Can a word sound like a shape before you have seen it?

Sound-shape mapping prior to conscious awareness.

Supplementary materials

1. Enlarged subfigures of Fig. 4
2. A brief review on CFS studies
3. Supplementary analyses of Experiment 3
Fig. S1. Enlarged figure: Individual participants’ suppression times (circles, left y-axis) in the two conditions (congruent, incongruent) of Experiment 1. From left to right, the participants are ordered according to the strength of the effect. Error bars represent SEM. The bars show the difference in suppression times (right y-axis) between the conditions.
**Fig. S2.** Enlarged figure: Individual participants’ suppression times (circles, left y-axis) in the two conditions (congruent, incongruent) of Experiment 2a. From left to right, the participants are ordered according to the strength of the effect. Error bars represent SEM. The bars show the difference in suppression times (right y-axis) between the conditions.
Fig. S3. Enlarged figure: Individual participants’ suppression times (circles, left y-axis) in the two conditions (congruent, incongruent) of Experiment 2b. From left to right, the participants are ordered according to the strength of the effect. Error bars represent SEM. The bars show the difference in suppression times (right y-axis) between the conditions.
2. A brief review on CFS studies

How do our present results from Experiments 1 and 2 and compare to other studies on written word processing under CFS suppression? Familiar word forms have been shown to break suppression faster than unfamiliar word forms, with speakers of Chinese showing a bias for Chinese letters over Hebrew, and vice versa for speakers of Hebrew (Jiang et al., 2007). The bias for word form familiarity is visually detailed, with Chinese speakers showing an advantage for correctly presented characters, over the same characters reversed or re-organized (Yang & Yeh, 2011), implying enhanced processing for familiar configurations of visual elements. Words and expressions with emotional valence have also been shown to influence suppression time (e.g. word: Yang & Yeh, 2011; face: Yang, Zald, & Blake, 2007), although the directions of the effects are somewhat mixed (Gayet et al., 2014). More controversially, semantically incongruent written statements (e.g., “I ironed coffee”) have been shown to break suppression faster than semantically congruent statements (Sklar et al., 2012), suggesting that familiar word forms are not only processed efficiently, but that semantic integration between words is under way before the stimulus is reported as consciously visible. Similarly, we have recently shown that syntactic congruence of written words can be processed unconsciously in the absence of semantics (Hung & Hsieh, 2015). Although questions remain about the interpretation of some of these findings (Gayet et al., 2014), they suggest that the visual system exhibits some level of automated word form decoding while visually presented words are under (or perhaps emerging from) CFS-induced suppression.

These studies of written-word processing align with non-linguistic studies of face-processing under CFS, which suggest that visual familiarity with a stimulus type (e.g.,
upright versus inverted faces: Jiang et al., 2007), or particular stimuli (familiar faces versus strangers: Gobbini et al., 2013), enhances the speed of breakthrough. It is clear that some level of configural information processing is automatic for stimuli under suppression (see Stein & Sterzer (2012) in which the authors point out that face configural properties could drive some unconscious emotion effect under CFS.). By contrast, studies measuring suppression times for photographs with incongruent semantic content (e.g., a picture of someone putting a chess-board in an oven), have shown that semantic incongruence in photos, as in written phrases, breaks suppression faster than congruence (Mudrik, Breska, Lamy, & Deouell, 2011). However, a recent study failed to find conclusive evidence for this level of integration without awareness (Moors, Boelens, van Overwalle, & Wagemans, 2016). Further experiments are required to settle the disagreement. Taken together, these studies suggest that familiar stimuli like words and faces are subject to a certain degree of automatic processing, which results in suppression being broken faster for more-familiar stimuli (e.g., a familiar script, upright human faces, common scenes), but for more complex stimuli (multiword phrases, photographs of complex scenes) the same processes result in integration, and in these cases, incongruence ‘pops out’ faster. In our Experiments 1 and 2, it is the congruent combinations of written words with shape outlines which break suppression faster, not the incongruent ones. Our findings therefore align better with familiarity effects, than with high-level integration effects.

