BRILL *Multisensory Research (2020)* DOI:10.1163/22134808-bja10016 brill.com/msr

MULTISENSORY RESEARCH

Review

Multisensory Integration of Emotion in Schizophrenic Patients

Yi Lin [1], **Hongwei Ding** [1,*] **and Yang Zhang** [2,*]

[1] Speech–Language–Hearing Center, School of Foreign Languages, Shanghai Jiao Tong University, 800 Dong Chuan Rd., Minhang District, Shanghai, 200240, China

[2] Department of Speech–Language–Hearing Sciences & Center for Neurobehavioral Development, University of Minnesota, Twin Cities, MN 55455, USA

Received 15 October 2019; accepted 24 March 2020

Abstract

Multisensory integration (MSI) of emotion has been increasingly recognized as an essential element of schizophrenic patients' impairments, leading to the breakdown of their interpersonal functioning. The present review provides an updated synopsis of schizophrenics' MSI abilities in emotion processing by examining relevant behavioral and neurological research. Existing behavioral studies have adopted well-established experimental paradigms to investigate how participants understand multisensory emotion stimuli, and interpret their reciprocal interactions. Yet it remains controversial with regard to congruence-induced facilitation effects, modality dominance effects, and generalized vs specific impairment hypotheses. Such inconsistencies are likely due to differences and variations in experimental manipulations, participants' clinical symptomatology, and cognitive abilities. Recent electrophysiological and neuroimaging research has revealed aberrant indices in event-related potential (ERP) and brain activation patterns, further suggesting impaired temporal processing and dysfunctional brain regions, connectivity and circuities at different stages of MSI in emotion processing. The limitations of existing studies and implications for future MSI work are discussed in light of research designs and techniques, study samples and stimuli, and clinical applications.

Keywords

Multisensory integration (MSI), emotion, schizophrenia

1. Introduction

Schizophrenia is a severe and lifelong neurodevelopmental disorder that affects over 21 million people in the world (National Institute of Mental Health,

* To whom correspondence should be addressed. E-mails: hwding@sjtu.edu.cn/zhanglab@umn.edu

2016; World Health Organization, 2018). Patients with schizophrenia have been frequently reported to experience interpersonal misunderstanding, generate inappropriate behaviors, and have difficulty in assimilating into society (Brüne, 2005; Green *et al*., 2018; Pawelczyk *et al*., 2018). Core symptoms that appear to disrupt schizophrenics involve dysfunctions in rational thinking, decision making, memory and language, and social cognition (Brüne, 2005; Couture *et al*., 2006; Penn *et al*., 2008).

As a key indicator of how people perceive themselves and others, and construct interpersonal relationships in the social world, social cognition has been of great interest in existing literature on schizophrenia (Green and Leitman, 2008; Green *et al*., 2005; Penn *et al*., 2008; Sterea, 2015). One of the important components of social cognition is emotion processing, which serves as an essential target for early diagnosis and intervention of schizophrenia (Corcoran *et al*., 2015; Varga *et al*., 2018a, b).

It is generally accepted that patients with schizophrenia show aberrant unimodal emotion cognition (Edwards *et al*., 2002), especially deficits in processing visual (Chan *et al*., 2010; Kohler *et al*., 2003, 2010) and auditory (Hoertnagl *et al*., 2014; Kantrowitz *et al*., 2013, 2015; McLachlan *et al*., 2013) emotional signals. However, emotions are often conveyed through multiple communication modalities in natural interpersonal settings, in which people need to combine both linguistic inputs (e.g., words, phrases, and sentences) and paralinguistic cues (e.g., facial expressions, tone of voice, lip movements, eye contact, gestures, and body movements) (Filippi *et al*., 2017; Pawelczyk *et al*., 2018). If one fails to integrate emotional signals from multisensory channels, he/she may misinterpret others' feelings, thoughts and intentions, and even encounter more severe interpersonal conflicts. Therefore, examining schizophrenic patients' abilities of integrating multisensory emotional information has been prioritized in the area of social cognitive research.

Multisensory integration (MSI) refers to a natural and automatic process in which a person utilizes information from multiple sensory channels to make interpretations, decisions and behavioral responses (Calvert *et al*., 2004; Stein *et al*., 2009). For successful social interactions, MSI not only enhances the discrimination of physical properties of the signals (Bertelson and de Gelder, 2004), but also helps clarify the intended meaning of an utterance, indicate speakers' thoughts and emotions, and resolve ambiguity in human speech (Green and Angelaki, 2010; Paulmann and Pell, 2011). During this process, our brain disposes of irrelevant stimuli (e.g., background noise) to integrate signals from separate sensory modalities as a single event in a more efficient way. There has been empirical evidence suggesting that the accuracy rate of a listening comprehension task can be improved by 60% in noisy environments with auditory stimuli and visual aids simultaneously presented, as compared with an auditory-only experimental condition (Erber, 1969; Summerfield, 1979). Similarly, it has been found that MSI is conducive to the daily communication of a number of groups of participants bearing disadvantages in language comprehension and production, including second- or foreign-language learners (Hardison, 2010; Hazan *et al.*, 2005; Zhang *et al.*, 2009), people with hearing impairments (Holt *et al.*, 2011; Moradi *et al.*, 2017; Rao *et al.*, 2017; Yu *et al.*, 2017), and children with language or learning disabilities (Irwin and DiBlasi, 2017; Veuillet *et al.*, 2007). In addition to speech communication, multimodal approaches have been demonstrated to facilitate various research and practical fields ranging from educational, forensic, and financial to engineering domains (Bornik *et al.*, 2018; Egger *et al.*, 2019; Hassett and Curwood, 2009; Mohr *et al.*, 2010; Vilko and Hallikas, 2012).

However, MSI does not always produce a facilitatory effect, and cross-modal biases may occur. One of the underlying attributes lies in the discrepancies, ambiguities, or even conflicts among different sensory stimuli. For instance, Heald and Nusbaum (2014) demonstrated that audiovisual information could slow down responses and distract the observer from the word recognition task when there were two possible talkers. When incongruent signals from multiple sensory modalities are concurrently presented, they tend to interfere with participants' performances, resulting in reduced response accuracy, increased reaction time, and even illusionary perceptual fusions (Bertelson and Radeau, 1981; McGurk and MacDonald, 1976; Tseng *et al.*, 2015). It is also likely that one may not always equally allocate their perceptual and cognitive resources when processing multisensory stimuli, and thus resort to overcoming the bombardment of information by exercising selective attention (de Jong *et al.*, 2010; Koelewijn *et al.*, 2010; Talsma *et al.*, 2010). In this case, one sensory channel or modality may not be so perceptually dominant over the others that it is inhibited during MSI (Spence *et al.*, 2011). Here age and sensory abilities also play important roles. For instance, Sekiyama *et al.* (2014) showed that older adults are slower in response to auditory stimuli and their speech perception is more influenced by visual information than younger adults. Thus, understanding to what extent MSI can produce a facilitatory or inhibitory effect and how the two opposing manifestations can be reconciled across studies requires a careful examination of the differing nature of participants' characteristics and experimental tasks (Mahoney *et al.*, 2011; Sinnett *et al.*, 2008). Such an exploration of the enhancement and inhibition effects can enrich our understanding of how MSI functions as a core ability in complex settings of natural communication (Calvert *et al.*, 2004), and as an effective indicator of the distinctions between diverse healthy and clinical populations (Murray *et al.*, 2018; Stone *et al.*, 2011).

In schizophrenia studies, earlier work demonstrated deficient MSI with emotionally neutral stimuli (de Gelder *et al.*, 2003; Pearl *et al.*, 2009; Ross *et al.*, 2007; Williams *et al.*, 2010). In recent years, a number of researchers have identified impaired integration of multimodal emotion signals in patients with psychotic diseases, including schizophrenia, autism, and bipolar disorder (de Gelder *et al.*, 2005; Feldman *et al.*, 2018; Van Rheenen and Rossell, 2014; Williams *et al.*, 2010). Despite numerous efforts dedicated to the MSI field, research findings have been indeterminate concerning under what circumstances and to what extent schizophrenics show impairments in multisensory processing. For instance, in one study, while patients have been found to exhibit normal cross-modal spatial integration when processing simple tones and light flashes, they experience difficulty in integrating more complex linguistic stimuli (de Gelder *et al.*, 2003). In another study, however, schizophrenic patients demonstrated a deficit in target detection which required integrating a simple tone with a visually presented letter (Williams *et al.*, 2010). It deserves to be explored whether MSI deficits are consistently established in patients' emotion perception, which is an essential component of social cognition. Moreover, very few studies have thoroughly reviewed the nature of potential deficits in multisensory integration of emotions in schizophrenic patients. Specifically, there has been a lack of consensus regarding common research objectives and experimental paradigms, central theories or hypotheses on the patients' emotional MSI patterns, the factors affecting patients' performances, and the neural mechanisms governing the integration process of multimodal emotion information.

The current review provides a comprehensive synopsis of MSI for emotion cognition in schizophrenic patients. By analyzing the relevant behavioral and neurological studies, we aim to (a) summarize the research focuses and corresponding experimental paradigms in extant studies, (b) provide theoretical explanations for patients' deficient MSI of emotion, (c) identify the potential factors influencing patients' MSI performance, and (d) elucidate the temporal dynamics and neuroanatomical correlates of human's processing of multisensory emotion signals. This clinical perspective can help deepen our understanding of the nature of MSI in emotion processing in schizophrenics with respect to perceptual and cognitive impairments that negatively impact their social functioning in today's complex, demanding and information-abundant living environments. The review also provides important insights for future basic research and clinical practice in MSI, which can contribute to the diagnosis, intervention, adaptation and socialization processes for patients with this psychotic disorder.

2. Behavioral Studies

2.1. Research Objectives and Experimental Paradigms

2.1.1. Recognizing Multimodal Emotional Information

As an essential part of emotional intelligence and social perception, the recognition of emotional information in multiple sensory modalities among schizophrenic patients has always been a focal point of research in this field. To this end, a line of studies has adopted emotion recognition tests with high reliability and validity (Bryson *et al.*, 1997; Fiszdon and Bell, 2009; Fiszdon and Johannesen, 2010). Although the experimental tasks have not been consistently named and represented across extant literature (e.g., 'recognition', 'identification', 'discrimination', 'matching', 'naming', and 'labeling'; Trémeau, 2006), these behavioral test protocols generally require participants to consciously recognize the emotional valence of external stimuli.

The Bell–Lysaker Emotion Recognition Task (BLERT) is one of the test protocols frequently employed in existing research. It was developed by Bell *et al.* (1997). The BLERT is an audiovisual emotion recognition task designed to measure participants' ability of interpreting seven emotions (i.e., happiness, sadness, fear, surprise, disgust, anger, and no emotion) based on the facial expression, tone of voice and body movement in 21 short video vignettes. Though the protocol is effective in measuring participants' ability of integrating multimodal cues, it is not very powerful in separating the perception of emotional cues in a mixture of sensory modalities. In other words, the mean accuracy of the experiment items in this task may not tell which sensory modalities participants rely more on in making correct responses, and which emotion cues they are less sensitive to during multisensory integration.

The Comprehensive Affect Testing System (CATS) is also a well-validated tool, which is widely used to measure emotion recognition via facial expression, prosody, and semantic content (Froming *et al.*, 2006). It has become available in seven languages (i.e., American English, British English, French, German, Italian, Portuguese, and Spanish). According to Schaffer *et al.* (2009), CATS employs internationally standardized and universally recognized male and female faces (Ekman and Friesen, 1976), which pose neutral or emotional (e.g., happy, sad, angry, surprised, fearful, or disgusted) expressions. The auditory prosodic stimuli are produced by a male actor, and consist of a set of sentences with emotional (e.g., happy, sad, neutral, frightened, or angry) or neutral prosody and semantic content. This testing system constitutes 13 subsets, which can be categorized into four main types in terms of emotional perception, including (a) identification tasks of facial or prosodic emotion, (b) discrimination tasks of facial or prosodic emotion, (c) multimodal emotion-matching tasks, and (d) Stroop tasks of auditory, visual and/or semantic information of emotion which involve emotion judgment with conflicting cues. Specifically, participants are asked to name or select the emotion expressed through a visual or auditory modality in an identification task, and differentiate the stimuli presented in pairs or groups of three by deciding whether two communication modalities convey the same emotion or not in a discrimination task. While emotion is generally expressed in either a visual or an auditory modality in identification and discrimination tasks, matching tasks and Stroop tasks can present emotional information in more than a single modality. In a multimodal emotion-matching experiment, participants are often asked to match the emotional cues with a congruent valence across different sensory modalities (Castagna *et al.*, 2013). This matching task requires more cognitive efforts compared with the BLERT, since participants are not allowed to rely on only one of the communication modalities, but focus their attention on all the given cues in order to make correct judgments. However, the task has yet to be effective in unmasking the interplay of different sensory modalities during emotional perception.

2.1.2. Disentangling the Interactions Among Multisensory Emotional Cues
An increasing number of researchers have adopted popular bimodal emotion Stroop tasks in the CATS to investigate how participants disentangle the interactions of different communication modalities during MSI of emotion (de Gelder *et al.*, 2005; de Jong *et al.*, 2009, 2010, 2013; Giannitelli *et al.*, 2015; Mangelinckx *et al.*, 2017; Seubert, 2010; Van den Stock *et al.*, 2011; Vogel *et al.*, 2016). Originated from the classic word–color Stroop effect (Stroop, 1935), the audiovisual emotion Stroop tasks require the participants to focus on the emotional information conveyed in one single modality, and ignore the signals presented through the other channel with either congruent or incongruent emotional valence(s).

