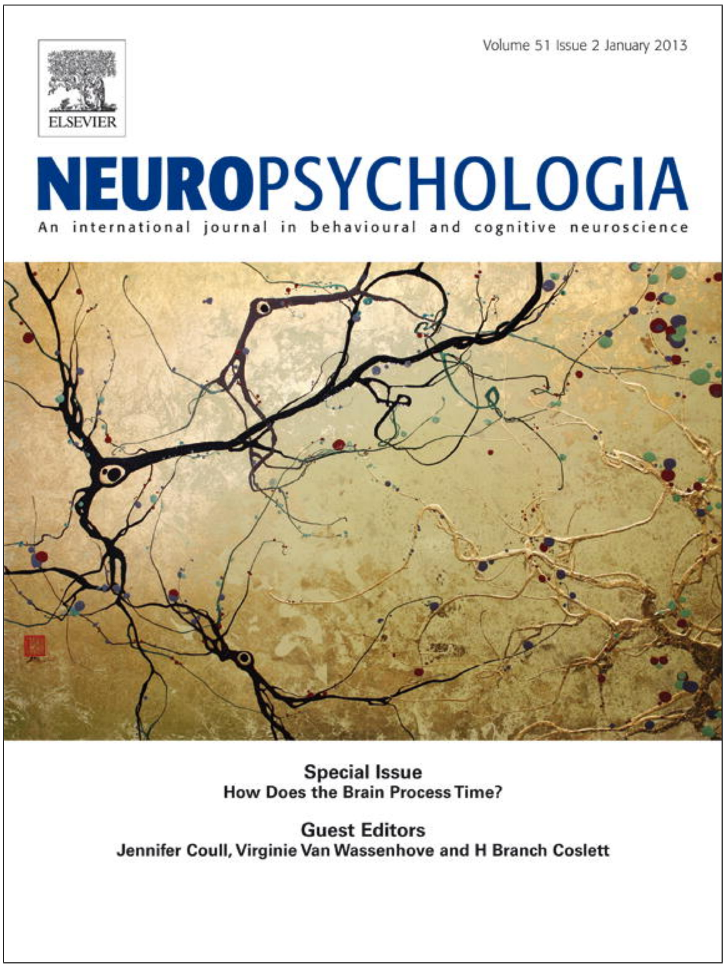


Volume 51 Issue 2 January 2013

ELSEVIER

NEUROPSYCHOLOGIA

An international journal in behavioural and cognitive neuroscience

Special Issue
How Does the Brain Process Time?

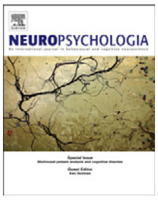
Guest Editors
Jennifer Coull, Virginie Van Wassenhove and H Branch Coslett

Neuropsychologia 51 (2013) 358–371

Contents lists available at SciVerse ScienceDirect

Neuropsychologia

journal homepage: www.elsevier.com/locate/neuropsychologia



Temporal event structure and timing in schizophrenia: Preserved binding in a longer "now"

Brice Martin[d,e,f], Anne Giersch[g], Caroline Huron[a,b,c], Virginie van Wassenhove[a,b,c,*]

[a] *INSERM, Cognitive Neuroimaging Unit, Bât 145 Point Courrier 156, 91191 Gif Yvette, France*
[b] *CEA, DSV/I2BM, NeuroSpin Center, F-91191 Gif/Yvette, France*
[c] *Univ Paris—Sud, Cognitive Neuroimaging Unit, F-91191 Gif/Yvette, France*
[d] *Univ Paris—Diderot, F-75205 Paris, France*
[e] *Univ Paul Sabatier, 31062 Toulouse, France*
[f] *Centre de Réhabilitation, Service Hospitalo—Universitaire de Réhabilitation, CHS Le Vinatier, 98 rue Boileau, 69006 Lyon, France*
[g] *INSERM, U666, Centre Hospitalier Régional Universitaire de Strasbourg, Département de Psychiatrie, 67091 Strasbourg, France*

ARTICLE INFO

Available online 16 July 2012

Keywords:
Multisensory integration
Audiovisual speech
Schizophrenia
Time perception
Fragmentation
Simultaneity
Psychophysics

