that extend beyond expression recognition (Bate et al., 2013). Should the motor system make a necessary causal contribution to expression perception – as suggested by reports of motor interference effects in healthy adults – one might expect the expression recognition ability of members of this population to be

disproportionately impaired relative to typical observers. Reconciling the findings from Möbius patients, with expression labeling and rating changes induced by motor manipulations in healthy adults, is potentially problematic, a fact highlighted in recent high-profile critiques of motor theories and their evidence base (Caramazza, Anzellotti, Strnad, & Lingnau, 2014; Hickok, 2014).

*The present study*

In sum, putative motor contributions to expression recognition have stimulated considerable interest, but the findings reported to date are equivocal, and motor theories remain enormously controversial. In an effort to advance this literature, the present study introduces a novel technique for investigating motor contributions to expression perception. Using a psychophysical paradigm that overcomes some of the limitations inherent in simple labeling and rating tasks, we sought to interrogate motor theories of expression perception more rigorously than has been possible to date. Having first confirmed that cues from both the eye and mouth regions contribute to sincerity judgments (Experiment 1), we show that the concurrent production of vowel sounds reduces the precision with which smiles are categorized as sincere or insincere without inducing systematic bias (Experiment 2). We then exclude the possibility that the performance decrement is caused by afferent auditory feedback (Experiment 3), and go on to show that comparable judgments of facial gender are unaffected by the motor load, irrespective of the presence or absence of a smile (Experiments 4 & 5).

## General Methods

*Stimuli*

Facial stimuli were drawn from morph continua, each comprising seven levels, varying attribute strength from 20% to 80% in increments of 10% (Figure 1). Image morphing was performed using Morpheus Photo Morpher Version 3.11 (Morpheus Software, Indianapolis, IN). Stimuli subtended 8° vertically when viewed at a distance of 57.3cm. The smile morphs used in Experiments 1-3 were created by blending one sincere 'enjoyment' smile and one insincere 'control' smile from the Smile Picture Set (Del Giudice & Colle, 2007). The set of gender morphs used in Experiments 4 blended one neutral male and one neutral female face taken from the Radboud Faces Database (Langner et al., 2010). The set of gender morphs used in Experiments 5 blended one

happy male and one happy female face with the same identity and from the same database as in Experiment 4.

Figure-1

*Procedure*

A common trial format was employed throughout the five experiments described (Figure 2a). Trials began with a 1000 Hz tone of 50 ms duration, followed by an interval of 500 ms, after which a facial stimulus was presented for 1000 ms. A response screen was presented 300 ms after stimulus offset, prompting participants to make a 2-AFC judgment with their dominant hand. In Experiments 1-3, participants judged whether the facial stimulus depicted a 'sincere' or 'insincere' smile. In Experiments 4 and 5 participants judged the gender of the face stimulus ('male' or 'female'). To make both of these judgments, observers must place a given exemplar in a natural category defined by characteristic facial variation; whereas males and females face differ in facial form and pigmentation (e.g., Cellerino, Borghetti & Sartucci, 2004), sincere and insincere smiles are defined by characteristic deformations around the eyes and mouth (Del Giudice & Colle, 2007; Niedenthal et al., 2011). The response screen was visible until a response was registered. The next trial began 1000 ms after the response was made. All experiments were programmed in MATLAB (The MathWorks, Natick, MA) using the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997).

Figure-2

Across all experiments, participants' responses were modeled by estimating psychometric functions using the Palamedes toolbox (Prins & Kingdom, 2009). Separate cumulative Gaussian functions were fitted for each condition based on 140 observations (20 presentations × 7 stimulus levels), for each participant. Each function estimated two key parameters: The point of subjective equality (PSE) and internal noise. The PSE is a measure of bias which represents the hypothetical stimulus strength equally likely to be judged as sincere or insincere (Experiments 1-3) or male or female (Experiments 4 and 5). Shifts in the PSE can result from changes in response and/or perceptual bias. The noise estimate is a measure of the precision with which stimuli are categorized and was defined

as the standard deviation of the symmetric Gaussian distribution underlying each cumulative Gaussian function. Noise estimates are inversely related to the slope of the psychometric function; steep and shallow slopes are associated with low and high noise estimates, respectively. Lower noise estimates indicate that observers can perceive subtle differences in stimulus strength and vary their responses accordingly. Greater noise estimates reveal that participants' responses are relatively invariant to changes in stimulus strength, indicative of imprecise categorization. Psychometric functions were modeled using a Maximum Likelihood criterion. PSE and slope measures were free to vary and estimated initially at 50% and 10%, respectively. Guess and lapse rates were fixed at zero. Raw proportions of sincere responses for Experiments 1-3 and male responses for Experiments 4-5, used to estimate psychometric functions are presented as a function of stimulus level in Table 1 and Table 2, respectively.

Where psychometric functions were found to vary across conditions, Signal Detection Theory (Green and Swets, 1966) was used to determine whether effects were seen across the stimulus range. For each observer, we estimated their ability to categorize i) the 20% and 80% levels, ii) the 30% and 70% levels, and iii) 40% and 60% levels. Hits and false alarms were defined according to the categorization task; sincere responses in the presence of the 60%, 70%, and 80% sincere stimuli were treated as hits, and sincere responses in the presence of 20% 30% and 40% sincere stimuli as false alarms[2]. Where participants made no misses / false alarms in a given condition, probabilities of 0.9995 and 0.0005 were assigned for the purposes of the *d'* calculation.

In each experiment, sample size was determined a priori based on i) the need to counterbalance the order of three conditions manipulated within-subjects, and ii) power analysis conducted assuming a large effect size (Cohen, 1988). Ethical clearance was granted by the local ethics committee and the study was conducted in line with the ethical guidelines laid down in the 6$^{th}$ (2008) Declaration of Helsinki. All participants gave informed consent.

**Experiment 1**

It has been suggested that manipulations that primarily interfere with upper (e.g., eyebrow contraction) or lower facial mimicry (e.g., biting on a pen), may disproportionately impair

perception of expressions characterized by changes in the eye-region and mouth-region, respectively (Ponari et al., 2012)[3]. It is well established that sincere smiles are associated with contractions of the muscles around the eyes (the orbicularis oculi and the pars lateralis), the so-called 'Duchenne marker' (Duchenne de Boulogne, 1862). Should judgments of smile sincerity be based solely on the Duchenne marker, blocking overt mimicry (or covert simulation) of the mouth through vowel production may induce little perceptual decrement. However, if sincerity judgments also rely on cues present in the mouth region, loading the motor structures associated with mouth movements ought to impair perceptual judgments (Ponari et al., 2012). To aid clear interpretation of our subsequent experiments, we therefore sought to confirm that observers use cues from both the eye and mouth regions when judging the sincerity of the smile morphs (Figure 1a).

*Methods*

Twenty-four healthy adults (7 males, $M_{age}$ = 31.29, 1 left handed) participated in Experiment 1. Psychometric functions were estimated for three viewing conditions. In the whole-face condition, participants were presented with a smiling face and were free to use cues from the eye-region, the mouth-region, or both to judge sincerity. In the mouth-only condition the eye-region was occluded, forcing observers to use cues from the mouth region. In the eyes-only condition, the mouth-region was occluded, forcing observers to use cues from the eye-region. Viewing condition (whole-face, eyes-only, mouth-only) was blocked. The order in which participants completed the three blocks was fully counterbalanced. Within each block of 140 trials, the seven levels of sincerity appeared 20 times, in a randomized order.