In order to examine whether the basic abilities of MSI of emotion was intact in schizophrenics, de Gelder *et al.* (2005) were one of the first research teams who adopted such a bimodal emotion categorization task. They compared how schizophrenic patients and healthy controls perceived facial expressions which were taken from an 11-step morphing continuum extending between two natural physiognomic facial tokens of happiness and sadness and co-presented with semantically neutral sentences spoken in happy, sad, and neutral tones of voice. De Jong *et al.* (2009) extended this paradigm by presenting facial expressions in a more explicit and dichotomous form, thus reducing the influence of attention and post-perceptual strategies. By enlarging the sample size and including nonschizophrenia psychosis patients as a second comparison group, they provided more robust evidence of deficient MSI of emotional information in schizophrenia. Furthermore, de Jong *et al.* (2010) utilized semantically neutral short phrases, and concurrently presented facial expressions in two experiments, in which the target emotions were happy/sad and happy/fearful respectively. They also modulated participants' modality-specific attention by presenting pairs of digits and pure tones as nonemotional visual and auditory distractors, which addressed the limited consideration of selective attention mechanisms in the previous two studies. Taken together, these studies discriminate among schizophrenic patients, nonschizophrenic psychotic patients and healthy people by providing a common cognitive assessment tool on multisensory integration patterns in the course of emotional processing.

A large and ever-growing body of literature on the interactions of cross-modal communication channels has been published in recent years (de Jong *et al.*, 2013; Giannitelli *et al.*, 2015; Mangelinckx *et al.*, 2017; Seubert, 2010; Van den Stock *et al.*, 2011; Vogel *et al.*, 2016). In these studies, the stimuli were not restricted to human faces and vocalizations, but have also been extended to body movements and animal sounds (Van den Stock *et al.*, 2011). Apart from visual faces and auditory prosodic cues, the effects of odorant primes on the emotional face recognition have also been explored (Seubert, 2010). Along with the diversification of multisensory stimuli, more researchers have delved into how the interaction patterns among various communication modalities are related to other cognitive functions and psychotic symptoms (de Jong *et al.*, 2013; Giannitelli *et al.*, 2015).

2.2. Theoretical Explanations for Deficits in MSI of Emotion

Integrating emotional information from multiple communication channels serves as a complex process. A vast majority of research in this field has demonstrated schizophrenic patients' dysfunctions in MSI of emotions (de Gelder *et al.*, 2005; de Jong *et al.*, 2009, 2010; Fiszdon and Bell, 2009; Giannitelli *et al.*, 2015; Mangelinckx *et al.*, 2017; Seubert, 2010). However, studies do not always share a consensus concerning (a) whether congruent (incongruent) information across different communication modalities facilitates or interferes with MSI of emotions, (b) whether auditory or visual emotional cues are more perceptually salient for schizophrenic patients, and (c) whether patients' deficits in integrating multimodal emotional cues stem from generalized neurocognitive impairments or belong to specific emotional dysfunctions that may be influenced by linguistic and cultural backgrounds.

2.2.1. Congruence-Induced Facilitation Effect and Incongruence-Induced Interference Effect

There is a clear pattern in MSI literature on healthy people. Participants respond faster and more accurately when encountering congruent information across modalities during MSI of emotion, and incongruent stimuli tend to elicit interference with emotion processing (Barnhart *et al.*, 2018; Collignon *et al.*, 2008; Liu *et al.*, 2015; McGurk and MacDonald, 1976; Niedenthal, 2007; Pell *et al.*, 2011; Schwartz and Pell, 2012). This is referred to as the congruence-induced facilitation effect and incongruence-induced interference effect, which also largely hold true for schizophrenic patients (de Jong *et al.*, 2009, 2010; Mangelinckx *et al.*, 2017; Tseng *et al.*, 2015).

However, it remains unclear to what extent congruent emotional signals co-occurring in different modalities facilitate MSI of emotion, and to what extent incongruence hampers cross-modal emotional processing. The findings are mixed as exaggerated (de Gelder *et al.*, 2005; de Jong *et al.*, 2010; Van den Stock *et al.*, 2011), diminished (de Gelder *et al.*, 2005; de Jong *et al.*, 2009), and unaffected (de Jong *et al.*, 2010; Müller *et al.*, 2014) multisensory integration in schizophrenics under the influence of cross-modal congruence or incongruence have all been reported.

Another inherently challenging issue arises concerning the asymmetry between emotionally congruent and incongruent trials in terms of required cognitive resources for conflict resolution, sensory inhibition, and response selection. When encountering cross-modal emotional conflicts, participants will by nature allocate more attention for conflict monitoring and resolution, which evokes greater activation in a cingulate-fronto-parietal network (Müller *et al.*, 2011). Thus, the studies contrasting cross-modal congruence and incongruence may not comprehensively represent schizophrenic patients' dysfunction of integrating multisensory emotional cues, but are likely to be confounded by their attentional and higher-level cognitive deficits (Pinheiro *et al.*, 2013; Tseng *et al.*, 2015; Zvyagintsev *et al.*, 2013).

2.2.2. Auditory vs Visual Dominance

Traditionally, multisensory interference effects are represented as misleading perceptual illusions. Some classic examples of audiovisual nonemotional processing confirm the influential role of visual channels such as the spatial ventriloquism effect (i.e., the localization of sounds biased by visual stimuli) (Bertelson and Radeau, 1981) and McGurk effect (i.e., the perception of speech sounds affected by observing lip movements) (McGurk and MacDonald, 1976). Others provide evidence for the interference by auditory signals, such as sound-induced flash illusions (i.e., fission or fusion visual illusion of a flash caused by a concurrent sound) (Shams *et al.*, 2000). Calvert *et al.* (2004), Colavita and Weisberg (1979), Spence (2009), Spence *et al.* (2011), and Vroomen and de Gelder (2000) have elaborated more details on perceptual illusions and modality dominance effects.

Modality dominance effects have also been extensively studied in recent research on emotional processing. The auditory dominance effect for patients with schizophrenia has been frequently observed, compared with healthy people who preferentially rely on visual signals when recognizing emotion (Collignon *et al.*, 2008). In the study conducted by Van den Stock *et al.* (2011), patients with schizophrenia tended to base their decisions on the auditory emotion cues more than controls, but only when human vocalizations instead of animal sounds were concurrently presented with dynamic, face-blurred and whole-body expressions. Apart from nonlinguistic emotional vocalizations, an auditory dominance effect has been found on linguistic inputs. For example, de Jong *et al.* (2009) revealed a diminished impact of facial expressions on vocal emotion recognition in schizophrenics when participants were asked to selectively attend to the vocal emotional phrase and ignore the emotional face presented simultaneously. Similarly, Vogel *et al.* (2016) adopted an audiovisual emotional sentence rating task, in which participants were asked to rely on all the available verbal and nonverbal cues to judge the emotional states of the speakers. The patients were found to be less influenced by nonverbal signals (i.e., facial expressions and tone of voice) and more by verbal information (i.e., spoken content) than healthy controls, even though the nonverbal dominance effect was present in both groups. Such perceptual reliance on auditory semantic cues could pose challenges to schizophrenics in complex social interactions when the linguistic information is ambiguous or when it is incongruent with paralinguistic signals such as gestures and facial expressions.

Interestingly, emotions expressed through visual channels can be perceptually more salient than auditory signals in some cases. According to de Gelder *et al.* (2005), the interactions between auditory and visual channels are bidirectional, with an increased effect of faces on vocal emotion categorization and a reduced effect of voices on facial emotion perception observed. This could partly be explained by vocal emotion recognition difficulty among schizophrenics, and partly by a stronger visual than auditory dominance effect in audiovisual perception in the patients. In line with this research, de Jong *et al.* (2010) corroborated the regulatory role of modality-specific attention on MSI of emotion by including auditory and visual distractors in the original bimodal emotion recognition paradigm. They indicated that compared with healthy people, schizophrenics experience a stronger impact of facial expressions on vocal emotion perception. Moreover, an exaggerated cross-modal influence occurs for schizophrenics when auditory distraction is presented, as opposed to attenuated integration for the control group, which may be associated with impairments in fundamental auditory processing among the patients.

2.2.3. Generalized Neurocognitive Impairments vs Specific Functional Deficits

Although emotion-processing deficits are evident in schizophrenics, it has been controversial whether patients' dysfunctional emotion processing represents global impairments or reflect distinct functional deficits (Bryson *et al.*, 1997; Dickinson *et al.*, 2004; Driver and Noesselt, 2008; Kerr and Neale, 1993; Kohler *et al.*, 2000; Kucharska-Pietura *et al.*, 2005; Lencz *et al.*, 2006; Mandal *et al.*, 1998; Mangelinckx *et al.*, 2017; Penn *et al.*, 2000; Silverstein, 2008). Many investigators argued that the deficits in processing emotion and integrating information from multiple modalities manifest a general tendency for schizophrenics to make poorer performances than healthy controls (Dickinson *et al.*, 2004; Heinrichs and Zakzanis, 1998; Huang *et al.*, 2013; Kerr and Neale, 1993; Lencz *et al.*, 2006). In the light of this view, the schizophrenia-associated impairments are shared across neuropsychological domains (Dickinson *et al.*, 2004), relatively static from illness onset (Hill *et al.*, 2004), by nature related to the illness without the confounding factors of medications and institutionalizations, and cannot be accounted for by lesions in a single anatomical region (Mohamed *et al.*, 1999).

However, the generalized deficit model has been challenged with clinical and empirical evidence. While some researchers agreed on a generalized perceptual impairment in schizophrenics, they also presented evidence for specific emotion-processing deficits in patients. In this case, links between emotion perception and social functioning will be independent of shared associations with neurocognitive abilities (Fiszdon and Johannesen, 2010). Such differential functional impairments have been demonstrated to be emotion-specific, patient-specific, task-specific, modality-specific and brain region-specific (Barkhof *et al.*, 2015; de Jong *et al.*, 2010; Driver and Noesselt, 2008; Hanlon *et al.*, 2005; Penn *et al.*, 2000; Silver *et al.*, 2002, 2009). While it is not difficult to determine cognitive dysfunctions in schizophrenics, one challenge is to assess a specific emotion-processing deficit unconfounded by other psychometric artifacts that affect participants' test performances (Silverstein, 2008). In MSI of emotions, another unanswered issue is whether patients' poorer performances are caused by the functional inability to integrate information from different channels or dysfunctions in correctly perceiving the emotions in a single modality (Castagna *et al.*, 2013).

2.3. Influential Factors

Factors influencing the controversial theoretical accounts for multimodal emotion-processing deficits in schizophrenics have been extensively discussed in the relevant studies. These factors largely fall into three categories: experimental manipulations, psychotic symptomatology and cognitive functioning.

2.3.1. Experimental Manipulations

As indicated in Section 2.1, existing studies vary from one to another in terms of research focuses, and the corresponding experimental paradigms are also different, thus resulting in controversial research outcomes. First, emotion valence has often been regarded as one of the potential factors. For example, de Jong *et al.* (2010) speculated that different response patterns can be activated possibly due to the changes in emotion types (i.e., happy vs sad, and happy vs fearful). Notably, patients have more difficulty in processing negatively-valenced signals, whereas they do not differ from healthy people in perceiving items with a positive emotion especially when stimuli are presented in the auditory channel (Fiszdon and Bell, 2009). In addition to stimulus valence and modality, task demands as a result of different task types, also trigger different responses by the patients (de Gelder *et al*., 2005; de Jong *et al*., 2009, 2010, 2013; Giannitelli *et al*., 2015; Simpson *et al*., 2013; Van den Stock *et al*., 2011; Zvyagintsev *et al*., 2013). It has been suggested that emotion identification tasks do not always generate similar performances as discrimination tasks (Penn *et al*., 2000; Lin *et al*., 2018). Such variances may be attributed to the differential cognitive efforts required by the tasks and the cognitive threshold of individual participants; in other words, schizophrenics can only process an absolute amount of information at a time and may not benefit from an increase in information during multimodal presentation to the same extent as healthy people (Fiszdon and Bell, 2009).

2.3.2. Psychotic Symptomatology

Participant-related factors serve as another important contributor to the inconsistent findings among existing literature. There has been an explosion of interest in how schizophrenic patients' clinical symptoms are related to their emotion perception deficits (Brazo *et al*., 2014; Ito *et al*., 2013; Mitchell and Rossell, 2014; Rossell and Boundy, 2005; Shea *et al*., 2007). For single-emotion processing, facial emotion recognition deficits are related to negative symptoms (e.g., blunted affect, emotional withdrawal, social avoidance) of the patients (Chan *et al*., 2010), and auditory affective processing deficits are more likely to be associated with positive symptoms (e.g., delusions, hallucinations, disorganized speech, confused thought) (Castagna *et al*., 2013; Rossell and Boundy, 2005; Shea *et al*., 2007).

When encountering emotional cues from auditory and visual channels, patients exhibit specific deficits in matching the multimodal information, which are affected by poor executive functioning and the negative symptoms of schizophrenia (Castagna *et al*., 2013). However, in the study conducted by Giannitelli *et al*. (2015), there was no clear evidence supporting an association between emotion identification deficits and either the positive or negative symptoms. Instead, a significant correlation between emotion processing and nonverbal communication disorders has been identified. It may be the case that patients with early-onset schizophrenia spectrum other than adult schizophrenics were recruited as participants in the experiment. Moreover, the effects of psychopathological comorbidities and pharmacological status on moderating schizophrenic patients' performances also deserve further examination in future research (Kawano *et al*., 2016; Mangelinckx *et al*., 2017; Vogel *et al*., 2016). By comparing how patients with schizophrenia and with other psychotic disorders perform multimodal emotion perception tasks, we can gain new insights into the diagnosis and rehabilitation of psychiatric illnesses in clinical practice (Brazo *et al.*, 2014; Ito *et al.*, 2013; Lysaker *et al.*, 2017; Postmes *et al.*, 2014).