In an alternative line of primed CFS, consciously presented linguistic cues have been shown to ‘prime’ faster response time and higher identification accuracy of an unseen object when audio labels (e.g., pumpkin) match the identity of a suppressed visual stimulus (Lupyan & Ward, 2013). Also, consciously presented written words prime
faster breakthrough for suppressed written words which share semantic association or partial orthographical overlap (e.g., *sock* primes both *shoe* and *shock* but not *tape*: Costello, Jiang, Baartman, McGlennen, & He, 2009). In a similar vein, our findings on unconscious syntactic processing show that consciously perceived words, which built up a syntactic context/expectation, led to shorter suppression time of the following incongruent target word, even when meaning was depleted by using pseudowords (Hung & Hsieh, 2015). These findings demonstrate that linguistic information can be integrated into the visual recognition under CFS, and that when linguistic cues are presented prior to suppressed stimuli, the phonological/semantic/syntactic associations facilitate on visual recognition. One possible account as noted by Lupyan and Ward (2013), interactive activation models of cognition allow higher-level conceptual information to influence the visual processing system via top-down ‘re-tuning’, or predictive processing, which may include selective pre-activation of diagnostic visual features, or other attentional mechanisms involved in the breakthrough process (e.g., following the word ‘pumpkin’, visual expectations about the visual features of a to-be-seen pumpkin). It is therefore clear that consciously presented linguistic primes can trigger activity in established pathways which result in recognition of familiar visual forms.

It was formerly suggested that bCFS (breaking CFS) could be used to measure the duration of unconscious processing under CFS (Jiang et al., 2007). However, as the main measure – time to respond – includes both the time taken to detect the stimulus, and the time taken to press the button, it is not possible to tell whether the paradigm measures unconscious processing time, or the time taken for a stimulus to transitions from fully unconscious processing to concrete, conscious awareness, or even a
reaction time enhancement due to higher levels of certainty during the transition from unconscious to conscious processing. Thus, bCFS may not necessarily provide direct information about the duration of unconscious processing (Gayet et al., 2014; Stein & Sterzer, 2014; Stein, Hebart, & Sterzer, 2011; Sterzer, Stein, Ludwig, Rothkirch, & Hesselmann, 2014; Yang, Brascamp, Kang, & Blake, 2014). Instead, bCFS is thought to index stimulus detectability, or access to awareness (Stein, End, & Sterzer, 2014; Stein, Seymour, Hebart, & Sterzer, 2014). In other words, time-to-break suppression may reflect the efficiency of the transition across the liminal zone from fully unconscious processing to conscious awareness of a visual stimulus. Given the possible confound between time-to-consciousness and time-to-respond in the breaking CFS paradigm, Experiment 3 is an important addition: The dependent measure here is not the time to report visibility, but whether or not the stimulus was visible at all. Thus we are confident that the lowering of the visibility threshold indicates multisensory processing in the absence of conscious awareness, the congruence of which allows a sub-threshold visual stimulus to achieve visibility. In addition, we identified that this congruence effect was only evident when the audio was presented prior to the onset of the visual stimulus, suggesting that auditory processing requires some lead-time before the phonological representation can be integrated with a very briefly presented visual shape.
3. Supplementary analyses of Experiment 3

a. Blocking of congruent and incongruent conditions

In Experiment 3, we blocked the congruent and incongruent conditions to minimize inter-trial response variability, and one may worry that participants would build an expectation according to what they heard. Results of the post-study briefing session of Experiment 3 showed that all participants reported themselves naïve to the relationship between the sound and shape, suggesting no prior knowledge of the pairing (p. 29 in Discussion). We have also checked and did not find the order effect of the block on the main interaction finding (Congruence x SOA interaction, CON first: (F(2,18) = 2.79, p = .09, $\eta_p^2 = .24$; InCON first: (F(2,18) = 2.75, p = .09, $\eta_p^2 = .23$). Both orders show the trend of the main effect but were not significant, possibly due to a lack of statistical power.

b. Results between the bottom-to-top and top-to-bottom staircases

Possibly due to lack of statistical power, the statistical result for each individual staircase was not significant when analyzed separately (bottom-to-top: F(2,38) = 2.79, $p = .07$, $\eta_p^2 = .13$; top-to-bottom: F(2,38) = 1.78, $p = .18$, $\eta_p^2 = .09$), and the top-to-bottom staircase seemed to be noisier. Since there are several possible reasons (e.g. lack of data), we prefer not to make any conclusion here.

c. Accuracy on unseen trials was not crucial for the key finding.
To test whether accuracy on unseen trials affects the key finding (lower visibility threshold when a congruent sound was presented 150 ms prior to the visual shape), we split the participants to a low accuracy group (n = 10, mean accuracy = 51.9%) and a high accuracy group (n = 10, mean accuracy = 61.8%) according to their performance on the unseen trials. Interestingly, the low accuracy group \( t_{\text{paired}}(9) = 2.83, p = .02 \) but not the high accuracy group \( t_{\text{paired}}(9) = 0.84, p = .42 \) exhibited a significant result, indicating that the key finding in Experiment 3 was not driven by those participants who had above chance-rate performance on the unseen trials.
References