Table-1

*Results and discussion*

If observers use cues from both regions when categorizing whole-face smiles, performance in the eyes-only and mouth-only conditions should independently predict whole-face performance. A multiple regression analysis was therefore conducted in which noise estimates from the whole-face condition were regressed onto the noise estimates from the eyes-only and mouth-only conditions. There was no significant correlation between the predictors [$r(23) = .27$, $p = .196$] and the dependent variable was normally

distributed [$W(24) = .96$, $p = .413$]. Crucially, variability in noise estimates observed in the eyes-only ($\beta = .441$, $p = .009$) and in the mouth-only ($\beta = .485$, $p = .005$) conditions predicted unique variance in whole-face performance. The combined regression model was highly significant [$F(1,22) = 12.620$, $p < .001$] explaining 54.5% of variability in whole-face noise estimates.

Consistent with the existing literature on the Duchenne marker (reviewed by Niedenthal et al., 2011), the results from Experiment 1 indicate that information present in the top half of the face is useful when judging smile sincerity. Critically, however, ability to judge smile sincerity from mouth cues also predicted independent variance in whole-face performance, confirming that whole-face sincerity judgments depend on cues derived from both the eye and mouth region. Existing simulation accounts of expression perception (Goldman & Sripada, 2005; Niedenthal et al., 2011), therefore predict that manipulations that load high-level motor areas responsible for planning mouth actions, should impair whole-face sincerity judgments (Experiments 2 and 3), irrespective of the nature of the top-down contribution; i.e., whether it is feature-specific, aiding interpretation of mouth variation only (e.g., Ponari et al., 2012), or global, aiding interpretation of the entire facial configuration (e.g., Friston, 2005; Gregory, 1997).

**Experiment 2**

Having determined that cues derived from the mouth area contribute to observers' judgments of smile sincerity, Experiment 2 examined whether a concurrent motor load would modulate performance on our psychophysical task. To induce motor interference, participants were asked to produce vowel sounds during stimulus presentation. Crucially, motor theories regard covert simulation in high-level motor areas as the core mechanism for action or expression understanding (Gallese & Sinigaglia, 2011; Goldman & Sripada, 2005; Kilner, 2011; Schütz-Bosbach & Prinz, 2007; Vitale et al., 2014; Wilson & Knoblich, 2005). Because vowel production has both planning and production components it is likely to load high-level motor structures such as the premotor cortex (Rizzolatti & Luppino, 2001).

Manipulations used to block mimicry in previous studies – biting on a pen (Niedenthal et al., 2001; Oberman et al., 2007) or on chopsticks (Ponari et al., 2012), wearing a mouth-guard (Rychlowska et al., 2014), 2014), sitting still (Stel & van Knippenberg, 2008), and botox injections (Neal & Chartrand, 2011) – block the peripheral motor system and may distort to some degree afferent feedback where observers make overt movements. However, it is unclear whether the foregoing manipulations effectively engage regions such as premotor cortex, recruited during the planning and coordination of action (Rizzolatti & Luppino, 2001). Interestingly, a manipulation more likely to load premotor cortex – chewing gum – failed to modulate performance on an expression labeling task (Oberman et al., 2007).

*Methods*

Twenty-four healthy adults (10 males, $M_{age}$ = 28.08, 2 left handed) participated in Experiment 2. Participants completed the whole-face task from Experiment 1 under three conditions: In the baseline condition, participants viewed the stimuli without the requirement to produce a vowel sound. In the remaining conditions participants produced one of two vowel sounds – either /i/ (pronounced '*eeh' as in 'cheese'*) or '/u/' (pronounced '*ooh' as in 'choose'*) – cued by the tone at the start of each trial. Participants were required to produce the vowel sound as soon as the tone was detected, and maintain the sound until the offset of the stimulus image. Auditory responses were recorded and response latencies analyzed using Audacity sound-editing software (http://audacity.sourceforge.net/). The order in which participants completed the conditions was fully counterbalanced across the sample.

*Results and discussion*

Analysis of the response latencies indicated that the vowel production task was performed well, with 98.3% of speech sounds produced within ±600 ms of the stimulus onset (Figure 3a). The noise and PSE estimates were analyzed using ANOVA with viewing condition (baseline, produce /i/, produce /u/) as a within-subjects factor. The analyses revealed a significant main effect of viewing condition on noise estimates [$F(2,46) = 9.06$, $p = .001$, $\eta^2 = .283$]. Planned pairwise comparisons revealed significantly lower noise estimates, indicative of greater categorization precision, in the baseline condition ($M = 8.67$, $SD = 4.02$) compared with both the produce /i/ ($M = 12.20$, $SD = 5.72$) [$t(23) = 3.57$, $p = .002$]

and produce /u/ ($M$ = 11.59, $SD$ = 4.63) conditions [$t(23) = 4.16$, $p < .001$]. The difference between the noise estimates in the produce /i/ and produce /u/ conditions was not significant [$t(23) = .66$, $p = .518$]. No significant effect of condition on PSE estimates was found [$F(2,46) = .40$, $p = .672$, $\eta^2$=.017] indicating that observers' bias did not differ across the baseline ($M$ = 51.17, $SD$ = 8.86), produce /u/ ($M$ = 50.17, $SD$ = 10.14) and produce /i/ conditions ($M$ = 50.29, $SD$ = 8.62).

Figure-3

To determine whether the motor load impaired categorization across the entire stimulus range, we estimated each observer's ability to categorize i) the 20% and 80% levels, ii) the 30% and 70% levels, and iii) the 40% and 60% levels using Signal Detection Theory (Green and Swets, 1966). Hits and false alarms were defined according to the categorization task; sincere responses in the presence of the 60%, 70%, and 80% sincere stimuli were treated as hits, and sincere responses in the presence of 20% 30% and 40% sincere stimuli as false alarms. Where participants made no misses or no false alarms in a given condition, probabilities of 0.9995 and 0.0005 were assigned for the purposes of the *d'* calculation. The resulting distributions of *d'* statistics were analyzed using ANOVA with Viewing Condition (baseline, produce /i/, produce /u/) and Stimulus Difference (20-80%, 30-70%, 40-60%).

The analysis revealed a main effect of Viewing Condition [$F(2,46) = 5.29$, $p = .009$, $\eta^2$ = .187]. Pairwise comparisons revealed significantly higher attribution sensitivity in the baseline condition ($M$ = 4.66, $SD$ = 0.96) compared to the produce /i/ ($M$ = 4.03, $SD$ = 1.31) [$t(23) = 2.55$, $p = .018$] and produce /u/ conditions ($M$ = 4.03, $SD$ = 1.16) [$t(23)$ = 3.32, $p = .003$]. Attribution sensitivity did not differ in the produce /i/ and produce /u/ conditions [$t(23) = .01$, $p = .992$]. The analysis also revealed a main effect of Stimulus Difference [$F(2,46) = 126.28$, $p < .001$, $\eta^2$=.846]. Attribution sensitivity for the 20% and 80% levels ($M$ = 5.45, $SD$ = 0.93) exceeded that for the 30% and 70% levels ($M$ = 4.61, $SD$ = 1.30), [$t(23) = 9.95$, $p < .001$], and for the 40% and 60% levels ($M$ = 2.66, $SD$ = 1.00) [$t(23) = 16.15$, $p < .001$]. Attribution sensitivity for the 30% and 70% levels also exceeded that of the 40% and 60% levels [$t(23) = 4.97$, $p < .001$]. Importantly, however, no interaction was seen between the Viewing Condition and Stimulus Difference [$F(4,92)$

= 1.54, $p$ = .196, $\eta^2$=.063], suggesting that motor load impaired categorization across the entire stimulus range.