2.3.3. Cognitive Functioning

The role of cognitive functioning has been well established in past research. According to Javitt (2009), bottom-up sensory impairments may serve as a hindrance in the correct identification of cognitive functioning deficits. It is futile to remediate social cognition without a thorough consideration of patients' early deficits in processing visual and auditory information (Hall *et al.*, 2004; Leitman *et al.*, 2005, 2007). In terms of MSI, a number of researchers have attempted to explain schizophrenic patients' abnormalities in multimodal emotion perception by means of their unisensory primary perceptual deficits. De Gelder *et al.* (2005) observed a reduced impact of emotional prosody on the categorization of facial expressions as well as an increased visual influence on emotional prosody recognition, which may indicate the patients' relative difficulty in auditory processing. In a recent study (Kraus *et al.*, 2019), auditory spectral processing deficits (including pitch processing and formant processing) were confirmed to be the most significant contributors to emotional processing deficits in schizophrenia. De Jong *et al.* (2010) also speculated that due to the potential fundamental auditory and visual processing deficits in schizophrenics, patients are likely to be overpowered by the co-occurring inputs that elicit a fierce competition for attentional resources. Mangelinckx *et al.* (2017) further underpinned emotion-processing impairments by basic perceptive-motor deficits (i.e., slower processing speed), which again confirms the role of low-level functioning. They adopted a subtractive method in statistical analyses, in which mean reaction time in a control task measuring perceptivo-motor deficits was subtracted from mean response time in the task with emotional stimuli. This quantitative method allowed the assessment of perceptive-motor delay in emotion processing. Though primary perceptual deficits can explain patients' poorer performance of integrating multimodal emotion cues to some extent, the effect of top-down neurocognitive dysfunctions (i.e., attention, memory, executive functioning) cannot be ruled out (Bryson *et al.*, 1997; Castagna *et al.*, 2013; Fiszdon and Johannesen, 2010; Mangelinckx *et al.*, 2017; Simpson *et al.*, 2013). For instance, selective attention has been regarded as a high-order and supramodal regulatory system, which can activate different response patterns due to the changes in emotions and tasks. Though it serves as an adaptive process for MSI of emotion, the deficient top-down attentional selection mechanism in schizophrenics can fuse their integration when they are confronted with concurrently presented conflicting information (de Jong *et al.*, 2010; Seubert, 2010; Zvyagintsev *et al.*, 2013). Besides, executive functions have also been demonstrated to take part in influencing not only emotional prosody perception, but also audiovisual emotion-matching ability (Bryson *et al.*, 1997; Castagna *et al.*, 2013). Different from bottom-up cognitive deficits that are largely associated with posterior cortical regions, the top-down mechanisms tend to involve higher-order (frontal and parietal) association cortices (Adcock *et al.*, 2009).

Explaining the emotion cognition deficits with either the bottom-up or top-down processing deficit can be helpful but insufficient to uncover the nature of multimodal emotion perception and unravel the mysteries of schizophrenia. Recent years have witnessed some researchers combining both low-level and high-order functioning in conceptualizing multisensory integration deficits in schizophrenic patients. Their basic notion is that the biases in audiovisual emotion judgment may stem from a fundamental sensorial deficit, a high-order cognitive dysfunction, or both (Dondaine *et al.*, 2014). De Jong *et al.* (2013) proposed a cohesive model to depict the three stages of multimodal emotion perception, namely sensory processing, neurocognition, and social cognition (Fig. 1a). During sensory processing of the stimuli, participants often integrate interpersonal signals presented auditorily and visually in the social environment. Then, the emotional stimuli interplay with and are shaped by attention, memory and executive functions. In the final phase, participants will interpret and evaluate the stimuli based on their social cognition. Patients may have the following four dysfunctions: (1) impaired sensory processing, (2) separation between sensory processing and neurocognition, (3) neurocognitive deficits, and (4) impaired social cognition. By means of structure equation modeling, the research team further pinpointed that it is the increasing disintegration between pre-attentive sensory processing and neural cognition that significantly distinguishes schizophrenic patients from healthy controls (de Jong *et al.*, 2013).

3. Electrophysiological and Neuroimaging Studies

The conceptual model described in Fig. 1 is enlightening as it shows how MSI works in terms of time sequence, which lays a solid foundation for further research on the underlying brain mechanisms in terms of the temporal and spatial brain activities.

3.1. Temporal Dynamics of MSI of Emotion

With the advances in the event-related potentials (ERP) methodology, many researchers have examined the electrophysiological indices of MSI. Multisensory integration is an automatic process first appearing as early as 110 ms after the stimuli are concurrently presented (Föcker and Röder, 2019; Pourtois *et al.*, 2000). Previous research has identified schizophrenic patients' reduced N170 to both face and face-voice stimuli, indicating abnormalities in decoding

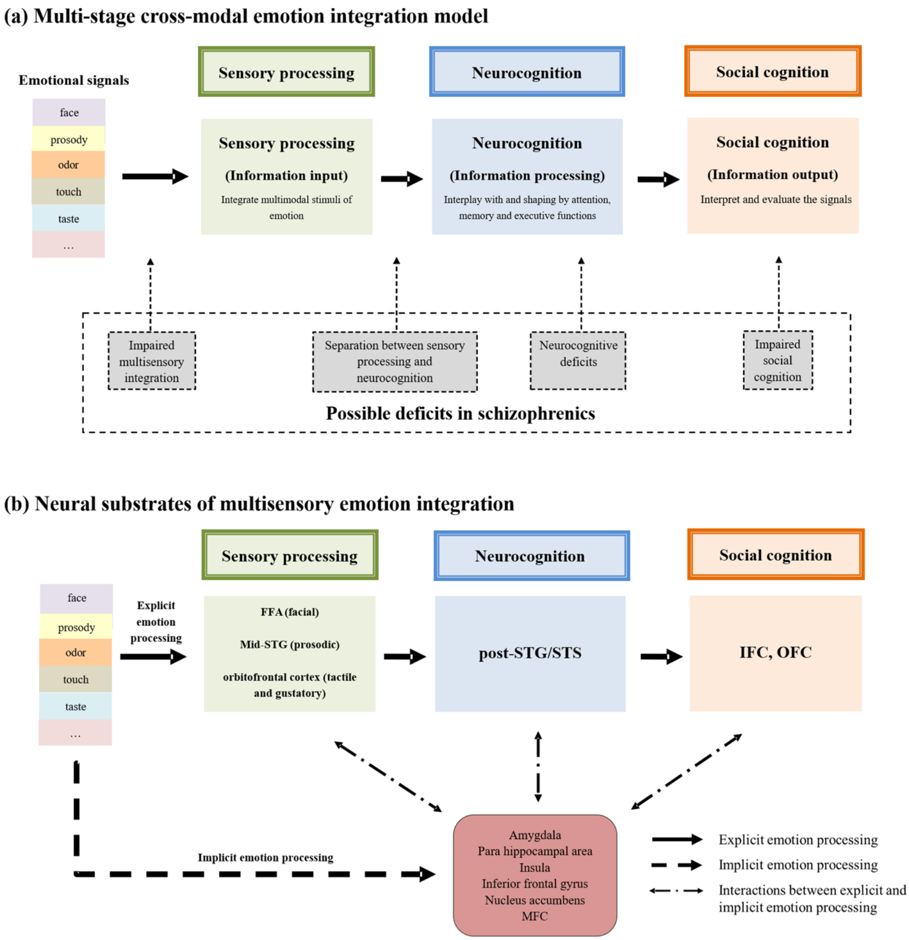

Figure 1. (a) Multi-stage cross-modal emotion integration model for healthy people and patients with schizophrenia based on de Jong *et al.*'s (2013) work. (b) Potential abnormal neural substrates during different stages of multisensory emotion integration in schizophrenics based on Adolphs's (2002), Tseng *et al.*'s (2015), and Wildgruber *et al.*'s (2009) work. *Note*. In Fig. 1(a), three stages of MSI of emotion, namely sensory processing, neurocognition, and social cognition, are respectively presented in green, blue and orange text boxes. Four likely dysfunctions in schizophrenics are marked by text boxes in gray with dashed outlines. In Fig. 1(b), brain regions associated with explicit emotion processing are shown according to the aforementioned three processing stages with solid arrows, and those related to implicit emotion processing are presented in a red text box with a dashed arrow. The reciprocal interactions between explicit and implicit processing are indicated by long dash dot arrows.

facial and audiovisual information (Liu *et al.*, 2016; Turetsky *et al.*, 2007). In a recent study on schizophrenics' audiovisual affective processing, Müller *et al.* (2014) employed an emotion rating task in which schizophrenics and healthy controls were asked to rate the happy, fearful or neutral faces concurrently

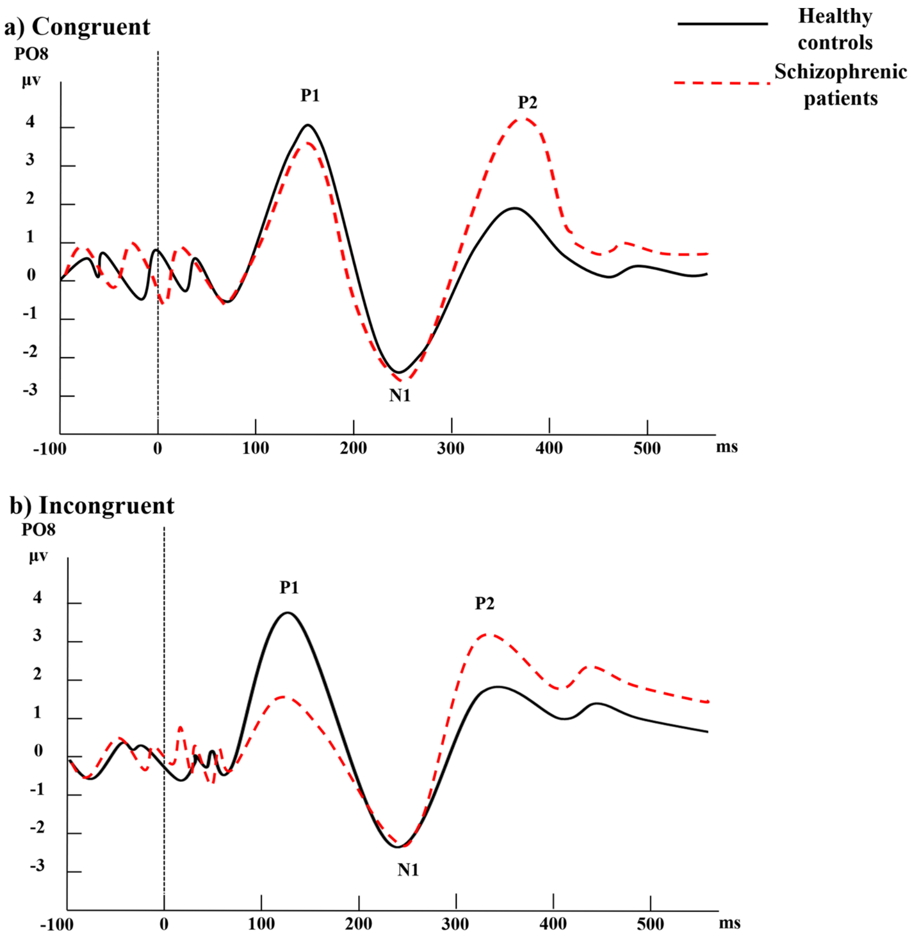

Figure 2. Event-related potentials in response to faces presented with emotionally (a) congruent and (b) incongruent sounds in schizophrenic patients (dashed lines) and healthy controls (solid line) at the PO8 electrode; based on Müller *et al.*'s (2014) work.

presented with emotional or neutral vocal sounds in an ERP study (Fig. 2). This electrophysiological study has provided a coherent picture with previous behavioral investigations, as an increased P2 was observed in schizophrenics, indicating early visual emotion-processing deficits. Also, the patients revealed a significant reduced visual P1 response when emotionally incongruent sounds from different sensory channels were simultaneously presented, which was negatively correlated with general symptom severity. These results suggest that early affective face processing deficits, though associated with symptom severity, can be attenuated by emotionally congruent sounds.

Though efforts have been dedicated to investigating the neural correlates that underlie schizophrenic deficits of audiovisual non-emotional speech integration (Ross *et al.*, 2007; Stekelenburg *et al.*, 2013; Stone *et al.*, 2011), the research conducted by Müller *et al.* (2014) has been one of the very few electrophysiological studies unfolding how emotion face processing is modulated by emotional sounds in time in schizophrenics. Their work serves as a breakthrough exploring the temporal dynamics of emotional MSI in schizophrenic patients. In addition to examining the influence of auditory information on visual affect processing, a reverse experimental paradigm can be developed to investigate how emotional information conveyed visually facilitates or interferes with auditory processing in terms of temporal dynamics. Furthermore, we can extend well-established research paradigms originally targeting healthy participants to patients with psychotic disorders (Diamond and Zhang, 2016; Filippi *et al.*, 2017), which would allow us to gain a more comprehensive understanding of how patients and healthy participants differ in MSI of emotion in terms of neurophysiological representations.

3.2. Neuroanatomical Mechanisms of MSI of Emotion

Apart from temporal dynamics, abnormal neuroanatomical functioning during multisensory emotion integration in schizophrenics also receives widespread attention among researchers. Resolving emotion-related conflicts involves intact callosal functioning, limbic regions (e.g., amygdala and insula) and dorsolateral prefrontal cortex for information transfer across hemispheres (Mitchell and Rossell, 2014; Müller *et al.*, 2011). However, a reduced corpus callosum (Arnone *et al.*, 2008), hyper- or hypoactive amygdala (Müller *et al.*, 2011; Strauss *et al.*, 2008), aberrant neural responses in the insula (Morgane *et al.*, 2005), and attenuated dorsolateral prefrontal activity (Vercammen *et al.*, 2012) have been observed in schizophrenic patients. In addition to a single brain region, there has been an increasing interest in functional connectivity (i.e., the temporal correlation between the time series activities of different brain regions) among clinical populations in recent resting-state and task-state studies (Anticevic *et al.*, 2011; Carballedo *et al.*, 2011; Das *et al.*, 2007; Fox and Greicius, 2010; Lynall *et al.*, 2010).