ABSTRACT

Patients with schizophrenia experience a loss of temporal continuity or subjective fragmentation along the temporal dimension. Here, we develop the hypothesis that impaired temporal awareness results from a perturbed structuring of events in time—i.e., canonical neural dynamics. To address this, 26 patients and their matched controls took part in two psychophysical studies using desynchronized audiovisual speech. Two tasks were used and compared: first, an identification task testing for multisensory binding impairments in which participants reported what they heard while looking at a speaker's face; in a second task, we tested the perceived simultaneity of the same audiovisual speech stimuli. In both tasks, we used McGurk fusion and combination that are classic ecologically valid multisensory illusions. First, and contrary to previous reports, our results show that patients do not significantly differ from controls in their rate of illusory reports. Second, the illusory reports of patients in the identification task were more sensitive to audiovisual speech desynchronies than those of controls. Third, and surprisingly, patients considered audiovisual speech to be synchronized for longer delays than controls. As such, the temporal tolerance profile observed in a temporal judgement task was less of a predictor for sensory binding in schizophrenia than for that obtained in controls. We interpret our results as an impairment of temporal event structuring in schizophrenia which does not specifically affect sensory binding operations but rather, the explicit access to timing information associated here with audiovisual speech processing. Our findings are discussed in the context of curent neurophysiological frameworks for the binding and the structuring of sensory events in time.

1. Introduction

A core distinction in cognitive neurosciences is the dissociation between automatic processes and attention-driven processes that implicate higher-order operations such as "top-down" control (i.e., distinction between implicit or explicit processes, respectively). The set of automatic operations implicated in the temporal organization of information is called "temporal event-structure" (Zacks & Tversky, 2001) and necessitates the segmentation of temporal units of information (Zacks, Speer, Swallow, Braver, & Reynolds, 2007). Functionally, these temporal units are time segments or temporal windows of various duration within which information is integrated in the brain (Theunissen & Miller, 1995; van Wassenhove, 2009; Wittmann, 2011). Neurophysiologically, temporal windows are the natural outcome of synaptic delays at the neuronal level or neural oscillations at the population level (for review see: Wang, 2010; Buzsáki, 2006, 2010). The automatic and implicit temporal segmentation thus provides the building blocks for more abstract levels of representations and has crucial implications for the qualitative and phenomenological aspect of conscious experience. However, it is unclear whether implicit and explicit temporal event structuring share similar functional properties or rely on entirely different neural mechanisms. This distinction is crucial for patients with schizophrenia: schizophrenia is typically characterized by a loss of experiential continuity, consisting of the subjective fragmentation of the experienced world, including its temporal dimension, and this, we argue, could be accounted for by impaired temporal event-structuring.

---

[*] Corresponding author at: INSERM, Cognitive Neuroimaging Unit, Bât 145 Point Courrier 156, 91191 Gif Yvette, France. Tel.: +33 16 908 1667.
E-mail addresses: Virginie.van-Wassenhove@cea.fr, Virginie.van.Wassenhove@gmail.com (V. van Wassenhove).

0028-3932/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.neuropsychologia.2012.07.002

Several psychiatrists consider the experienced loss of continuity in the sense of time as a key factor in the pathophysiology of schizophrenia (Andreasen, 1999; Minkowski, 1933); what does this precisely entail? Although self-reports ought to be taken with caution, we cite one case illustrating alterations that have been clinically described (Fuchs, 2007; Kimura, 1994; Minkowski, 1933; Vogeley & Kupke, 2007), *"Time splits up and doesn't run forward anymore. These arise uncountable disparate now, now, now, all crazy and without rule or order"* (quoted in Kimura, 1994). Other similar reports can be found illustrating the need to integrate phenomenological reports with current cognitive neuroscientific approaches (Uhlhaas & Mishara, 2007).