These results indicate that inducing a motor load through vowel production decreases the precision with which smiles are categorized as sincere or insincere, without introducing systematic bias, consistent with motor theories of expression perception (Goldman & Sripada, 2005; Niedenthal et al., 2011; Vitale et al., 2014). In light of the findings from Experiment 1, the motor load may have prevented observers using information around the mouth region to inform their sincerity judgements, thereby reducing categorization precision. Alternatively, the load may have detracted from a top-down contribution in the form of a global interpretation of the facial configuration. The detrimental effect of motor load was not restricted to the highly ambiguous stimuli around the middle of the morph continuum; effects were seen across the range of stimulus intensities.

**Experiment 3**

The results of Experiment 2 suggest that concurrent vowel production increases the internal noise associated with expression perception and interpretation, consistent with hypothesized motor contributions to expression perception. It is possible, however, that the decrease in categorization precision did not reflect the presence of the motor load per se, but rather distraction caused by afferent auditory feedback; in other words, the performance decrement may have been induced by the resultant speech sounds, not the speech production itself. Experiment 3 sought to test this alternative account. Participants judged the sincerity of whole-face smiles in a baseline viewing condition, a production condition in which participants produced the vowel /i/, and a passive condition in which observers heard the vowel /i/ during stimulus presentation. If the performance decrement observed in Experiment 2 was a product of afferent auditory feedback, the loss of sensitivity should also be induced by the passive auditory signals.

*Methods*

Twenty-four healthy adults (2 males, $M_{age}$ = 20.17, 1 left handed) participated in Experiment 3. The baseline and produce /i/ conditions were identical to those described in Experiment 2. To provide a conservative test of the auditory feedback account, we sought to maximize the salience of the auditory stimuli presented in the passive /i/ condition.

During the passive block, each participant encountered twenty-five tokens of /i/ (five examples of vowels produced by five participants in Experiment 2). Male and female participants heard tokens produced by male and female actors, respectively. The auditory stimulus could occur at 0 ms, ±100 ms, ±200 ms, ± 300 ms, ± 400 ms, ±500 ms, relative to the onset of the visual stimulus. The distribution of vowel onset asynchronies was yoked to that seen in the production condition in Experiment 2. The offsets of the auditory and visual stimuli in the passive condition were always synchronized. The order in which participants completed the three conditions was fully counterbalanced.

*Results and discussion*

Analysis of the response latencies indicated that the vowel production task in the motor condition was performed well, with 99.1% of speech sounds produced within ±600 ms of the stimulus onset (Figure 3b). Noise and PSE estimates were analyzed using ANOVA with Viewing Condition (baseline, produce /i/, passive /i/) as a within-subjects factor. The analyses revealed a significant main effect of Viewing Condition on noise estimates [$F(2,46) = 3.74$, $p = .031$, $\eta^2 = .140$]. Pairwise comparisons revealed significantly lower noise estimates, indicative of greater categorization precision, in the baseline condition ($M = 10.28$, $SD = 4.45$) compared to the produce /i/ condition ($M = 13.40$, $SD = 7.81$) [$t(23) = 2.51$, $p = .019$], replicating the interference effect seen in Experiment 2. Crucially, however, the noise estimates seen in the passive /i/ condition ($M = 11.23$, $SD = 5.26$) did not differ from those seen in the baseline condition [$t(23) = .96$, $p < .346$], arguing against an afferent auditory feedback account of the precision decrement. There was no effect of Viewing Condition on PSE estimates [$F(2,46) = .28$, $p = .755$, $\eta^2 = .012$], indicating that observers' bias was comparable in the baseline ($M = 52.71$, $SD = 6.35$), produce /i/ ($M = 53.08$, $SD = 7.20$) and passive /i/ conditions ($M = 52.21$, $SD = 6.12$).

To determine whether the motor load impaired categorization across the entire stimulus range, we estimated observers' ability to categorize the 20% and 80% levels, the 30% and 70% levels, and the 40% and 60% levels. The resulting distributions of *d'* statistics were analyzed using ANOVA with Viewing Condition (baseline, produce /i/, passive /i/) and Stimulus Difference (20-80%, 30-70%, 40-60%) as within-subjects factors. The analysis revealed a main effect of Viewing Condition [$F(2,46) = 4.36$, $p = .018$, $\eta^2 = .159$]. Pairwise comparisons revealed significantly lower attribution sensitivity in the produce /i/ condition

($M$ = 3.84, $SD$ = 1.52) than in the baseline ($M$ = 4.46, $SD$ = 1.04) [$t(23) = 2.53$, $p = .019$] and the passive /i/ conditions ($M$ = 4.32, $SD$ = 1.18) [$t(23) = 2.08$, $p < .05$]. The difference between the baseline and passive /i/ conditions was not significant [$t(23) = 0.78$, $p = .446$]. The analysis also revealed a main effect of Stimulus Difference [$F(2,46) = 126.28$, $p < .001$, $\eta^2$=.846]. Attribution sensitivity for the 20% and 80% levels ($M$ = 5.36, $SD$ = 1.27) exceeded that seen for the 70% and 30% levels ($M$ = 4.76, $SD$ = 1.37) [$t(23) = 2.90$, $p = .008$] and for the 60% and 40% levels ($M$ = 2.50, $SD$ = 1.01) [$t(23) = 17.19$, $p < .001$]. Attribution sensitivity for the 70% and 30% levels also exceeded that seen for the 60% and 40% levels [$t(23) = 11.30$, $p < .001$]. Once again, no interaction was seen between the Viewing Condition and Stimulus Difference [$F(4,92) = 1.54$, $p = .213$, $\eta^2$=.063], suggesting that motor load impaired categorization across the entire stimulus range.

**Experiment 4**

The pattern of results observed in Experiments 2 and 3 is consistent with the view that motor processes aid the perception and interpretation of facial expressions via top-down influence (Goldman & Sripada, 2005; Oberman et al., 2007; Vitale et al., 2014). However, these results might also reflect the additional task demands associated with the vowel production task (akin to generic distraction). Afferent somatosensory feedback, or the planning and production components of the concurrent motor task, could load the wider cognitive system, thereby disrupting performance on a wide range of tasks, extending beyond expression perception. Experiment 4 sought to distinguish between these possibilities by testing whether vowel production also modulates categorization precision when observers judge facial gender. Because these tasks have similar demands, a generic distraction account predicts that vowel production should induce comparable precision decrements for gender classification.

*Methods*

Twenty-four healthy adults (9 males, $M_{age}$ = 30.42, 1 left handed) participated in Experiment 4. Participants were required to judge the gender of whole-face stimuli (responding ‘male’ or ‘female’) drawn from a continuum blending a male and female face (Figure 1b). Both faces exhibited a ‘neutral’ expression; i.e. neither appeared to express emotion. Psychometric functions were modelled under three viewing conditions, baseline, production of /i/ and production of /u/. The order in which participants completed the

conditions was again fully counterbalanced. With the exception of the stimuli presented (drawn from a male-female morph) and the judgment made (judging facial gender), the methods were identical to those employed in Experiment 2.

Table-2

*Results and discussion*

Analysis of the response latencies indicated that the vowel production task was again performed well, with 97.1% of speech sounds produced within ±600 ms of the onset of the stimulus (Figure 4a). This level of performance is comparable with that seen in Experiment 2. Noise and PSE estimates were analyzed using ANOVA, with condition (baseline, produce /i/ and produce /u/) as a within-subjects factor. In contrast to Experiment 2, the analysis revealed no effect of condition on noise estimates [$F(2,46) = .37$, $p = .693$, $\eta^2 = .016$], indicating that categorization precision in the baseline ($M = 8.32\%$, $SD = 3.39\%$), produce /i/ ($M = 7.66\%$, $SD = 4.31\%$) and produce /u/ ($M = 7.94\%$, $SD = 3.59\%$) conditions was comparable. Similarly, PSE estimates did not vary across condition [$F(2,46) = 1.96$, $p = .153$, $\eta^2 = .078$], indicating that observers' bias did not differ in the baseline ($M = 44.23$, $SD = 4.77$), produce /i/ ($M = 44.60$, $SD = 4.36$) and produce /u/ ($M = 46.21$, $SD = 6.31$) conditions. The results from Experiment 4 indicate that the precision decrements induced by vowel production in Experiments 2 and 3 do not extend to judgments of facial gender, arguing against a generic distraction account.