Following the research trends, Müller *et al.* (2013) undertook a functional Magnetic Resonance Imaging (fMRI) study where participants had to rate emotional faces while ignoring vocal sounds, and employed both task-independent (resting state), and task-dependent meta-analytic connectivity modeling (MACM) analyses. The MACM approach characterizes the co-activation profile of a certain seed region in regard to its connectivity and function across a large number of functional neuroimaging studies. In the analysis, all experiments that activate the seed region are first identified from large-scale databases [e.g., BrainMap (Laird *et al.*, 2009a) and Neurosynth

(Yarkoni *et al.*, 2011)] storing standardized coordinate information. Subsequently, quantitative meta-analysis is adopted to examine the convergence across the foci reported in these experiments. Significant convergence across different brain regions suggests consistent co-activation (i.e., functional activity with the seed; Laird *et al.*, 2009b; Robinson *et al.*, 2010). In the final stage, a conjunction analysis between the resting-state results and MACM analyses is conducted to show both task-independent and task-related connectivity of the seed region. In line with previous work, the new analysis by Müller *et al.* (2013) indicated a dysregulated left inferior parietal cortex (IPC) in both schizophrenic and depressive patients during MSI of emotion. They also demonstrated a functional network of left IPC, precuneus and posterior cingulate cortex (PrC/PCC), medial orbitofrontal cortex (mOFC), left middle frontal (MFG) as well as inferior frontal (IFG) gyrus. These results not only suggest impaired audiovisual integration, but also imply dysfunctional emotional and cognitive processing in the patient groups. Similarly, Dzafic *et al.* (2018) further revealed less activated superior temporal gyri (STG) and aberrant neural circuitry in patients involving inferior frontoparietal regions, the left insula and right amygdala functional network by employing an experiment exploring dynamic emotion perception. Thus, deficient emotion-related conflict perception in schizophrenics may not be attributed to dysfunctions in a single area, but involves multiple brain regions, connectivity, and circuities that constitute a complex, dynamic cognitive system. Interestingly and importantly, these studies have replicated the findings of previous MSI research using non-emotional stimuli (Straube *et al.*, 2014; Szycik *et al.*, 2013): there appears to be altered connectivity in IFG and STG networks responsible for high-order integrative functions in schizophrenics, thereby causing difficulty for the patients to integrate both linguistically abstract and emotionally contradictory audiovisual stimuli.

The findings from behavioral and neuroimaging studies have collectively enriched the multi-stage cross-modal emotion integration model for both normal people and schizophrenic patients by indicating the relevant neural substrates (Fig. 1b) (Adolphs, 2002; Calvert, 2001; Regenbogen *et al.*, 2015; Tseng *et al.*, 2015; Wildgruber *et al.*, 2009). Initially, emotional cues from different communication channels are decoded at modality-specific primary or secondary cortices [e.g., facial: fusiform face area (FFA), prosodic: middle section of superior temporal gyrus (mid-STG), tactile/gustatory: orbitofrontal cortex]. Multisensory emotional signals are then blended and meaningfully identified with the help of multimodal association areas [e.g., posterior section of superior temporal gyrus or sulcus (post-STG/STS)]. In the final stage, cognitive evaluations are made within the inferior frontal cortex (IFC) and the orbitofrontal cortex (OFC). These three stages constitute explicit emotion processing, which reciprocally interacts with implicit (automatic) processing during the whole process. While cognitively controlled and explicit emotion evaluation mostly happens at a cortical level, implicit emotion processing is generally conducted in subcortical limbic regions [e.g., amygdala, parahippocampal area, insula, IFG, nucleus accumbens, medial frontal cortex (MFC)]. In line with this model, a series of neuroimaging studies have been conducted showing that participants' emotion-processing performances are largely modulated by other cognitive abilities (e.g., attention, memory, conflict resolution and monitoring, language comprehension, theory of mind) that are mostly associated with these cortical or subcortical regions (Dzafic *et al.*, 2018; Lysaker *et al.*, 2017; Müller *et al.*, 2013; Razafimandimby *et al.*, 2016).

4. Discussion

The current review has explored the extent and nature of MSI of emotion in schizophrenics by summarizing the latest development in behavioral and neurological studies. In general, schizophrenic patients show impaired MSI ability in emotion processing. Experimental paradigms adopted in the existing literature vary from one study to the next, but they investigated schizophrenics' performances from two main perspectives: how participants understand multimodal emotional information, and how they resolve the interplay among different sensory modalities. Existing behavioral studies have been inconsistent toward congruence-induced facilitation (incongruence-induced interference) effects, modality dominance effects, and generalized vs specific impairment hypotheses, which can be attributed to variations in experimental manipulations, patients' clinical profiles, and their cognitive abilities. There have been a relatively smaller number of studies using neurological techniques compared to behavioral ones. Nevertheless, abnormal electrophysiological indices (e.g., P1, P2), and activation patterns in cortical regions (e.g., mid-STS, post-STS/STG, IFC, OFC) and subcortical circuities (e.g., amygdala, insula, inferior frontoparietal regions) serve as essential biomarkers for impaired integration of multisensory emotional signals in schizophrenics, indicating dysfunctional temporal dynamics and neuroanatomical substrates during patients' MSI of emotion.

4.1. Limitations of Existing Research

There are several limitations in the literature that need to be considered when interpreting the findings. First, extant studies tend to be limited in experimental design types and measures, with many more cross-sectional studies than longitudinal ones, and considerably more research using behavioral methods than electrophysiological and neuroimaging techniques (Herniman *et al.*, 2017; Tseng *et al.*, 2015). Though some key ERP indices and spatial localization of brain functions during MSI of emotion have been identified in the related studies, there are many unknowns regarding the cortical and subcortical dynamics responsible for the generation of the possible ERP components in schizophrenics. In other words, more thorough investigations should be made, which can combine the timing and source localization of neural responses in the MSI of emotion (Diamond and Zhang, 2016).

Secondly, many studies tend to be biased in terms of the inclusion of study samples, with chronic schizophrenics more frequently involved compared to patients with acute symptoms or early-onset schizophrenia spectrums. Even fewer studies have examined schizophrenic patients with comorbid diseases, compared patients of different schizophrenic subtypes, or included participants of other special groups such as nonschizophrenic psychotic patients and aging populations (Chaby *et al.*, 2015; Kohler *et al.*, 2000; Kret and Ploeger, 2015; Weniger *et al.*, 2004; Williams *et al.*, 2007). Thus, it is unclear to what extent the results are generalizable to a broader population influenced by schizophrenia throughout the development and maintenance of this psychotic disorder, and to what extent schizophrenics exhibit different behavioral and neural responses due to diversified symptom profiles (Bonfils *et al.*, 2019; Gee *et al.*, 2012).

Another limitation concerning stimulus design is that existing cross-modal studies are mostly restricted to audiovisual emotion processing, especially the perception of the human face and voice, and few have developed stimuli of other modalities (e.g., tactile, gustatory) or forms (e.g., body movements, gestures, music). As a procedural operation, existing Stroop studies generally involve binary contrasts of emotional stimuli (e.g., facial vs prosodic, prosodic vs semantic), thereby underrepresenting the natural communicative settings where we are simultaneously exposed to more than two communication channels of emotion expression (Lin and Ding, 2019). It is furthermore worth noting that despite a widespread employment of linguistic stimuli, very few researchers have thoroughly investigated how patients can be distinguished from the comparison group when processing linguistic inputs that are phonetically, phonologically, morphologically, syntactically, semantically or pragmatically different.

Moreover, since most relevant studies have been conducted in western stressed-time nontonal languages (e.g., English, German, Dutch), it remains to be tested whether the findings are applicable to eastern languages like Mandarin Chinese that bears unique features of lexical tones, a syllable-timed rhythmic pattern, hieroglyphics, null arguments, and a high-context culture (Chong *et al.*, 2015; Hall, 1976). In one study (Chan *et al.*, 2008), Mandarin-speaking patients with nonparanoid schizophrenia were found to demonstrate deficits in both facial and prosodic emotion recognition, and spatial perception ability was found to correlate with facial emotion recognition performance while attentional processing predicted prosodic emotion processing in both identification and discrimination tasks. Tseng *et al.* (2013) further examined the functional correlates of emotion recognition deficits in Mandarin-speaking patients with schizophrenia. They found that processing of a specific emotion category (i.e., happy emotion) together with executive-function deficit can predict severity of their psychotic symptoms in the positive dimension. There is also a pioneering ERP study (Chen *et al.*, 2016) that employed the passive listening oddball paradigm to investigate the neural markers of emotional voice processing deficits in association with behavioral symptomatology of Chinese-speaking schizophrenic patients. The findings confirmed deficits of emotional voice processing with weaker mismatch negativity amplitudes associated with more positive schizophrenic symptoms. More importantly, Chen *et al.* (2016) further revealed that in the nonattentive condition, the mismatch negativity response and P3a response (an index of involuntary attentional orientation to stimulus differences in the oddball paradigm) to the angry (not happy) voice were good predictors of the clinical diagnosis. As suggested by Tseng *et al.* (2012) and Yang *et al.* (2012), the potential consequences of the universal neuropsychological deficits among schizophrenic patients may be language- and culture-specific with respect to the relationships between severity of diagnostic dimensions and deficits in specific categories and modalities of emotion processing. Although none of these studies employed MSI protocols, the behavioral and ERP results have demonstrated informative deficits in auditory as well as visual modalities in line with studies on patients who speak nontonal languages. The potential disparities in the two modalities and their relative dominance effects need to be taken into serious consideration when designing MSI experiments (or intervention) with Chinese schizophrenic patients.

To our knowledge, the feasibility and efficacy of multimodal training approaches have been demonstrated in several studies in the field of second-language acquisition and speech pathology (Hardison, 2010; Heinonen-Guzejev *et al.*, 2014; Moradi *et al.*, 2017; Rao *et al.*, 2017; Yu *et al.*, 2017; Zhang *et al.*, 2009). However, there is very little application of cross-modal emotion integration tests in clinical practice for schizophrenics, despite numerous efforts contributed to examining multimodal emotion processing in mentally impaired patients. Since emotion cognition is intertwined with other cognitive abilities (e.g., theory of mind, attributional style, perspective shifting) (Penn *et al.*, 2008), it deserves careful examination how multisensory emotion processing can improve or can be improved by other domains of social cognitive skills in real-life practice so as to provide more effective cognitive training programs for both healthy people and mentally impaired patients (Barnett *et al.*, 2006; Bechi *et al.*, 2015; de Niear *et al.*, 2018; Iacoviello *et al.*, 2014).

4.2. Implications for Future Research and Practice

The current review highlights the key findings and issues on deficits in MSI of emotion in schizophrenia, which are critical to the development of MSI theories and models for characterizing this clinical population. The widely-acknowledged deficits in MSI of emotion demonstrated in the extant literature serve to propel our probe into the social cognitive mechanisms underlying the dysfunctions that largely disturb schizophrenic patients and induce their social alienation. Theoretical accounts specifically need to take into account the congruence-induced facilitation effects, modality dominance effects, generalized vs specific impairment hypotheses with empirical evidence on how schizophrenic patients integrate multisensory emotional signals, and resolve cross-modal discrepancies.

This review provides important implications for clinical practice. The behavioral and neural indices emerging during multimodal emotion processing can serve as a reference tool for comparing schizophrenia with other psychotic illnesses (e.g., depression, anxiety, autism) with regard to similarities and differences in core psychopathological symptoms and social functioning abilities. By carefully considering the variations in individual and experimental characteristics that give rise to different multimodal processing patterns, future practioners can better address the universal demands of most psychotic patients, and meanwhile suit the needs specific to a certain patient or subtype of patients. In this way, they can develop more personalized effective techniques and programs of psychotic diagnosis, intervention, and rehabilitation for the diverse clinical populations.

The current review points out limitations of existing studies and directions for future research and clinical practice. First, more efforts are needed to utilize both cross-sectional and longitudinal designs, and integrate behavioral, electrophysiological, and neuroimaging techniques when determining how individuals with schizophrenia differ from healthy people during MSI of emotion. Recent years have witnessed rapid advancements in the use of artificial intelligence and big data that combine multimodal behavioral, electrophysiological and neuroimaging measures in association with normative sample and clinical profiles including genetic data. To our knowledge, the findings from electrophysiological studies and neuroimaging ones have largely been disassociated, and there has been a lack of understanding of how schizophrenic patients, nonschizophrenic psychotic patients and healthy people differ in integrating multimodal emotional signals across the lifespan. Combining the behavioral and multimodal imaging research methods enables us to reveal the behavioral representations, time course, and functional localization of multimodal emotion processing in schizophrenics in a more objective, scientific and comprehensive way. Moreover, our understanding can be greatly advanced concerning the trajectories of multisensory emotional development or dysfunctions in patients with schizophrenia relative to patients with other mental illnesses or healthy controls (Corcoran *et al.*, 2015; Gee *et al.*, 2012; Hoertnagl *et al.*, 2014; Lin *et al.*, 2018).

It is also important for future work to cover a broader population of patients with different subtypes of schizophrenia or at different stages of the illness, and involve a wider range of participants as comparison groups. This helps us to better ascertain the extent to which impaired MSI performances in emotion processing are related to a series of symptoms and cognitive deficits associated with and specific to schizophrenia. Though the emotion-processing deficit has been consistently recognized as a core feature of schizophrenia spectrum disorders and across different subtypes of schizophrenia (Mandal *et al.*, 1998), few studies have taken into account the behavioral or neurological differences caused by various sample characteristics. Inconsistent findings across studies are possibly due to the variation in the subtypes of schizophrenia (catatonic, paranoid, disorganized, undifferentiated, or residual schizophrenia), symptom types (positive vs negative symptom), onset and duration of the illness (early vs late adulthood; long vs short), course of psychopathology (acute vs chronic), and type and dose of medication (Bonfils *et al.*, 2019). Studying these features of schizophrenia can enhance our understanding of patients' advantages and disadvantages at different phases of the diseases in cross-modal emotion cognition (Simpson *et al.*, 2013; Tseng *et al.*, 2015), which may provide a solid foundation for further exploring the pathogenetic and psychological mechanisms of schizophrenia.

Additional efforts are needed to examine how variations in stimulus design and task demands influence participants' multisensory emotion-processing performances. Past literature has been controversial toward congruence-induced facilitation effects, modality dominance effects, and generalized vs specific impairment hypotheses. Such inconsistencies can be further investigated with more empirical evidence adopting ecologically valid experimental materials that involve more communication channels (e.g., face, prosody, and semantic content presented simultaneously), modalities (e.g., tactile, olfactory), cues (e.g., body movements, eye contact, gestures) and forms (e.g., video clips, static photos, music), and cover more emotion categories (e.g., fear, surprise, disgust), levels of arousal or clarity (e.g., high, mid, low), and modes of presentation (e.g., emotion-laden vs emotion-label) (Bach *et al.*, 2009; Herbener *et al.*, 2007; Lin and Ding, 2019; Lin *et al.*, 2020; Mitchell and Rossell, 2014). This can help reveal the complex nature of MSI of emotion processing in natural communication situations. It is also interesting to examine to what extent the inconsistent findings can be accounted for by variations in processing requirements (e.g., spontaneous vs posed) and response patterns (e.g., forced-choice, matching, open-ended) (Cassels and Birch, 2014; Putnam and Kring, 2007).