In addition to clinical descriptions and self-reports, a number of studies have reported impairments of duration perception (Davalos, Kisley, & Freedman, 2005; Elvevåg et al., 2003; Volz et al., 2001) and a perturbed discrimination of simultaneous vs. synchronous events (Foucher, Lacambre, Pham, Giersch, & Elliott, 2007; Giersch et al., 2009; Schmidt, McFarland, Ahmed, McDonald, & Elliott, 2011). The latter studies show that for patients to become aware of the asynchrony between two sensory events, these events have to be separated by longer delays than for controls. The range of temporal delays that lies below the asynchrony detection threshold constitutes the actual temporal window of integration; within that window, events are considered to be simultaneous. Hence, the enlarged temporal window observed in patients suggests that they are binding or integrating events for a longer time or "in excess" compared to controls. These enlarged temporal windows are observed when explicitly accessing time information (i.e., when patients are asked to report the temporal characteristics of stimuli) and may be at the core of the general inability in organizing events in time.

Besides these explicit temporal impairments, recent results also suggest that patients with schizophrenia are sensitive to desynchronies at an implicit level: it has notably been shown that patients' responses are influenced by short and unconscious asynchronies (Giersch et al., 2009; Lalanne, van Assche, & Giersch, 2012, submitted). Sensitivity to short asynchronies does not tell us how different events are integrated in time, especially at an implicit level. However, the "unity assumption" in multisensory research posits that events are most likely to bind if they are perceived as belonging to a unique underlying cause: in other words, events perceived to be simultaneous should be more likely to bind together (Vatakis & Spence, 2007; Welch & Warren, 1980). For instance, in a populated room, the auditory utterance and the movements of a speaker's face that perceived to be in-sync are more likely to bind together in a single stream of speech. In schizophrenia, impaired audiovisual (AV) integration has previously been reported (de Gelder, Vroomen, Annen, Masthof, & Hodiamont, 2003; de Gelder et al., 2005; Ross et al., 2007) but impairments are not uniform (Pearl et al., 2009; Surguladze et al., 2001) and speech-specific (de Gelder et al., 2003).

Taken all together then, patients with schizophrenia would show less integration despite an enlarged temporal window of integration. This is clearly inconsistent: enlarged temporal windows should be associated with more, and not less, integration. Here, we thus aim at disentangling this conundrum by testing the possible dissociation between implicit and explicit temporal processing and by defining which specific impairments lead to the time distortions experienced by patients with schizophrenia.

For this, we focused on the possible consequences of temporal event-structure impairment in the perceptual binding of ecologically relevant stimuli such as AV speech—which bear obvious daily life relevance. We predicted that such a temporal-event structure deficit would affect the known temporal constraints of AV speech integration and that the subjective temporal estimation of these constraints would be perturbed. The former hypothesis can be addressed using an identification task (ID) in which participants report their perception of AV speech stimuli implicating the integration of visual and auditory information: this is equivalent to measuring the implicit timing of perceptual binding operations. The latter hypothesis can be tested using a simultaneity judgment task (SIM) in which participants report their perceived simultaneity of auditory and visual components of speech events: this assesses the explicit access to the encoding of temporal information. Using these approaches concomitantly (e.g., Conrey & Pisoni, 2006; van Wassenhove, Grant, & Poeppel, 2007) empirically addresses a tricky theoretical issue at the core of temporal perception research: namely, can we experimentally dissociate the temporal content of a representation (explicit time encoding) from the temporal characteristics of a representation (implicit time) (Dennett & Kinsbourne, 1992; van Wassenhove, 2009)?

Well known ecologically relevant illusions necessitating the binding of information across auditory and visual sensory modalities are the McGurk effects (McGurk & MacDonald, 1976). In McGurk/illusory fusion, dubbing an auditory "ba" (A_b) onto a visual place of articulation "ga" (V_g) leads to the illusory fused percept "da"; in McGurk illusion/combination, dubbing an auditory "ga" (A_g) onto a visual place of articulation "ba" (V_b) leads to the illusory combination percept "bga". Fusion is used as an index of automatic *AV* speech integration (Sams et al., 1991; van Wassenhove, Grant, & Poeppel, 2005) because it leads to a unique perceptual outcome that is nothing like any of the original sensory inputs (i.e., neither "ga" nor "ba"). Combination has been much less studied: unlike fusion, the resulting percept is not unique but the product of co-articulated AV speech information (such as "bga"). Fusion and combination stimuli were specifically chosen for the identification task to provide an insight on the binding mechanisms of speech: since auditory and visual speech stimuli and perceptual reports differ from each other, an index of multisensory integration is clearly obtained when desynchronizing the auditory and visual speech stimuli. AV speech integration has been shown to tolerate asynchronies in the order of 200 to 300 ms in healthy population (Conrey & Pisoni, 2006; Munhall, Gribble, Sacco, & Ward, 1996; Maier, Di Luca, & Noppeney, 2011; van Wassenhove et al., 2007). These temporal windows reflect precise neurophysiological correlates that have recently been described within a predictive coding framework for AV speech processing (Arnal, Morillon, Kell, & Giraud, 2009; Arnal, Wyart, & Giraud, 2011; van Wassenhove et al., 2005) and are in line with temporal units necessary for speech parsing (Poeppel, 2003; Giraud & Poeppel, 2012). Thus, AV speech makes an ideal ecological test for our question.