Figure-4

**Experiment 5**

Experiment 4 indicates that precision decrements induced by vowel production in Experiments 2 and 3 do not simply reflect the increased task demands associated with the concurrent vowel production task. One interpretation of these results is that the categorization of smile sincerity benefits from a top-down contribution from the motor system, one that is not recruited by categorization of facial gender. Consequently, the concurrent motor load impairs sensitivity to smile sincerity, but not facial gender. However, another possibility is that vowel production induces an executive load that is not present in the gender judgment task. It is well known that the sight of actions (e.g., Heyes,

2011) and expressions (e.g., Sato & Yoshikawa, 2007) primes imitative responding. When judging smile sincerity, the requirement to produce vowel sounds is in direct conflict with the tendency to imitate. Because the motor programs to produce a vowel and imitate the smile cannot be discharged simultaneously, an executive process must intervene to resolve the response competition. When judging expression-neutral gender stimuli, however, there is no response competition, and consequently no executive load. The differential effects of vowel production on attribution of smile sincerity and facial gender might reflect the respective presence and absence of this executive load, rather than an expression-specific motor contribution to perception.

Experiment 5 sought to test this executive load account of the effect of vowel production on sincerity judgments. Should the increase in noise seen in Experiments 2 and 3 reflect executive load, a similar decrement should be seen when judging the gender of smiling faces. Importantly, vowel production in the presence of a smiling face induces response competition and consequently an executive load – the conflict between vowel production and facial imitation must be resolved.

*Methods*

Twenty-four healthy adults (9 males, $M_{age}$ = 27.46, 2 left handed) participated in Experiment 5. Participants were required to judge the gender of whole-face facial stimuli drawn from a continuum blending the same male and female faces used in Experiment 4. However, unlike the continuum used in Experiment 4, both of the morphed faces exhibited happy expressions (Figure 1c). Psychometric functions were again modelled under three viewing conditions: baseline, production of /i/ and production of /u/. The order in which participants completed the conditions was fully counterbalanced. With the exception of the morph stimuli, the methods were identical to those employed in Experiment 4.

*Results and discussion*

Analysis of the response latencies indicated that the vowel production task was again performed well, with 99.1% of speech sounds produced within ±600 ms of the onset of the stimulus (Figure 4b). This level of performance is comparable with that seen in Experiments 2-4. Noise and PSE estimates were analyzed using ANOVA, with condition as a within-subjects factor. As in Experiment 4, the analysis revealed no effect of

condition on noise estimates [$F(2,46) = .16$, $p = .852$, $\eta^2 = .007$], indicating that categorization precision was comparable in the baseline ($M = 9.23\%$, $SD = 4.78\%$), produce /i/ ($M = 9.46\%$, $SD = 7.58\%$) and produce /u/ ($M = 9.82\%$, $SD = 5.67\%$) conditions. PSE estimates were also similar in the baseline ($M = 46.97$, $SD = 4.47$), produce /i/ and produce /u/ ($M = 47.00$, $SD = 5.77$) conditions ($M = 46.91$, $SD = 8.34$), indicating similar degrees of bias [$F(2,46) = .002$, $p = .998$, $\eta^2 = .000$]. The fact that vowel production failed to interfere with attribution of facial gender, irrespective of the presence of a smiling expression, argues against an executive load account of the motor interference effect observed in Experiments 2 and 3.

Finally, we compared noise estimates measured in the baseline conditions across Experiments 2-5. Should the judgments of facial gender be easier than judgments of smile sincerity, the additional task demands (Experiment 4), or the additional executive load (Experiment 5) induced by vowel production, might be sufficient to interfere with sincerity judgments, but insufficient to impair judgments of facial gender. Importantly, however, ANOVA with Experiment (2-5) as a between-subjects factor confirmed that the noise estimates obtained in the baseline conditions did not vary significantly [$F(3,92) = 1.01$, $p = .394$, $\eta^2 = .032$], confirming that the gender and smile sincerity tasks were of comparable difficulty.

**General discussion**

The present study introduces a novel paradigm for investigating motor contributions to the perception of facial expressions, one that adopts a psychophysical approach, permitting disambiguation of decrements associated with internal noise (e.g. associated with perception and interpretation) from response bias. Having confirmed that observers' judgments of smile sincerity were based on cues from both the eye and mouth regions (Experiment 1), we went on to show that the presence of a concurrent motor load, induced by vowel production, causes a decrease in the precision with which smiles are categorized; responses varied less as a function of stimulus level (Experiment 2). In Experiment 3 we replicated this effect when observers were required to produce vowels, but not when they passively listened to the same vowel sounds. In Experiments 4 and 5 we found that similar judgments of facial gender were unaffected by vowel production, irrespective of the presence of a smiling expression.

Previous studies have reported that blocking facial mimicry alters expression labeling (Niedenthal et al., 2001; Oberman et al., 2007; Ponari et al., 2012; Stel & van Knippenberg, 2008) and rating performance (Maringer et al., 2011; Rychlowska et al., 2014). Remarkably, however, this is the first study to demonstrate i) that a concurrent motor task increases the level of internal noise present during categorization of facial expressions, and ii) that reduction in precision does not extend to other types of facial judgment. Crucially, our novel psychophysical paradigm enabled us to exclude the possibility that motor-induced performance decrements reflect changes in response bias. These findings are a timely contribution in light of high-profile challenges to the existing evidence base for motor theories (Caramazza et al., 2014; Hickok, 2014).

The increase in internal noise induced by vowel production is consistent with simulation accounts of expression perception (Goldman & Sripada, 2005; Niedenthal et al., 2011; Vitale et al., 2014). These models propose that the motor system exerts a top-down influence on perception, useful when interpreting ambiguous expressions. This contribution may be mediated by a reverse simulation process whereby covert mimicry of observed facial expressions aids interpretation by activating motor, somatosensory and affective states associated with observed expressions (Goldman & Sripada, 2005; Niedenthal et al., 2011). Alternatively, observers may form an initial hypothesis about the emotion conveyed through a visual analysis, which is subsequently tested via covert simulation; for example, the anticipated sensory consequences of the hypothesized state may be compared with feedback from the simulation (see Vitale et al., 2014). Importantly, both reverse simulation and generate-and-test models predict that a concurrent motor load ought to result in reduced precision (e.g. greater internal noise), without introducing a systematic perceptual or decision bias.

An alternative account of the precision decrement seen in Experiments 2 and 3 posits that afferent auditory feedback, and not the motor load itself, induced the loss of sensitivity. Importantly, however, experiencing vowel sounds passively did not interfere with smile categorization (Experiment 3). This finding, together with the fact that predicted sensory consequences of actions are thought to be less salient than unpredicted events triggered exogenously (e.g., Brown, Adams, Parees, Edwards, & Friston, 2013), argues against an afferent auditory feedback account of the perceptual decrement. It is also important to

remember that the motor system may contribute to perception through various mechanisms; for example, by modulating processing in low-level visual areas (including V1 and Lateral Geniculate Nucleus; Erisken et al., 2014), by directing the allocation of attention (Bekkering & Neggers, 2002), and by aiding mental rotation (Wexler, Kosslyn, & Berthoz, 1998). Moreover, concurrent vowel production can impair performance through generic distraction (e.g., caused by the additional task demands) and executive load. However, the fact that vowel production did not induce perceptual decrements when judging facial gender (Experiment 4), irrespective of the presence of a smiling expression (Experiments 5), suggests that the interference effect was not a product of low-level sensory modulation, attentional allocation, mental rotation, generic distraction or executive load.