Considering the imbalance in the number of studies conducted in different languages and cultures, additional research is needed to shed light on cross-linguistic and cross-cultural differences in multisensory emotion processing in schizophrenics. We therefore highly advocate that stimuli in future research reflect the uniqueness of different languages and cultures, especially those that are underrepresented in the current literature, such as Mandarin Chinese. For instance, it is an interesting testbed to examine whether impaired tone matching abilities, which is associated with auditory word processing deficits in Mandarin-speaking schizophrenics (Yang *et al.*, 2012), affect the understanding of other communication modalities and the integration of multimodal stimuli or not.

Finally, translational efforts are needed to apply basic research findings to clinical practice with schizophrenic patients and other clinical populations. Unclear diagnostic criteria, complex subtypes and comorbidities, and obscure epidemiological symptoms of the psychiatric illnesses all pose great challenges for accurate detection and effective intervention of mental diseases (Phillips and Kupfer, 2013; Ross *et al.*, 2006), which can hardly be alleviated by unimodal strategies. Thus, results from multimodal emotion-processing test batteries can be utilized as an indicator of psychotic abnormalities. Different behavioral and neurological representations in multisensory emotion perception among experimental groups can also provide evidential support for differentiating schizophrenics of different subtypes or at different illness stages (Weniger *et al.*, 2004; Williams *et al.*, 2007), and distinguishing schizophrenics from other clinical populations (Müller *et al.*, 2013). On the other hand, we can develop more therapeutic training protocols including well-validated cross-modal emotion-processing tasks, and intervention programs integrating both top-down and bottom-up strategies (Adcock *et al.*, 2009; Javitt, 2009; Nuechterlein *et al.*, 2014). By sharpening patients' abilities of understanding multimodal emotional information and disentangling emotional cues in multiple modalities, we expect to strengthen not only their emotion cognition, but also other related perceptual abilities and cognitive functions, and even neuroplasticity that exerts protective effects against the mental disorder (Barnett *et al.*, 2006; Bechi *et al.*, 2015, 2016; Clewett and Murty, 2019; Silverstein *et al.*, 2012).

5. Conclusion

In natural communication settings, effective and efficient integration of emotion signals from multiple sensory modalities serves as a cornerstone of successful social functioning. However, when schizophrenic patients are bombarded with multimodal emotion inputs, they generally demonstrate dysfunctions in this interpersonal skill. Despite extensive investigations on patients' understanding of multisensory emotion stimuli and their reciprocal interactions by means of various well-validated protocols, researchers have yet to reach a consensus on congruence-induced facilitation (incongruence-induced interference) effects, modality dominance effects, and generalized vs specific impairment hypotheses, probably due to experimental, clinical, and cognitive factors. Electrophysiological and neuroimaging studies further provide biomarkers for patients' impaired temporal processing and dysfunctional brain regions, connectivity and circuities at different stages of MSI of emotion. Further work is needed to incorporate more research designs and techniques, enlarge study samples involving more schizophrenic subtypes and illness stages, enrich stimulus properties and response patterns, and introduce research outcomes into practical fields of diagnosis and intervention.

Acknowledgements

This study was supported by grants from the major project of National Social Science Foundation of China (18ZDA293). Zhang additionally received the Brain Imaging Grant and Grand Challenges Exploratory Research Grant from the University of Minnesota in support of the international collaboration. We also thank the anonymous reviewers for providing us with insightful suggestions to improve our review, and the editors for efficient handling and professional editing of our manuscript.

References

Adcock, R. A., Dale, C., Fisher, M., Aldebot, S., Genevsky, A., Simpson, G. V., Nagarajan, S. and Vinogradov, S. (2009). When top-down meets bottom-up: auditory training enhances verbal memory in schizophrenia, *Schizophr. Bull.* **35**, 1132–1141. DOI:10.1093/schbul/sbp068.

Adolphs, R. (2002). Neural systems for recognizing emotion, *Curr. Opin. Neurobiol.* **12**, 169–177. DOI:10.1016/S0959-4388(02)00301-X.

Anticevic, A., Repovs, G. and Barch, D. M. (2011). Emotion effects on attention, amygdala activation, and functional connectivity in schizophrenia, *Schizophr. Bull.* **38**, 967–980. DOI:10.1093/schbul/sbq168.

Arnone, D., McIntosh, A. M., Tan, G. M. Y. and Ebmeier, K. P. (2008). Meta-analysis of magnetic resonance imaging studies of the corpus callosum in schizophrenia, *Schizophr. Res.* **101**, 124–132. DOI:10.1016/j.schres.2008.01.005.

Bach, D. R., Buxtorf, K., Grandjean, D. and Strik, W. K. (2009). The influence of emotion clarity on emotional prosody identification in paranoid schizophrenia, *Psychol. Med.* **39**, 927–938. DOI:10.1017/S0033291708004704.

Barkhof, E., de Sonneville, L. M. J., Meijer, C. J. and de Haan, L. (2015). Specificity of facial emotion recognition impairments in patients with multi-episode schizophrenia, *Schizophr. Res. Cogn.* **2**, 12–19. DOI:10.1016/j.scog.2015.01.001.

Barnett, J. H., Salmond, C. H., Jones, P. B. and Sahakian, B. J. (2006). Cognitive reserve in neuropsychiatry, *Psychol. Med.* **36**, 1053–1064. DOI:10.1017/S0033291706007501.

Barnhart, W. R., Rivera, S. and Robinson, C. W. (2018). Different patterns of modality dominance across development, *Acta Psychol. (Amst.)* **182**, 154–165. DOI:10.1016/j.actpsy.2017.11.017.

Bechi, M., Bosia, M., Spangaro, M., Buonocore, M., Cocchi, F., Pigoni, A., Piantanida, M., Guglielmino, C., Bianchi, L., Smeraldi, E. and Cavallaro, R. (2015). Combined social cognitive and neurocognitive rehabilitation strategies in schizophrenia: neuropsychological and psychopathological influences on Theory of Mind improvement, *Psychol. Med.* **45**, 3147–3157. DOI:10.1017/S0033291715001129.

Bechi, M., Bosia, M., Spangaro, M., Pigoni, A., Buonocore, M., Scrofani, D., Cocchi, F., Bianchi, L., Guglielmino, C., Smeraldi, E. and Cavallaro, R. (2016). Visual and audio emotion processing training for outpatients with schizophrenia: an integrated multisensory approach, *Neuropsychol. Rehabil.* **28**, 1131–1144. DOI:10.1080/09602011.2016.1240698.

Bell, M., Bryson, G. and Lysaker, P. (1997). Positive and negative affect recognition in schizophrenia: a comparison with substance abuse and normal control subjects, *Psychiatry Res.* **73**, 73–82. DOI:10.1016/S0165-1781(97)00111-X.

Bertelson, P. and de Gelder, B. (2004). The psychology of multimodal perception, in: *Crossmodal Space and Crossmodal Attention*, C. Spence and J. Driver (Eds), pp. 141–177. Oxford University Press, New York, NY, USA.

Bertelson, P. and Radeau, M. (1981). Cross-modal bias and perceptual fusion with auditory–visual spatial discordance, *Percept. Psychophys.* **29**, 578–584. DOI:10.3758/BF03207374.

Bonfils, K. A., Ventura, J., Subotnik, K. L. and Nuechterlein, K. H. (2019). Affective prosody and facial emotion recognition in first-episode schizophrenia: associations with functioning & symptoms, *Schizophr. Res.: Cogn.* **18**, 100153. DOI:10.1016/j.scog.2019.100153.

Bornik, A., Urschler, M., Schmalstieg, D., Bischof, H., Krauskopf, A., Schwark, T., Scheurer, E. and Yen, K. (2018). Integrated computer-aided forensic case analysis, presentation, and documentation based on multimodal 3D data, *Forensic Sci. Int.* **287**, 12–24. DOI:10.1016/j.forsciint.2018.03.031.

Brazo, P., Beaucousin, V., Lecardeur, L., Razafimandimby, A. and Dollfus, S. (2014). Social cognition in schizophrenic patients: the effect of semantic content and emotional prosody in the comprehension of emotional discourse, *Front. Psychiatry* **5**, 120. DOI:10.3389/fpsyt.2014.00120.

Brüne, M. (2005). Emotion recognition, 'theory of mind,' and social behavior in schizophrenia, *Psychiatry Res.* **133**, 135–147. DOI:10.1016/j.psychres.2004.10.007.

Bryson, G., Bell, M. and Lysaker, P. (1997). Affect recognition in schizophrenia: a function of global impairment or a specific cognitive deficit, *Psychiatry Res.* **71**, 105–113. DOI:10.1016/S0165-1781(97)00050-4.

Calvert, G. A. (2001). Crossmodal processing in the human brain: insights from functional neuroimaging studies, *Cereb. Cortex* **11**, 1110–1123. DOI:10.1093/cercor/11.12.1110.

Calvert, G. A., Spence, C. and Stein, B. E. (2004). *The Handbook of Multisensory Processing*. MIT Press, Cambridge, MA, USA.

Carballedo, A., Scheuerecker, J., Meisenzahl, E., Schoepf, V., Bokde, A., Möller, H.-J., Doyle, M., Wiesmann, M. and Frodl, T. (2011). Functional connectivity of emotional processing in depression, *J. Affect. Disord.* **134**, 272–279. DOI:10.1016/j.jad.2011.06.021.

Cassels, T. G. and Birch, S. A. J. (2014). Comparisons of an open-ended vs. forced-choice 'mind reading' task: implications for measuring perspective-taking and emotion recognition, *PLoS ONE* **9**, e93653. DOI:10.1371/journal.pone.0093653.

Castagna, F., Montemagni, C., Milani, A. M., Rocca, G., Rocca, P., Casacchia, M. and Bogetto, F. (2013). Prosody recognition and audiovisual emotion matching in schizophrenia: the contribution of cognition and psychopathology, *Psychiatry Res.* **205**, 192–198. DOI:10.1016/j.psychres.2012.08.038.

Chaby, L., Luherne-du Boullay, V., Chetouani, M. and Plaza, M. (2015). Compensating for age limits through emotional crossmodal integration, *Front. Psychol.* **6**, 691. DOI:10.3389/fpsyg.2015.00691.

Chan, C. C. H., Wong, R., Wang, K. and Lee, T. M. C. (2008). Emotion recognition in Chinese people with schizophrenia, *Psychiatry Res.* **157**, 67–76. DOI:10.1016/j.psychres.2006.03.028.

Chan, R. C. K., Li, H., Cheung, E. F. C. and Gong, Q. (2010). Impaired facial emotion perception in schizophrenia: a meta-analysis, *Psychiatry Res.* **178**, 381–390. DOI:10.1016/j.psychres.2009.03.035.

Chen, C., Liu, C.-C., Weng, P.-Y. and Cheng, Y. (2016). Mismatch negativity to threatening voices associated with positive symptoms in schizophrenia, *Front. Hum Neurosci.* **10**, 362. DOI:10.3389/fnhum.2016.00362.

Chong, C. S., Kim, J. and Davis, C. (2015). Exploring acoustic differences between Cantonese (tonal) and English (non-tonal) spoken expressions of emotions, in: *INTERSPEECH 2015. 16th Annual Conference of the International Speech Communication Association*, Dresden, Germany, pp. 1522–1526.

Clewett, D. and Murty, V. P. (2019). Echoes of emotions past: how neuromodulators determine what we recollect, *eNeuro* **6**, ENEURO.0108-18.2019. DOI:10.1523/ENEURO.0108-18.2019.

Colavita, F. B. and Weisberg, D. (1979). A further investigation of visual dominance, *Percept. Psychophys.* **25**, 345–347. DOI:10.3758/BF03198814.

Collignon, O., Girard, S., Gosselin, F., Roy, S., Saint-Amour, D., Lassonde, M. and Lepore, F. (2008). Audio–visual integration of emotion expression, *Brain Res.* **1242**, 126–135. DOI:10.1016/j.brainres.2008.04.023.

Corcoran, C. M., Keilp, J. G., Kayser, J., Klim, C., Butler, P. D., Bruder, G. E., Gur, R. C. and Javitt, D. C. (2015). Emotion recognition deficits as predictors of transition in individuals at clinical high risk for schizophrenia: a neurodevelopmental perspective, *Psychol. Med.* **45**, 2959–2973. DOI:10.1017/S0033291715000902.

Couture, S. M., Penn, D. L. and Roberts, D. L. (2006). The functional significance of social cognition in schizophrenia: a review, *Schizophr. Bull.* **32**(Suppl._1), S44–S63. DOI:10.1093/schbul/sbl029.

Das, P., Kemp, A. H., Flynn, G., Harris, A. W. F., Liddell, B. J., Whitford, T. J., Peduto, A., Gordon, E. and Williams, L. M. (2007). Functional disconnections in the direct and indirect amygdala pathways for fear processing in schizophrenia, *Schizophr. Res.* **90**, 284–294. DOI:10.1016/j.schres.2006.11.023.

de Gelder, B., Vroomen, J., Annen, L., Masthof, E. and Hodiamont, P. (2003). Audio–visual integration in schizophrenia, *Schizophr. Res.* **59**, 211–218. DOI:10.1016/S0920-9964(01)00344-9.

de Gelder, B., Vroomen, J., de Jong, S. J., Masthoff, E. D., Trompenaars, F. J. and Hodiamont, P. (2005). Multisensory integration of emotional faces and voices in schizophrenics, *Schizophr. Res.* **72**, 195–203. DOI:10.1016/j.schres.2004.02.013.

de Jong, J. J., Hodiamont, P. P. G., Van den Stock, J. and de Gelder, B. (2009). Audiovisual emotion recognition in schizophrenia: reduced integration of facial and vocal affect, *Schizophr. Res.* **107**, 286–293. DOI:10.1016/j.schres.2008.10.001.

de Jong, J. J., Hodiamont, P. P. G. and de Gelder, B. (2010). Modality-specific attention and multisensory integration of emotions in schizophrenia: reduced regulatory effects, *Schizophr. Res.* **122**, 136–143. DOI:10.1016/j.schres.2010.04.010.

de Jong, J. J., de Gelder, B. and Hodiamont, P. P. G. (2013). Sensory processing, neurocognition, and social cognition in schizophrenia: towards a cohesive cognitive model, *Schizophr. Res.* **146**, 209–216. DOI:10.1016/j.schres.2013.02.034.

de Niear, M. A., Gupta, P. B., Baum, S. H. and Wallace, M. T. (2018). Perceptual training enhances temporal acuity for multisensory speech, *Neurobiol. Learn. Mem.* **147**, 9–17. DOI:10.1016/j.nlm.2017.10.016.