We speculate that correlated sensorimotor experience may be necessary for the emergence of motor contributions to expression perception (Cook, Bird, Catmur, Press, & Heyes, 2014; Press & Cook, 2015). In the typically developing population, individuals experience a wealth of correlated 'seeing' and 'doing' during ontogeny; for example, performing a smile or frown frequently predicts the sight of smiling or frowning interactants, respectively. Following this kind of contingent sensorimotor experience, motor programs responsible for expression planning and execution may come to excite predicted visual consequences. Interestingly, perception of arbitrary stimuli including Gabor patches (Cardoso-Leite, Mamassian, Schültz-Bosbach, & Waszak, 2010) and houses (Kühn, Seurinck, Fias, & Waszak, 2010) can be influenced by motor contributions if paired contingently with action performance during sensorimotor training. In Experiments 4 and 5, vowel production had no influence on judgments of facial gender suggesting that this perceptual decision does not ordinarily recruit motor processes. However, whether judgments of facial gender have the potential to recruit motor contributions to perception, conceptually similar to those recruited by smile judgments, remains an open empirical question (Press & Cook, 2015).

When healthy adult observers experience a concurrent motor load, the precision with which they categorized the smile morphs was reduced. This finding is seemingly inconsistent with previous reports that individuals with Möbius syndrome, a congenital condition associated with partial or complete paralysis of the facial muscles, show unimpaired facial expression recognition (Calder, Keane et al., 2000; Rives Bogart &

Matsumoto, 2009), or non-specific visual deficits (Bate et al., 2013). Drawing conclusions about typical cognitive functioning from atypically developing populations is notoriously difficult (Karmiloff-Smith, 1998). It is possible that this apparent inconsistency reflects the use of insensitive tasks that were unable to detect subtle differences in the expression recognition abilities of Möbius patients. Highly ambiguous judgments, such as categorizing smiles with similar configurations, may be more likely to detect modulations in top-down influence, than tasks requiring the categorization of stereotypical basic emotions (Bate et al., 2013; Calder, Keane et al., 2000; Rives Bogart & Matsumoto, 2009). Whether the loss of precision observed here extends to other types of expression categorizations remains an open empirical question. Alternatively, expression recognition in individuals with Möbius syndrome may be achieved via a qualitatively different route, drawing on compensatory processes. For example, individuals born with Möbius syndrome, unlike members of the typically developing population, do not experience contingencies between expression production and expression observation, and may therefore develop other sources of top-down modulation.

In summary, the present study reports important new evidence for motor contributions to the perception of facial expressions. Using a novel psychophysical paradigm we have shown that inducing a concurrent motor load through vowel production causes a loss of perceptual sensitivity, without introducing response bias, when healthy adult observers judge smile sincerity. The perceptual decrement is not a product of afferent auditory feedback and does not extend to similar judgments of facial gender. These findings are consistent with models proposing that the motor system makes a causal contribution to the perception and interpretation of facial expressions (Goldman & Sripada, 2005; Niedenthal et al., 2011; Vitale et al., 2014).

**References**


Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" Test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry, 42*(2), 241-251.

Bate, S., Cook, S. J., Mole, J., & Cole, J. (2013). First report of generalized face processing difficulties in mobius sequence. *PLoS One, 8*(4), e62656.

Bekkering, H., & Neggers, S. (2002). Visual search is modulated by action intentions. *Psychological Science, 13*(4), 370-374.

Bird, G., & Cook, R. (2013). Mixed emotions: the contribution of alexithymia to the emotional symptoms of autism. *Translational Psychiatry, 3*, e285.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*(4), 433-436.

Brown, H., Adams, R. A., Parees, I., Edwards, M., & Friston, K. (2013). Active inference, sensory attenuation and illusions. *Cognitive Processing, 14*(4), 411-427.

Calder, A. J., Keane, J., Cole, J., Campbell, R., & Young, A. W. (2000). Facial expression recognition by people with mobius syndrome. *Cognitive Neuropsychology, 17*(1), 73-87.

Calder, A. J., Young, A. W., Keane, J., & Dean, M. (2000). Configural information in facial expression perception. *Journal of Experimental Psychology: Human Perception Performance, 26*(2), 527-551.

Caramazza, A., Anzellotti, S., Strnad, L., & Lingnau, A. (2014). Embodied cognition and mirror neurons: a critical assessment. *Annual Review of Neuroscience, 37*, 1-15.

Cardoso-Leite, P., Mamassian, P., Schültz-Bosbach, S., & Waszak, F. (2010). A new look at sensory attenuation action-effect anticipation affects sensitivity, not response bias. *Psychological Science, 21*(12), 1740-1745.

Cellerino, A., Borghetti, D., & Sartucci, F. (2004). Sex differences in face gender recognition in humans. *Brain Research Bulletin, 63*(6), 443-449.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences (Second Edition)*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Cook, R., Bird, G., Catmur, C., Press, C., & Heyes, C. M. (2014). Mirror neurons: from origin to function. *Behavioral and Brain Sciences, 37*(2), 177-192.

Dapretto, M., Davies, M. S., Pfeifer, J. H., Scott, A. A., Sigman, M., Bookheimer, S. Y., et al. (2006). Understanding emotions in others: mirror neuron dysfunction in children with autism spectrum disorders. *Nature Neuroscience, 9*(1), 28-30.

Del Giudice, M., & Colle, L. (2007). Differences between children and adults in the recognition of enjoyment smiles. *Developmental Psychology, 43* (3), 796-803.

Duchenne de Boulogne, G. B. (1862). *The mechanism of human facial expression*. Paris: Jules Renard.

Erisken, S., Vaiceliunaite, A., Jurjut, O., Fiorini, M., Katzner, S., & Busse, L. (2014). Effects of locomotion extend throughout the mouse early visual system. *Current Biology, 24*(24), 2899-2907.

Ferrari, P. F., Gallese, V., Rizzolatti, G., & Fogassi, L. (2003). Mirror neurons responding to the observation of ingestive and communicative mouth actions in the monkey ventral premotor cortex. *European Journal of Neuroscience, 17*(8), 1703-1714.

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B- Biological Sciences, 360*(1456), 815-836.

Gallese, V., & Sinigaglia, C. (2011). What is so special about embodied simulation? *Trends in Cognitive Sciences, 15*(11), 512-519.

Gilbert, C. D., & Li, W. (2013). Top-down influences on visual processing. *Nature Reviews Neuroscience, 14*(5), 350-363.

Goldman, A. I., & Sripada, C. S. (2005). Simulationist models of face-based emotion recognition. *Cognition, 94*(3), 193-213.

Green, D. M., and Swets, J. A. (1966). Signal Detection Theory and Psychophysics. New York, NY: Wiley.

Gregory, R. L. (1997). Knowledge in perception and illusion. *Philosophical Transactions of the Royal Society B- Biological Sciences, 352*(1358), 1121-1127.

Hamilton, A. F. (2012). Reflecting on the mirror neuron system in autism: a systematic review of current theories. *Developmental Cognitive Neuroscience, 3*, 91-105.

Heyes, C. M. (2011). Automatic imitation. *Psychological Bulletin, 137*(3), 463-483.

Hickok, G. (2014). *The Myth of Mirror Neurons: The Real Neuroscience of Communication and Cognition*. New York: W. W. Norton & Company, Inc.

Jack, R. E., Garrod, O. G. B., & Schyns, P. G. (2014). Dynamic facial expressions of emotion transmit an evolving hierarchy of signals over time. *Current Biology, 24*(2), 187-192.

Karmiloff-Smith, A. (1998). Development itself is the key to understanding developmental disorders. *Trends in Cognitive Sciences, 2*(10), 389-398.

Kilner, J. M. (2011). More than one pathway to action understanding. *Trends in Cognitive Sciences, 15*(8), 352-357.

Kühn, S., Seurinck, R., Fias, W., & Waszak, F. (2010). The internal anticipation of sensory action effects: when action induces FFA and PPA activity. *Frontiers in Human Neuroscience, 4*.

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S. T., & van Knippenberg, A. (2010). Presentation and validation of the Radboud Faces Database. *Cognition & Emotion, 24*(8), 1377-1388.

Maringer, M., Krumhuber, E. G., Fischer, A. H., & Niedenthal, P. M. (2011). Beyond smile dynamics: mimicry and beliefs in judgments of smiles. *Emotion, 11*(1), 181-187.

Marneweck, M., Palermo, R., & Hammond, G. (2014). Discrimination and recognition of facial expressions of emotion and their links with voluntary control of facial musculature in Parkinson's disease. *Neuropsychology, 28*(6), 917-928.

Neal, D. T., & Chartrand, T. L. (2011). Embodied emotion perception amplifying and dampening facial feedback modulates emotion perception accuracy. *Social Psychological and Personality Science, 2*(6), 673-678.

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 & Emotion, 15* (6), 853-864.

Niedenthal, P. M., Mermillod, M., Maringer, M., & Hess, U. (2011). The Simulation of Smiles (SIMS) model: Embodied simulation and the meaning of facial expression. *Behavioral and Brain Sciences, 33*(6), 417-433.

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.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision, 10*(4), 437-442.

Pistoia, F., Conson, M., Trojano, L., Grossi, D., Ponari, M., Colonnese, C., et al. (2010). Impaired conscious recognition of negative facial expressions in patients with locked-in syndrome. *Journal of Neuroscience, 30*(23), 7838-7844.

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*(6), 1398-1403.

Press, C., & Cook, R. (2015). Beyond action-specific simulation: domain-general motor contributions to perception. *Trends in Cognitive Sciences, 19*(4), 176-178.

Prins, N., & Kingdom, F. A. A. (2009). Palamedes: Matlab routines for analyzing psychophysical data. . *http://www.palamedestoolbox.org*.

Rives Bogart, K., & Matsumoto, D. (2009). Facial mimicry is not necessary to recognize emotion: Facial expression recognition by people with Moebius syndrome. *Social Neuroscience, 5*(2), 241-251.

Rizzolatti, G., & Luppino, G. (2001). The cortical motor system. *Neuron, 31*(6), 889-901.

Rychlowska, M., Canadas, E., Wood, A., Krumhuber, E. G., Fischer, A., & Niedenthal, P. M. (2014). Blocking mimicry makes true and false smiles look the same. *PLoS One, 9*(3), e90876.

Sato, W., & Yoshikawa, S. (2007). Spontaneous facial mimicry in response to dynamic facial expressions. *Cognition, 104*(1), 1-18.

Schütz-Bosbach, S., & Prinz, W. (2007). Perceptual resonance: action-induced modulation of perception. *Trends in Cognitive Sciences, 11*(8), 349-355.

Stel, M., & van Knippenberg, A. (2008). The role of facial mimicry in the recognition of affect. *Psychological Science, 19*(10), 984-985.

Vitale, J., Williams, M., Johnston, B., & Boccignone, G. (2014). Affective facial expression processing via simulation: A probabilistic model. *Biologically Inspired Cognitive Architectures, 10*, 30-41.

Wexler, M., Kosslyn, S. M., & Berthoz, A. (1998). Motor processes in mental rotation. *Cognition, 68*(1), 77-94.

Wilson, M., & Knoblich, G. (2005). The case for motor involvement in perceiving conspecifics. *Psychological Bulletin, 131*(3), 460-473.

**Acknowledgements**

The research reported in this article was supported by a Future Research Leaders grant (ES/K008226/1) awarded by the Economic and Social Research Council (ESRC) to R.C. The authors gratefully acknowledge the help of Miss Jennifer Murphy who assisted in the analysis of the vocal response latencies.

**Footnotes**

[1] When the concurrent motor task is unrelated to the perceptual judgment (e.g., performing vowel sounds when judging smile sincerity), existing motor theories predict a loss of categorization precision. When concurrent motor tasks are closely-related to the judgment (e.g., performing an insincere smile when judging smile sincerity), these models predict perceptual bias.

[2] This treatment reflects the nature of the attribution task; i.e., participants judged whether stimuli belong to one of two natural categories (sincere or insincere smiles). This is a departure from a standard signal detection task whereby a signal is absent on some trials and present in others. In the latter case it is common to estimate false alarms in the absence of a stimulus and the proportion of hits at varying levels of stimulus intensity.

[3] The suggestion that a top-down contribution can selectively modulate perception of the eye or mouth regions, is at odds with the finding that expressions are typically defined by correlated eye and mouth variation (Jack et al., 2014), and that expressions are represented holistically (Calder, Young et al., 2000).


*Figure 1:* (a) The smile stimuli used in Experiments 1-3, morphing incrementally from 20% sincere (left) to 80% sincere (right). The morph continuum was created using images taken from the Smile Picture Set (Del Giudice & Colle, 2007). (b) The gender stimuli used in Experiment 4 morphing incrementally from 20% male (left) to 80% male (right). (c) The gender stimuli used in Experiment 5 morphing incrementally from 20% male (left) to 80% male (right). The morph continua from Experiment 4 and 5 were created using images taken from the Radboud Faces Database (Langner et al., 2010).