Diamond, E. and Zhang, Y. (2016). Cortical processing of phonetic and emotional information in speech: a cross-modal priming study, *Neuropsychologia* **82**, 110–122. DOI:10.1016/j.neuropsychologia.2016.01.019.

Dickinson, D., Iannone, V. N., Wilk, C. M. and Gold, J. M. (2004). General and specific cognitive deficits in schizophrenia, *Biol. Psychiatry* **55**, 826–833. DOI:10.1016/j.biopsych.2003.12.010.

Dondaine, T., Robert, G., Péron, J., Grandjean, D., Vérin, M., Drapier, D. and Millet, B. (2014). Biases in facial and vocal emotion recognition in chronic schizophrenia, *Front. Psychol.* **5**, 900. DOI:10.3389/fpsyg.2014.00900.

Driver, J. and Noesselt, T. (2008). Multisensory interplay reveals crossmodal influences on 'sensory-specific' brain regions, neural responses, and judgments, *Neuron* **57**, 11–23. DOI:10.1016/j.neuron.2007.12.013.

Dzafic, I., Burianová, H., Martin, A. K. and Mowry, B. (2018). Neural correlates of dynamic emotion perception in schizophrenia and the influence of prior expectations, *Schizophr. Res.* **202**, 129–137. DOI:10.1016/j.schres.2018.06.015.

Edwards, J., Jackson, H. J. and Pattison, P. E. (2002). Emotion recognition via facial expression and affective prosody in schizophrenia: a methodological review, *Clin. Psychol. Rev.* **22**, 789–832. DOI:10.1016/S0272-7358(02)00130-7.

Egger, M., Ley, M. and Hanke, S. (2019). Emotion recognition from physiological signal analysis: a review, *Electron. Notes Theor. Comput. Sci.* **343**, 35–55. DOI:10.1016/j.entcs.2019.04.009.

Ekman, P. and Friesen, W. V. (1976). Measuring facial movement, *Environ. Psychol. Nonverbal Behav.* **1**, 56–75. DOI:10.1007/BF01115465.

Erber, N. P. (1969). Interaction of audition and vision in the recognition of oral speech stimuli, *J. Speech. Lang. Hear. Res.* **12**(2), 423–425. DOI:10.1044/jshr.1202.423.

Feldman, J. I., Dunham, K., Cassidy, M., Wallace, M. T., Liu, Y. and Woynaroski, T. G. (2018). Audiovisual multisensory integration in individuals with autism spectrum disorder: a systematic review and meta-analysis, *Neurosci. Biobehav. Rev.* **95**, 220–234. DOI:10.1016/j.neubiorev.2018.09.020.

Filippi, P., Ocklenburg, S., Bowling, D. L., Heege, L., Güntürkün, O., Newen, A. and de Boer, B. (2017). More than words (and faces): evidence for a Stroop effect of prosody in emotion word processing, *Cogn. Emot.* **31**, 879–891. DOI:10.1080/02699931.2016.1177489.

Fiszdon, J. M. and Bell, M. D. (2009). Effects of presentation modality and valence on affect recognition performance in schizophrenia and healthy controls, *Psychiatry Res.* **170**, 114–118. DOI:10.1016/j.psychres.2008.11.014.

Fiszdon, J. M. and Johannesen, J. K. (2010). Functional significance of preserved affect recognition in schizophrenia, *Psychiatry Res.* **176**, 120–125. DOI:10.1016/j.psychres.2009.08.006.

Föcker, J. and Röder, B. (2019). Event-related potentials reveal evidence for late integration of emotional prosody and facial expression in dynamic stimuli: an ERP study, *Multisens. Res.* **32**, 473–497. DOI:10.1163/22134808-20191332.

Fox, M. D. and Greicius, M. (2010). Clinical applications of resting state functional connectivity, *Front. Syst. Neurosci.* **4**, 19. DOI:10.3389/fnsys.2010.00019.

Froming, K., Levy, M., Schaffer, S. and Ekman, P. (2006). *The Comprehensive Affect Testing System*. Psychology Software, Inc., Pittsburgh, PA, USA.

Gee, D. G., Karlsgodt, K. H., van Erp, T. G. M., Bearden, C. E., Lieberman, M. D., Belger, A., Perkins, D. O., Olvet, D. M., Cornblatt, B. A., Constable, T., Woods, S. W., Addington, J., Cadenhead, K. S., McGlashan, T. H., Seidman, L. J., Tsuang, M. T., Walker, E. F. and Cannon, T. D., on behalf of the NAPLS Consortium (2012). Altered age-related trajectories of amygdala–prefrontal circuitry in adolescents at clinical high risk for psychosis: a preliminary study, *Schizophr. Res.* **134**, 1–9. DOI:10.1016/j.schres.2011.10.005.

Giannitelli, M., Xavier, J., François, A., Bodeau, N., Laurent, C., Cohen, D. and Chaby, L. (2015). Facial, vocal and cross-modal emotion processing in early-onset schizophrenia spectrum disorders, *Schizophr. Res.* **168**, 252–259. DOI:10.1016/j.schres.2015.07.039.

Green, A. M. and Angelaki, D. E. (2010). Multisensory integration: resolving sensory ambiguities to build novel representations, *Curr. Opin. Neurobiol.* **20**, 353–360. DOI:10.1016/j.conb.2010.04.009.

Green, M. F. and Leitman, D. I. (2008). Social cognition in schizophrenia, *Schizophr. Bull.* **34**, 670–672. DOI:10.1093/schbul/sbn045.

Green, M. F., Olivier, B., Crawley, J. N., Penn, D. L. and Silverstein, S. (2005). Social cognition in schizophrenia: recommendations from the Measurement and Treatment Research to Improve Cognition in Schizophrenia New Approaches Conference, *Schizophr. Bull.* **31**, 882–887. DOI:10.1093/schbul/sbi049.

Green, M. F., Horan, W. P., Lee, J., McCleery, A., Reddy, L. F. and Wynn, J. K. (2018). Social disconnection in schizophrenia and the general community, *Schizophr. Bull.* **44**, 242–249. DOI:10.1093/schbul/sbx082.

Hall, E. T. (1976). *Beyond Culture*. Anchor Press/Doubleday, Garden City/New York, NY, USA.

Hall, J., Harris, J. M., Sprengelmeyer, R., Sprengelmeyer, A., Young, A. W., Santos, I. M., Johnstone, E. C. and Lawrie, S. M. (2004). Social cognition and face processing in schizophrenia, *Br. J. Psychiat.* **185**, 169–170. DOI:10.1192/bjp.185.2.169.

Hanlon, F. M., Weisend, M. P., Yeo, R. A., Huang, M., Lee, R. R., Thoma, R. J., Moses, S. N., Paulson, K. M., Miller, G. A. and Cañive, J. M. (2005). A specific test of hippocampal deficit in schizophrenia, *Behav. Neurosci.* **119**, 863–875. DOI:10.1037/0735-7044.119.4.863.

Hardison, D. M. (2010). Visual and auditory input in second-language speech processing, *Lang. Teach.* **43**, 84–95. DOI:10.1017/S0261444809990176.

Hassett, D. D. and Curwood, J. S. (2009). Theories and practices of multimodal education: the instructional dynamics of picture books and primary classrooms, *Read. Teach.* **63**, 270–282. DOI:10.1598/RT.63.4.2.

Hazan, V., Sennema, A., Iba, M. and Faulkner, A. (2005). Effect of audiovisual perceptual training on the perception and production of consonants by Japanese learners of English, *Speech Commun.* **47**, 360–378. DOI:10.1016/j.specom.2005.04.007.

Heald, S. L. M. and Nusbaum, H. C. (2014). Talker variability in audio–visual speech perception, *Front. Psychol.* **5**, 698. DOI:10.3389/fpsyg.2014.00698.

Heinonen-Guzejev, M., Klyuchko, M., Heikkilä, K., Spinosa, V., Tervaniemi, M. and Brattico, E. (2014). Noise sensitivity modulates the auditory-cortex discrimination of sound feature changes, in: *Inter.Noise 2014*, Melbourne, Australia.

Heinrichs, R. W. and Zakzanis, K. K. (1998). Neurocognitive deficit in schizophrenia: a quantitative review of the evidence, *Neuropsychology* **12**, 426–445. DOI:10.1037/0894-4105.12.3.426.

Herbener, E. S., Rosen, C., Khine, T. and Sweeney, J. A. (2007). Failure of positive but not negative emotional valence to enhance memory in schizophrenia, *J. Abnorm. Psychol.* **116**, 43–55. DOI:10.1037/0021-843X.116.1.43.

Herniman, S. E., Allott, K. A., Killackey, E., Hester, R. and Cotton, S. M. (2017). The effect of comorbid depression on facial and prosody emotion recognition in first-episode schizophrenia spectrum, *J. Affect. Disord.* **208**, 223–229. DOI:10.1016/j.jad.2016.08.068.

Hill, S. K., Schuepbach, D., Herbener, E. S., Keshavan, M. S. and Sweeney, J. A. (2004). Pretreatment and longitudinal studies of neuropsychological deficits in antipsychotic-naïve patients with schizophrenia, *Schizophr. Res.* **68**, 49–63. DOI:10.1016/S0920-9964(03)00213-5.

Hoertnagl, C. M., Yalcin-Siedentopf, N., Baumgartner, S., Biedermann, F., Deisenhammer, E. A., Hausmann, A., Kaufmann, A., Kemmler, G., Mühlbacher, M., Rauch, A.-S., Fleischhacker, W. W. and Hofer, A. (2014). Affective prosody perception in symptomatically remitted patients with schizophrenia and bipolar disorder, *Schizophr. Res.* **158**, 100–104.

Holt, R. F., Kirk, K. I. and Hay-McCutcheon, M. (2011). Assessing multimodal spoken word-in-sentence recognition in children with normal hearing and children with cochlear implants, *J. Speech. Lang. Hear. Res.* **54**, 632–657. DOI:10.1044/1092-4388(2010/09-0148).

Huang, C. L.-C., Hsiao, S., Hwu, H.-G. and Howng, S.-L. (2013). Are there differential deficits in facial emotion recognition between paranoid and non-paranoid schizophrenia? A signal detection analysis, *Psychiatry Res.* **209**, 424–430. DOI:10.1016/j.psychres.2013.03.026.

Iacoviello, B. M., Wu, G., Alvarez, E., Huryk, K., Collins, K. A., Murrough, J. W., Iosifescu, D. V. and Charney, D. S. (2014). Cognitive-emotional training as an intervention for major depressive disorder, *Depress. Anxiety* **31**, 699–706. DOI:10.1002/da.22266.

Irwin, J. and DiBlasi, L. (2017). Audiovisual speech perception: a new approach and implications for clinical populations, *Lang. Linguist. Compass* **11**, 77–91. DOI:10.1111/lnc3.12237.

Ito, F., Matsumoto, K., Miyakoshi, T., Ohmuro, N., Uchida, T. and Matsuoka, H. (2013). Emotional processing during speech communication and positive symptoms in schizophrenia, *Psychiatry Clin. Neurosci.* **67**, 526–531. DOI:10.1111/pcn.12103.

Javitt, D. C. (2009). Sensory processing in schizophrenia: neither simple nor intact, *Schizophr. Bull.* **35**, 1059–1064. DOI:10.1093/schbul/sbp110.

Kantrowitz, J. T., Leitman, D. I., Lehrfeld, J. M., Laukka, P., Juslin, P. N., Butler, P. D., Silipo, G. and Javitt, D. C. (2013). Reduction in tonal discriminations predicts receptive emotion processing deficits in schizophrenia and schizoaffective disorder, *Schizophr. Bull.* **39**, 86–93. DOI:10.1093/schbul/sbr060.

Kantrowitz, J. T., Hoptman, M. J., Leitman, D. I., Moreno-Ortega, M., Lehrfeld, J. M., Dias, E., Sehatpour, P., Laukka, P., Silipo, G. and Javitt, D. C. (2015). Neural substrates of auditory emotion recognition deficits in schizophrenia, *J. Neurosci.* **35**, 14909–14921. DOI:10.1523/JNEUROSCI.4603-14.2015.

Kawano, M., Oshibuchi, H., Kawano, T., Muraoka, H., Tsutsumi, T., Yamada, M., Inada, K. and Ishigooka, J. (2016). Dopamine dynamics during emotional cognitive processing: implications of the specific actions of clozapine compared with haloperidol, *Eur. J. Pharmacol.* **781**, 148–156. DOI:10.1016/j.ejphar.2016.04.013.

Kerr, S. L. and Neale, J. M. (1993). Emotion perception in schizophrenia: specific deficit or further evidence of generalized poor performance?, *J. Abnorm. Psychol.* **102**, 312–318. DOI:10.1037/0021-843X.102.2.312.

Koelewijn, T., Bronkhorst, A. and Theeuwes, J. (2010). Attention and the multiple stages of multisensory integration: a review of audiovisual studies, *Acta Psychol. (Amst.)* **134**, 372–384. DOI:10.1016/j.actpsy.2010.03.010.

Kohler, C. G., Bilker, W., Hagendoorn, M., Gur, R. E. and Gur, R. C. (2000). Emotion recognition deficit in schizophrenia: association with symptomatology and cognition, *Biol. Psychiatry* **48**, 127–136. DOI:10.1016/S0006-3223(00)00847-7.