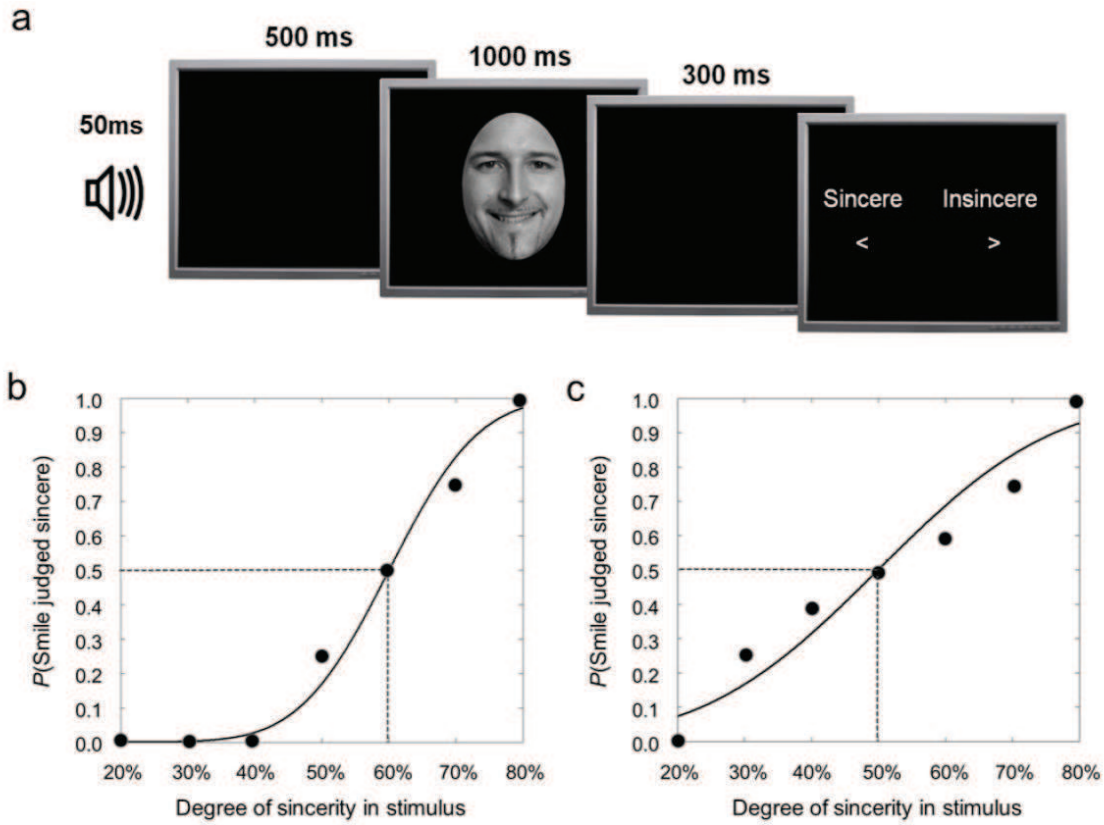


*Figure 2:* (a) The arrangement of experimental trials. In Experiment 1 the tone was task-irrelevant. In Experiments 2-5, observers were required to produce a vowel sound, either /u/ or /i/, cued by the tone. In Experiment 3, observers also passively listened to, a vowel (/i/) played through speakers after the tone. (a) Illustration of bias. This observer is relatively sensitive to physical changes in the stimulus but is prone to making "insincere" responses. (b) Illustration of insensitivity. This observer exhibits no bias, but their responses vary less with physical changes in stimulus strength.

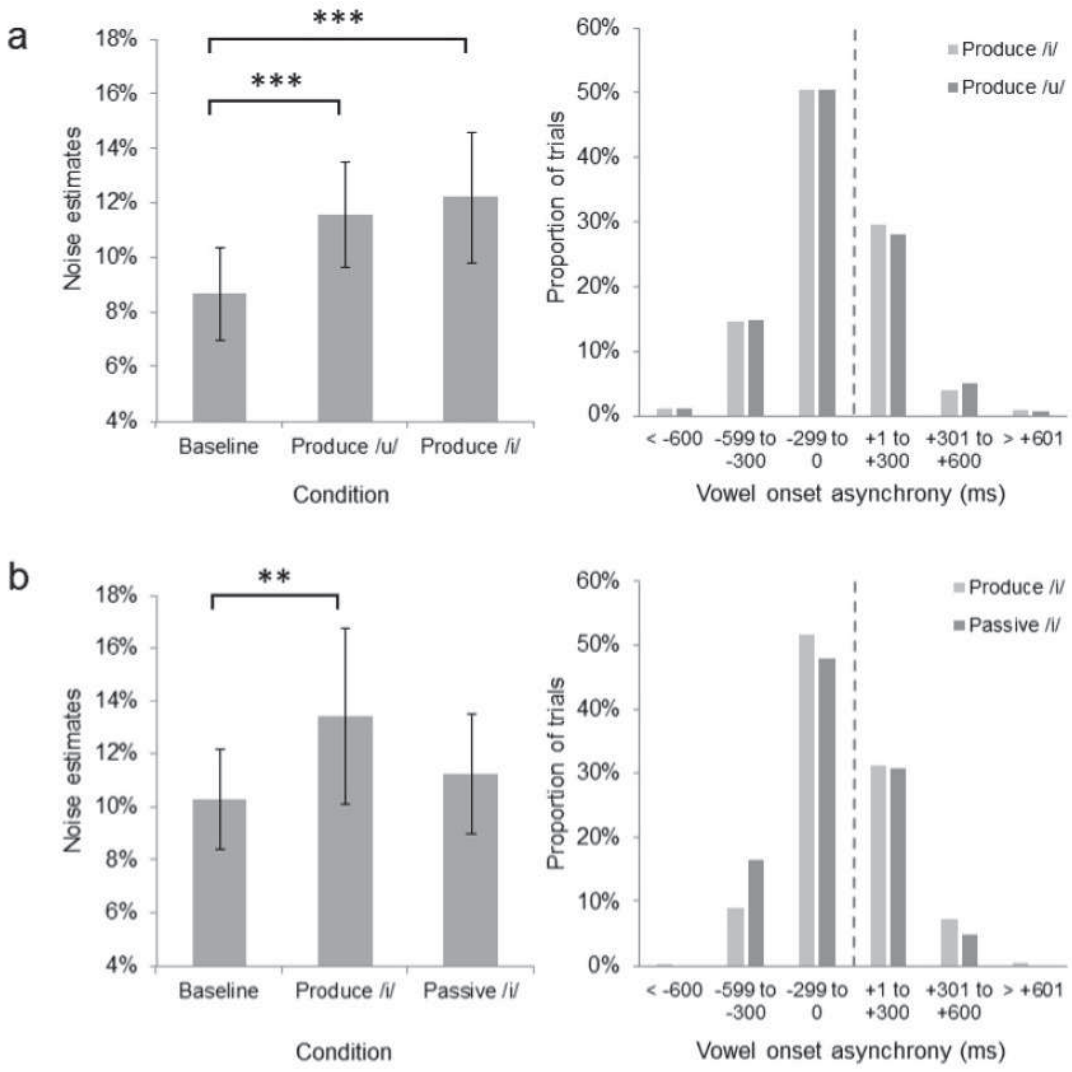


*Figure 3:* (a) The noise estimates and vowel onset asynchronies from Experiment 2. (b) The noise estimates and vowel onset asynchronies from Experiment 3. Negative asynchronies indicate that vowel production commenced before the onset of the image. Error bars represent 95% Confidence Intervals. *** denotes $p < .01$, ** denotes $p < .025$.