Kohler, C. G., Turner, T. H., Bilker, W. B., Brensinger, C. M., Siegel, S. J., Kanes, S. J., Gur, R. E. and Gur, R. C. (2003). Facial emotion recognition in schizophrenia: intensity effects and error pattern, *Am. J. Psychiatry* **160**, 1768–1774. DOI:10.1176/appi.ajp.160.10.1768.

Kohler, C. G., Walker, J. B., Martin, E. A., Healey, K. M. and Moberg, P. J. (2010). Facial emotion perception in schizophrenia: a meta-analytic review, *Schizophr. Bull.* **36**, 1009–1019. DOI:10.1093/schbul/sbn192.

Kraus, M. S., Walker, T. M., Jarskog, L. F., Millet, R. A. and Keefe, R. S. E. (2019). Basic auditory processing deficits and their association with auditory emotion recognition in schizophrenia, *Schizophr. Res.* **204**, 155–161. DOI:10.1016/j.schres.2018.08.031.

Kret, M. E. and Ploeger, A. (2015). Emotion processing deficits: a liability spectrum providing insight into comorbidity of mental disorders, *Neurosci. Biobehav. Rev.* **52**, 153–171. DOI:10.1016/j.neubiorev.2015.02.011.

Kucharska-Pietura, K., David, A. S., Masiak, M. and Phillips, M. L. (2005). Perception of facial and vocal affect by people with schizophrenia in early and late stages of illness, *Br. J. Psychiatry* **187**, 523–528. DOI:10.1192/bjp.187.6.523.

Laird, A. R., Eickhoff, S. B., Kurth, F., Fox, P. M., Uecker, A. M., Turner, J. A., Robinson, J. L., Lancaster, J. L. and Fox, P. T. (2009a). ALE meta-analysis workflows via the BrainMap database: progress towards a probabilistic functional brain atlas, *Front. Neuroinform.* **3**, 23. DOI:10.3389/neuro.11.023.2009.

Laird, A. R., Eickhoff, S. B., Li, K., Robin, D. A., Glahn, D. C. and Fox, P. T. (2009b). Investigating the functional heterogeneity of the default mode network using coordinate-based meta-analytic modeling, *J. Neurosci.* **29**, 14496–14505. DOI:10.1523/JNEUROSCI.4004-09.2009.

Leitman, D. I., Foxe, J. J., Butler, P. D., Saperstein, A., Revheim, N. and Javitt, D. C. (2005). Sensory contributions to impaired prosodic processing in schizophrenia, *Biol. Psychiatry* **58**, 56–61. DOI:10.1016/j.biopsych.2005.02.034.

Leitman, D. I., Hoptman, M. J., Foxe, J. J., Saccente, E., Wylie, G. R., Nierenberg, J., Jalbrzikowski, M., Lim, K. O. and Javitt, D. C. (2007). The neural substrates of impaired prosodic detection in schizophrenia and its sensorial antecedents, *Am. J. Psychiatry* **164**, 474–482.

Lencz, T., Smith, C. W., McLaughlin, D., Auther, A., Nakayama, E., Hovey, L. and Cornblatt, B. A. (2006). Generalized and specific neurocognitive deficits in prodromal schizophrenia, *Biol. Psychiatry* **59**, 863–871. DOI:10.1016/j.biopsych.2005.09.005.

Lin, Y. and Ding, H. (2019). Multisensory integration of emotions in a face–prosody–semantics Stroop task, in: *ICMI '19: NeuroManagement and Intelligent Computing Method on Multimodal Interaction*, Suzhou, China, pp. 1–5. DOI:10.1145/3357160.3357668.

Lin, Y., Ding, H. and Zhang, Y. (2018). Emotional prosody processing in schizophrenic patients: a selective review and meta-analysis, *J. Clin. Med.* **7**, 363. DOI:10.3390/jcm7100363.

Lin, Y., Ding, H. and Zhang, Y. (2020). Prosody dominates over semantics in emotion word processing: evidence from cross-channel and cross-modal Stroop effects, *J. Speech. Lang. Hear. Res.* **63**, 896–912. DOI:10.1044/2020_jslhr-19-00258.

Liu, P., Rigoulot, S. and Pell, M. D. (2015). Culture modulates the brain response to human expressions of emotion: electrophysiological evidence, *Neuropsychologia* **67**, 1–13. DOI:10.1016/j.neuropsychologia.2014.11.034.

Liu, T., Pinheiro, A. P., Zhao, Z., Nestor, P. G., McCarley, R. W. and Niznikiewicz, M. (2016). Simultaneous face and voice processing in schizophrenia, *Behav. Brain Res.* **305**, 76–86. DOI:10.1016/j.bbr.2016.01.039.

Lynall, M.-E., Bassett, D. S., Kerwin, R., McKenna, P. J., Kitzbichler, M., Muller, U. and Bullmore, E. (2010). Functional connectivity and brain networks in schizophrenia, *J. Neurosci.* **30**, 9477–9487. DOI:10.1523/JNEUROSCI.0333-10.2010.

Lysaker, P. H., George, S., Chaudoin–Patzoldt, K. A., Pec, O., Bob, P., Leonhardt, B. L., Vohs, J. L., James, A. V., Wickett, A., Buck, K. D. and Dimaggio, G. (2017). Contrasting metacognitive, social cognitive and alexithymia profiles in adults with borderline personality disorder, schizophrenia and substance use disorder, *Psychiatry Res.* **257**, 393–399. DOI:10.1016/j.psychres.2017.08.001.

Mahoney, J. R., Li, P. C. C., Oh-Park, M., Verghese, J. and Holtzer, R. (2011). Multisensory integration across the senses in young and old adults, *Brain Res.* **1426**, 43–53. DOI:10.1016/j.brainres.2011.09.017.

Mandal, M. K., Pandey, R. and Prasad, A. B. (1998). Facial expressions of emotions and schizophrenia: a review, *Schizophr. Bull.* **24**, 399–412. DOI:10.1093/oxfordjournals.schbul.a033335.

Mangelinckx, C., Belge, J. B., Maurage, P. and Constant, E. (2017). Impaired facial and vocal emotion decoding in schizophrenia is underpinned by basic perceptivo-motor deficits, *Cogn. Neuropsychiatry* **22**, 461–467. DOI:10.1080/13546805.2017.1382342.

McGurk, H. and MacDonald, J. (1976). Hearing lips and seeing voices, *Nature* **264**, 746–748. DOI:10.1038/264746a0.

McLachlan, N. M., Phillips, D. S., Rossell, S. L. and Wilson, S. J. (2013). Auditory processing and hallucinations in schizophrenia, *Schizophr. Res.* **150**, 380–385. DOI:10.1016/j.schres.2013.08.039.

Mitchell, R. L. C. and Rossell, S. L. (2014). Perception of emotion-related conflict in human communications: what are the effects of schizophrenia?, *Psychiatry Res.* **220**, 135–144. DOI:10.1016/j.psychres.2014.07.077.

Mohamed, S., Paulsen, J. S., O'Leary, D., Arndt, S. and Andreasen, N. (1999). Generalized cognitive deficits in schizophrenia: a study of first-episode patients, *Arch. Gen. Psychiatry* **56**, 749–754. DOI:10.1001/archpsyc.56.8.749.

Mohr, D. C., Duffecy, J., Jin, L., Ludman, E. J., Lewis, A., Begale, M. and McCarthy Jr, M. (2010). Multimodal e-mental health treatment for depression: a feasibility trial, *J. Med. Internet Res.* **12**, e48. DOI:10.2196/jmir.1370.

Moradi, S., Wahlin, A., Hällgren, M., Rönnberg, J. and Lidestam, B. (2017). The efficacy of short-term gated audiovisual speech training for improving auditory sentence identification in noise in elderly hearing aid users, *Front. Psychol.* **8**, 368. DOI:10.3389/fpsyg.2017.00368.

Morgane, P. J., Galler, J. R. and Mokler, D. J. (2005). A review of systems and networks of the limbic forebrain/limbic midbrain, *Prog. Neurobiol.* **75**, 143–160. DOI:10.1016/j.pneurobio.2005.01.001.

Müller, V. I., Habel, U., Derntl, B., Schneider, F., Zilles, K., Turetsky, B. I. and Eickhoff, S. B. (2011). Incongruence effects in crossmodal emotional integration, *Neuroimage* **54**, 2257–2266. DOI:10.1016/j.neuroimage.2010.10.047.

Müller, V. I., Cieslik, E. C., Laird, A. R., Fox, P. T. and Eickhoff, S. B. (2013). Dysregulated left inferior parietal activity in schizophrenia and depression: functional connectivity and characterization, *Front. Hum. Neurosci.* **7**, 268. DOI:10.3389/fnhum.2013.00268.

Müller, V. I., Kellermann, T. S., Seligman, S. C., Turetsky, B. I. and Eickhoff, S. B. (2014). Modulation of affective face processing deficits in schizophrenia by congruent emotional sounds, *Soc. Cogn. Affect. Neurosci.* **9**, 436–444. DOI:10.1093/scan/nss107.

Murray, M. M., Eardley, A. F., Edginton, T., Oyekan, R., Smyth, E. and Matusz, P. J. (2018). Sensory dominance and multisensory integration as screening tools in aging, *Sci. Rep.* **8**, 8901. DOI:10.1038/s41598-018-27288-2.

National Institute of Mental Health (2016). *Schizophrenia (NIH Publication No. 19-MH-8082)*. Retrieved from https://www.nimh.nih.gov/health/publications/schizophrenia/19-mh-8082-schizophrenia_155669.pdf.

Niedenthal, P. M. (2007). Embodying emotion, *Science* **316**, 1002–1005. DOI:10.1126/science.1136930.

Nuechterlein, K. H., Ventura, J., Subotnik, K. L., Hayata, J. N., Medalia, A. and Bell, M. D. (2014). Developing a cognitive training strategy for first-episode schizophrenia: integrating bottom-up and top-down approaches, *Am. J. Psychiatr. Rehabil.* **17**, 225–253. DOI:10.1080/15487768.2014.935674.

Paulmann, S. and Pell, M. D. (2011). Is there an advantage for recognizing multi-modal emotional stimuli?, *Motiv. Emot.* **35**(2), 192–201. DOI:10.1007/s11031-011-9206-0.

Pawelczyk, A., Kotlicka-Antczak, M., Lojek, E., Ruszpel, A. and Pawelczyk, T. (2018). Schizophrenia patients have higher-order language and extralinguistic impairments, *Schizophr. Res.* **192**, 274–280. DOI:10.1016/j.schres.2017.04.030.

Pearl, D., Yodashkin-Porat, D., Katz, N., Valevski, A., Aizenberg, D., Sigler, M., Weizman, A. and Kikinzon, L. (2009). Differences in audiovisual integration, as measured by McGurk phenomenon, among adult and adolescent patients with schizophrenia and age-matched healthy control groups, *Compr. Psychiatry* **50**, 186–192. DOI:10.1016/j.comppsych.2008.06.004.

Pell, M. D., Jaywant, A., Monetta, L. and Kotz, S. A. (2011). Emotional speech processing: disentangling the effects of prosody and semantic cues, *Cogn. Emot.* **25**, 834–853. DOI:10.1080/02699931.2010.516915.

Penn, D. L., Combs, D. R., Ritchie, M., Francis, J., Cassisi, J., Morris, S. and Townsend, M. (2000). Emotion recognition in schizophrenia: further investigation of generalized versus specific deficit models, *J. Abnorm. Psychol.* **109**, 512–516. DOI:10.1037/0021-843X.109.3.512.

Penn, D. L., Sanna, L. J. and Roberts, D. L. (2008). Social cognition in schizophrenia: an overview, *Schizophr. Bull.* **34**, 408–411. DOI:10.1093/schbul/sbn014.

Phillips, M. L. and Kupfer, D. J. (2013). Bipolar disorder diagnosis: challenges and future directions, *The Lancet* **381**, 1663–1671. DOI:10.1016/S0140-6736(13)60989-7.

Pinheiro, A. P., del Re, E., Mezin, J., Nestor, P. G., Rauber, A., McCarley, R. W., Gonçalves, Ó. F. and Niznikiewicz, M. A. (2013). Sensory-based and higher-order operations contribute to abnormal emotional prosody processing in schizophrenia: an electrophysiological investigation, *Psychol. Med.* **43**, 603–618. DOI:10.1017/S003329171200133X.

Postmes, L., Sno, H. N., Goedhart, S., van der Stel, J., Heering, H. D. and de Haan, L. (2014). Schizophrenia as a self-disorder due to perceptual incoherence, *Schizophr. Res.* **152**, 41–50. DOI:10.1016/j.schres.2013.07.027.

Pourtois, G., de Gelder, B., Vroomen, J., Rossion, B. and Crommelinck, M. (2000). The time-course of intermodal binding between seeing and hearing affective information, *NeuroReport* **11**, 1329–1333.

Putnam, K. M. and Kring, A. M. (2007). Accuracy and intensity of posed emotional expressions in unmedicated schizophrenia patients: vocal and facial channels, *Psychiatry Res.* **151**, 67–76. DOI:10.1016/j.psychres.2006.09.010.

Rao, A., Rishiq, D., Yu, L., Zhang, Y. and Abrams, H. (2017). Neural correlates of selective attention with hearing aid use followed by ReadMyQuips auditory training program, *Ear Hear.* **38**, 28–41. DOI:10.1097/AUD.0000000000000348.

Razafimandimby, A., Hervé, P.-Y., Marzloff, V., Brazo, P., Tzourio-Mazoyer, N. and Dollfus, S. (2016). Functional deficit of the medial prefrontal cortex during emotional sentence attribution in schizophrenia, *Schizophr. Res.* **178**, 86–93. DOI:10.1016/j.schres.2016.09.004.

Regenbogen, C., Kellermann, T., Seubert, J., Schneider, D. A., Gur, R. E., Derntl, B., Schneider, F. and Habel, U. (2015). Neural responses to dynamic multimodal stimuli and pathology-specific impairments of social cognition in schizophrenia and depression, *Br. J. Psychiatry* **206**, 198–205. DOI:10.1192/bjp.bp.113.143040.

Robinson, J. L., Laird, A. R., Glahn, D. C., Lovallo, W. R. and Fox, P. T. (2010). Metaanalytic connectivity modeling: delineating the functional connectivity of the human amygdala, *Hum. Brain Mapp.* **31**, 173–184. DOI:10.1002/hbm.20854.

Ross, L. A., Saint-Amour, D., Leavitt, V. M., Molholm, S., Javitt, D. C. and Foxe, J. J. (2007). Impaired multisensory processing in schizophrenia: deficits in the visual enhancement of speech comprehension under noisy environmental conditions, *Schizophr. Res.* **97**, 173–183. DOI:10.1016/j.schres.2007.08.008.

Ross, R. G., Heinlein, S. and Tregellas, H. (2006). High rates of comorbidity are found in childhood-onset schizophrenia, *Schizophr. Res.* **88**, 90–95. DOI:10.1016/j.schres.2006.07.006.

Rossell, S. L. and Boundy, C. L. (2005). Are auditory–verbal hallucinations associated with auditory affective processing deficits?, *Schizophr. Res.* **78**, 95–106. DOI:10.1016/j.schres.2005.06.002.

Schaffer, S. G., Wisniewski, A., Dahdah, M. and Froming, K. B. (2009). The comprehensive affect testing system — abbreviated: effects of age on performance, *Arch. Clin. Neuropsychol.* **24**, 89–104. DOI:10.1093/arclin/acp012.

Schwartz, R. and Pell, M. D. (2012). Emotional speech processing at the intersection of prosody and semantics, *PLoS ONE* **7**, e47279. DOI:10.1371/journal.pone.0047279.

Sekiyama, K., Soshi, T. and Sakamoto, S. (2014). Enhanced audiovisual integration with aging in speech perception: a heightened McGurk effect in older adults, *Front. Psychol.* **5**, 323. DOI:10.3389/fpsyg.2014.00323.

Seubert, J., Loughead, J., Kellermann, T., Boers, F., Brensinger, C. M. and Habel, U. (2010). Multisensory integration of emotionally valenced olfactory–visual information in patients with schizophrenia and healthy controls, *J. Psychiatry Neurosci.* **35**, 185–194. DOI:10.1503/jpn.090094.

Shams, L., Kamitani, Y. and Shimojo, S. (2000). Illusions: what you see is what you hear, *Nature* **408**, 788. DOI:10.1038/35048669.

Shea, T. L., Sergejew, A. A., Burnham, D., Jones, C., Rossell, S. L., Copolov, D. L. and Egan, G. F. (2007). Emotional prosodic processing in auditory hallucinations, *Schizophr. Res.* **90**, 214–220. DOI:10.1016/j.schres.2006.09.021.

Silver, H., Shlomo, N., Turner, T. and Gur, R. C. (2002). Perception of happy and sad facial expressions in chronic schizophrenia: evidence for two evaluative systems, *Schizophr. Res.* **55**, 171–177. DOI:10.1016/S0920-9964(01)00208-0.

Silver, H., Bilker, W. and Goodman, C. (2009). Impaired recognition of happy, sad and neutral expressions in schizophrenia is emotion, but not valence, specific and context dependent, *Psychiatry Res.* **169**, 101–106. DOI:10.1016/j.psychres.2008.11.017.

Silverstein, S. M. (2008). Measuring specific, rather than generalized, cognitive deficits and maximizing between-group effect size in studies of cognition and cognitive change, *Schizophr. Bull.* **34**, 645–655. DOI:10.1093/schbul/sbn032.

Silverstein, S. M., Wang, Y. and Keane, B. P. (2012). Cognitive and neuroplasticity mechanisms by which congenital or early blindness may confer a protective effect against schizophrenia, *Front. Psychol.* **3**, 624. DOI:10.3389/fpsyg.2012.00624.

Simpson, C., Pinkham, A. E., Kelsven, S. and Sasson, N. J. (2013). Emotion recognition abilities across stimulus modalities in schizophrenia and the role of visual attention, *Schizophr. Res.* **151**, 102–106. DOI:10.1016/j.schres.2013.09.026.

Sinnett, S., Soto-Faraco, S. and Spence, C. (2008). The co-occurrence of multisensory competition and facilitation, *Acta Psychol. (Amst.)* **128**, 153–161. DOI:10.1016/j.actpsy.2007.12.002.

Spence, C. (2009). Explaining the Colavita visual dominance effect, *Prog. Brain Res.* **176**, 245–258. DOI:10.1016/S0079-6123(09)17615-X.

Spence, C., Parise, C. and Chen, Y. C. (2011). The Colavita visual dominance effect, in: *The Neural Bases of Multisensory Processes*, M. M. Murray and M. T. Wallace (Eds), pp. 523–550. CRC Press/Taylor & Francis, Boca Raton, FL, USA.

Stein, B. E., Stanford, T. R. and Rowland, B. A. (2009). The neural basis of multisensory integration in the midbrain: its organization and maturation, *Hear. Res.* **258**, 4–15. DOI:10.1016/j.heares.2009.03.012.

Stekelenburg, J. J., Maes, J. P., Van Gool, A. R., Sitskoorn, M. and Vroomen, J. (2013). Deficient multisensory integration in schizophrenia: an event-related potential study, *Schizophr. Res.* **147**, 253–261. DOI:10.1016/j.schres.2013.04.038.

Sterea, R. (2015). The relationship between social cognition and functional outcomes in schizophrenia, *Procedia Soc. Behav. Sci.* **187**, 256–260.

Stone, D. B., Urrea, L. J., Aine, C. J., Bustillo, J. R., Clark, V. P. and Stephen, J. M. (2011). Unisensory processing and multisensory integration in schizophrenia: a high-density electrical mapping study, *Neuropsychologia* **49**, 3178–3187. DOI:10.1016/j.neuropsychologia.2011.07.017.

Straube, B., Green, A., Sass, K. and Kircher, T. (2014). Superior temporal sulcus disconnectivity during processing of metaphoric gestures in schizophrenia, *Schizophr. Bull.* **40**, 936–944. DOI:10.1093/schbul/sbt110.

Strauss, G. P., Allen, D. N., Duke, L. A., Ross, S. A. and Schwartz, J. (2008). Automatic affective processing impairments in patients with deficit syndrome schizophrenia, *Schizophr. Res.* **102**, 76–87. DOI:10.1016/j.schres.2008.01.014.

Stroop, J. R. (1935). Studies of interferences in serial verbal reactions, *Exp. Psychol.* **18**, 643–662. DOI:10.1037/h0054651.

Summerfield, Q. (1979). Use of visual information for phonetic perception, *Phonetica* **36**, 314–331. DOI:10.1159/000259969.

Szycik, G. R., Ye, Z., Mohammadi, B., Dillo, W., te Wildt, B. T., Samii, A., Frieling, H., Bleich, S. and Münte, T. F. (2013). Maladaptive connectivity of Broca's area in schizophrenia during audiovisual speech perception: an fMRI study, *Neuroscience* **253**, 274–282. DOI:10.1016/j.neuroscience.2013.08.041.

Talsma, D., Senkowski, D., Soto-Faraco, S. and Woldorff, M. G. (2010). The multifaceted interplay between attention and multisensory integration, *Trends Cogn. Sci.* **14**, 400–410. DOI:10.1016/j.tics.2010.06.008.

Trémeau, F. (2006). A review of emotion deficits in schizophrenia, *Dialogues Clin. Neurosci.* **8**, 59–70.

Tseng, H.-H., Chen, S.-H. and Huang, Y.-L. (2012). Dissimilar deficits of facial and prosodic emotion recognition in patients with schizophrenia in Taiwan, *Taiwanese J. Psychiatry* **26**, 105–113.

Tseng, H.-H., Chen, S.-H., Liu, C.-M., Howes, O., Huang, Y.-L., Hsieh, M. H., Liu, C.-C., Shan, J.-C., Lin, Y.-T. and Hwu, H.-G. (2013). Facial and prosodic emotion recognition deficits associate with specific clusters of psychotic symptoms in schizophrenia, *PLoS ONE* **8**, e66571. DOI:10.1371/journal.pone.0066571.

Tseng, H.-H., Bossong, M. G., Modinos, G., Chen, K.-M., McGuire, P. and Allen, P. (2015). A systematic review of multisensory cognitive-affective integration in schizophrenia, *Neurosci. Biobehav. Rev.* **55**, 444–452. DOI:10.1016/j.neubiorev.2015.04.019.

Turetsky, B. I., Kohler, C. G., Indersmitten, T., Bhati, M. T., Charbonnier, D. and Gur, R. C. (2007). Facial emotion recognition in schizophrenia: when and why does it go awry?, *Schizophr. Res.* **94**, 253–263. DOI:10.1016/j.schres.2007.05.001.

Van den Stock, J., de Jong, S. J., Hodiamont, P. P. G. and de Gelder, B. (2011). Perceiving emotions from bodily expressions and multisensory integration of emotion cues in schizophrenia, *Soc. Neurosci.* **6**, 537–547. DOI:10.1080/17470919.2011.568790.

Van Rheenen, T. E. and Rossell, S. L. (2014). Multimodal emotion integration in bipolar disorder: an investigation of involuntary cross-modal influences between facial and prosodic channels, *J. Int. Neuropsychol. Soc.* **20**, 525–533. DOI:10.1017/S1355617714000253.

Varga, E., Endre, S., Bugya, T., Tényi, T. and Herold, R. (2018a). Community-based psychosocial treatment has an impact on social processing and functional outcome in schizophrenia, *Front. Psychiatry* **9**, 247. DOI:10.3389/fpsyt.2018.00247.

Varga, E., Herold, R., Tényi, T. and Bugya, T. (2018b). Social Cognition Analyzer Application (SCAN) — a new approach to analyze social cognition in schizophrenia, *Eur. Neuropsychopharmacol.* **28**, S55–S56.

Vercammen, A., Morris, R., Green, M. J., Lenroot, R., Kulkarni, J., Carr, V. J., Weickert, C. S. and Weickert, T. W. (2012). Reduced neural activity of the prefrontal cognitive control circuitry during response inhibition to negative words in people with schizophrenia, *J. Psychiatry Neurosci.* **37**, 379–388. DOI:10.1503/jpn.110088.

Veuillet, E., Magnan, A., Ecalle, J., Thai-Van, H. and Collet, L. (2007). Auditory processing disorder in children with reading disabilities: effect of audiovisual training, *Brain* **130**, 2915–2928. DOI:10.1093/brain/awm235.

Vilko, J. P. P. and Hallikas, J. M. (2012). Risk assessment in multimodal supply chains, *Int. J. Prod. Econ.* **140**, 586–595. DOI:10.1016/j.ijpe.2011.09.010.

Vogel, B., Brück, C., Jacob, H., Eberle, M. and Wildgruber, D. (2016). Integration of verbal and nonverbal emotional signals in patients with schizophrenia: decreased nonverbal dominance, *Psychiatry Res.* **241**, 98–103. DOI:10.1016/j.psychres.2016.03.050.

Vroomen, J. and de Gelder, B. (2000). Sound enhances visual perception: cross-modal effects of auditory organization on vision, *J. Exp. Psychol. Hum. Percept. Perform.* **26**, 1583–1590. DOI:10.1037/0096-1523.26.5.1583.

Weniger, G., Lange, C., Rüther, E. and Irle, E. (2004). Differential impairments of facial affect recognition in schizophrenia subtypes and major depression, *Psychiatry Res.* **128**, 135–146. DOI:10.1016/j.psychres.2003.12.027.

Wildgruber, D., Ethofer, T., Grandjean, D. and Kreifelts, B. (2009). A cerebral network model of speech prosody comprehension, *Int. J. Speech Lang. Pathol.* **11**, 277–281. DOI:10.1080/17549500902943043.

Williams, L. E., Light, G. A., Braff, D. L. and Ramachandran, V. S. (2010). Reduced multisensory integration in patients with schizophrenia on a target detection task, *Neuropsychologia* **48**, 3128–3136. DOI:10.1016/j.neuropsychologia.2010.06.028.

Williams, L. M., Das, P., Liddell, B. J., Olivieri, G., Peduto, A. S., David, A. S., Gordon, E. and Harris, A. W. F. (2007). Fronto-limbic and autonomic disjunctions to negative emotion distinguish schizophrenia subtypes, *Psychiatry Res. Neuroimaging* **155**, 29–44. DOI:10.1016/j.pscychresns.2006.12.018.

World Health Organization (2018). *Schizophrenia fact sheet*. Retrieved from www.who.int/news-room/fact-sheets/detail/schizophrenia.

Yang, L., Chen, S., Chen, C.-M., Khan, F., Forchelli, G. and Javitt, D. C. (2012). Schizophrenia, culture and neuropsychology: sensory deficits, language impairments and social functioning in Chinese-speaking schizophrenia patients, *Psychol. Med.* **42**, 1485–1494. DOI:10.1017/S0033291711002224.

Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C. and Wager, T. D. (2011). Large-scale automated synthesis of human functional neuroimaging data, *Nat. Methods* **8**, 665–670. DOI:10.1038/nmeth.1635.

Yu, L., Rao, A., Zhang, Y., Burton, P. C., Rishiq, D. and Abrams, H. (2017). Neuromodulatory effects of auditory training and hearing aid use on audiovisual speech perception in elderly individuals, *Front. Aging Neurosci.* **9**, 30. DOI:10.3389/fnagi.2017.00030.

Zhang, Y., Kuhl, P. K., Imada, T., Iverson, P., Pruitt, J., Stevens, E. B., Kawakatsu, M., Tohkura, Y. and Nemoto, I. (2009). Neural signatures of phonetic learning in adulthood: a magnetoencephalography study, *Neuroimage* **46**, 226–240. DOI:10.1016/j.neuroimage.2009.01.028.

Zvyagintsev, M., Parisi, C., Chechko, N., Nikolaev, A. R. and Mathiak, K. (2013). Attention and multisensory integration of emotions in schizophrenia, *Front. Hum. Neurosci.* **7**, 674. DOI:10.3389/fnhum.2013.00674.