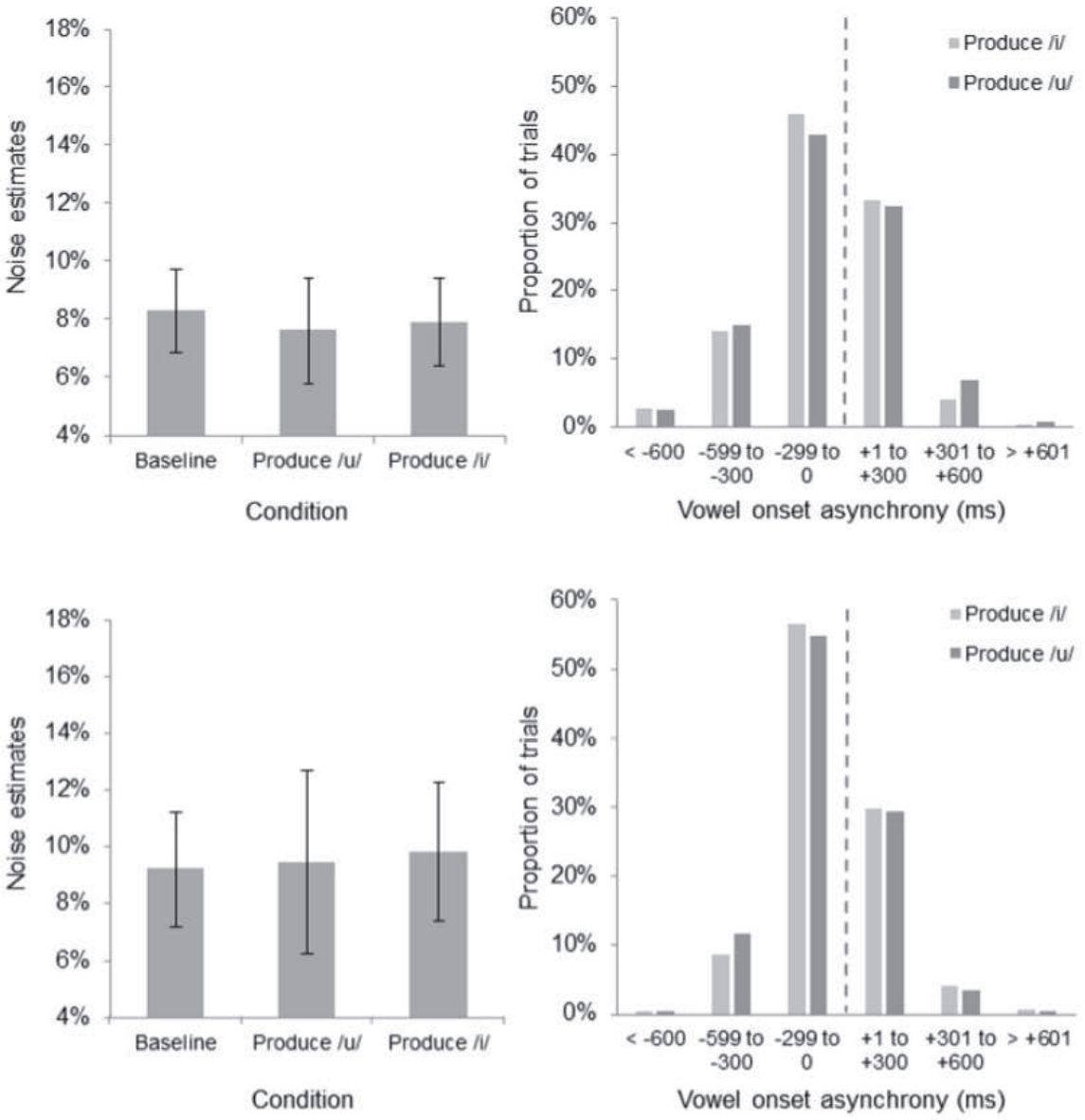


*Figure 4:* (a) The noise estimates and vowel onset asynchronies from Experiment 4. (b) The noise estimates and vowel onset asynchronies from Experiment 5. Negative asynchronies indicate that vowel production commenced before the onset of the image. Error bars represent 95% Confidence Intervals.

**Table 1**

Table 1: Mean proportion of sincere responses to each stimulus level, in each condition, for Experiments 1-3. Standard deviations are shown in italics inside parentheses.

| | | Smile morph level (% sincere) | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 20% | | 30% | | 40% | | 50% | | 60% | | 70% | | 80% | |
| **Experiment 1** | Whole-face | 2.7% | *(1.2)* | 2.3% | *(0.7)* | 22.3% | *(3.3)* | 69.4% | *(4.2)* | 90.2% | *(2.7)* | 95.4% | *(1.4)* | 98.1% | *(0.7)* |
| | Eyes-only | 4.0% | *(1.4)* | 6.9% | *(1.8)* | 21.5% | *(2.9)* | 61.0% | *(5.3)* | 82.9% | *(3.3)* | 93.8% | *(1.6)* | 95.0% | *(1.2)* |
| | Mouth-only | 8.8% | *(2.0)* | 16.7% | *(3.1)* | 38.5% | *(4.9)* | 61.5% | *(5.7)* | 78.8% | *(3.9)* | 88.3% | *(3.2)* | 95.0% | *(1.2)* |
| **Experiment 2** | Baseline | 1.5% | *(0.7)* | 5.8% | *(2.6)* | 20.2% | *(4.4)* | 58.5% | *(6.4)* | 83.8% | *(4.1)* | 95.0% | *(2.3)* | 97.5% | *(1.5)* |
| | Produce /u/ | 2.7% | *(1.4)* | 6.7% | *(2.0)* | 22.9% | *(5.0)* | 57.1% | *(5.0)* | 75.8% | *(5.0)* | 91.7% | *(2.1)* | 95.2% | *(2.1)* |
| | Produce /i/ | 3.1% | *(1.2)* | 8.8% | *(3.4)* | 22.1% | *(4.4)* | 52.9% | *(6.0)* | 80.0% | *(4.7)* | 90.8% | *(2.6)* | 95.2% | *(2.1)* |
| **Experiment 3** | Baseline | 1.7% | *(0.8)* | 5.6% | *(1.8)* | 24.4% | *(4.5)* | 64.0% | *(4.7)* | 89.0% | *(2.3)* | 96.5% | *(1.5)* | 96.5% | *(1.7)* |
| | Produce /i/ | 5.4% | *(2.3)* | 7.3% | *(2.8)* | 26.7% | *(4.5)* | 65.2% | *(3.9)* | 88.3% | *(2.3)* | 92.1% | *(2.0)* | 95.2% | *(1.7)* |
| | Passive /i/ | 2.5% | *(2.1)* | 7.1% | *(3.7)* | 23.8% | *(5.4)* | 59.8% | *(4.4)* | 84.2% | *(2.7)* | 96.9% | *(1.0)* | 97.3% | *(1.0)* |

*Table 1*: Mean proportion of sincere responses to each stimulus level, in each condition, for Experiments 1-3. Standard deviations are shown in italics inside parentheses.

**Table 2**

Table 2: Mean proportion of male responses to each stimulus level, in each condition, for Experiments 4-5. Standard deviations are shown in italics inside parentheses.

| | | Gender morph level (% male) | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 20% | | 30% | | 40% | | 50% | | 60% | | 70% | | 80% | |
| **Experiment 4** | Baseline | 1.88% | *(0.79)* | 5.00% | *(1.53)* | 27.50% | *(3.73)* | 80.42% | *(4.40)* | 95.21% | *(1.69)* | 98.13% | *(0.79)* | 98.54% | *(0.64)* |
| | Produce /u/ | 2.50% | *(0.85)* | 2.71% | *(1.13)* | 26.88% | *(3.44)* | 80.21% | *(4.65)* | 96.25% | *(1.25)* | 98.33% | *(0.83)* | 98.54% | *(0.70)* |
| | Produce /i/ | 2.29% | *(0.80)* | 2.50% | *(1.09)* | 21.67% | *(3.99)* | 72.50% | *(5.90)* | 91.25% | *(2.95)* | 97.29% | *(1.75)* | 97.92% | *(0.99)* |
| **Experiment 5** | Baseline | 2.08% | *(0.85)* | 3.75% | *(1.10)* | 20.00% | *(3.19)* | 67.71% | *(3.90)* | 90.83% | *(1.99)* | 97.08% | *(1.16)* | 98.54% | *(0.88)* |
| | Produce /u/ | 3.13% | *(1.85)* | 6.25% | *(2.76)* | 21.88% | *(4.04)* | 67.92% | *(5.01)* | 87.08% | *(3.90)* | 95.00% | *(1.83)* | 97.50% | *(1.00)* |
| | Produce /i/ | 2.71% | *(1.68)* | 3.75% | *(2.00)* | 18.96% | *(3.72)* | 66.04% | *(5.04)* | 92.50% | *(2.15)* | 96.88% | *(1.20)* | 98.13% | *(1.03)* |

*Table 2*: Mean proportion of male responses to each stimulus level, in each condition, for Experiments 4-5. Standard deviations are shown in italics inside parentheses.