In healthy participants, no major differences were observed when comparing the temporal windows obtained in an ID or a SIM task (Conrey & Pisoni, 2006; van Wassenhove et al., 2007): the temporal properties of AV speech integration appear to reflect directly the temporal information available for the conscious perception of AV speech simultaneity. As previously emphasized, this is in marked contrast with what is currently observed in patients with schizophrenia. Patients appear to have a deficit in integrating multisensory information whereas their explicit impairments would have predicted excessive integration. The limit of the current literature in schizophrenia is that explicit and implicit judgments have not been directly compared using multisensory information. This study fills this gap by directly comparing patients' AV speech integration and simultaneity ratings on the same stimuli and in two tasks. First, we proceeded with assessing AV speech integration in two groups of patients with schizophrenia using illusory McGurk fusion and combination. We then tested whether patients showed an enlarged tolerance to AV desynchrony when identifying the illusions—namely, do AV speech illusions tolerate more asynchrony in patients than in controls (ID task, implicit timing)? Third, we used

Table 1
Study design and parameters.

	Study 1 ($n=26$; 13 patients, 13 matched controls) AV asynchronies [−is A lead; +is A lag]+/−0, 80, 120, 200, 240, 280, 320, 360, 440 ms	Study 2 ($n=26$; 13 patients, 13 matched controls) AV asynchronies [−is A lead; +is A lag]−960, −560, −240, −80, 0, +80, +160, +240, +320, +400, +480, +560, +720, +1040, +1440 ms
McGurk fusion A_bV_g; male speaker	Identification (ID) simultaneity judgment (SIM)	Identification (ID) simultaneity judgment (SIM)
Mc Gurk combination A_bV_g; female speaker congruent speech A_bV_b; female speaker	Identification (ID) simultaneity judgment (SIM)	Identification (ID) simultaneity judgment (SIM) simultaneity judgment (SIM)
Congruent speech A_bV_b ; female speaker	N/A	Simultaneity judgment (SIM)

Table 2
Characteristic patients with schizophrenia and matched controls.

Study 1 ($n=26$)			Study 2 ($n=26$)	
	Patients ($n=13$)	Controls ($n=13$)	Patients ($n=13$)	Controls ($n=13$)
Age	35 (7.5)	33 (8.9)	39 (8.6)	44.2 (8.8)
Female	3	3	6	6
PANSS positive	14.8 (4.2)	–	17 (6.1)	–
PANSS negative	19.6 (8.4)	–	28.2 (11.3)	–
PANSS global	35.1 (15)	–	40.7 (10.55)	–
PANS total	63.8 (31.6)	–	85.8 (25.4)	–
Medication	Clozapine (30%) Aripiprazole (15%) Olanzapine (15%) Risperidone (15%) Haloperidol (8%) Zuclopentixol (8%) Fluphénazine (8%)	–	Olanzopine (13%) Risperidone (33%) Haloperidol (13%) Zuclopenthixol (13%) Fluphénazine (6%) Flupentixol (13%) Pipotiazine (6%)	

the same stimuli but this time asked participants to judge whether AV events were simultaneous or successive in time (SIM task, explicit timing) Table 1.

2. Material and methods

2.1. Participants

Participants were stabilized chronic outpatients individually matched in gender, age, and level of education with healthy controls. All participants were native speakers of French, had healthy or corrected-to-normal vision with no known speech or hearing disabilities. Patients were recruited from the Department of Psychiatry at Strasbourg University, France and from a local hospital (Association Elan Retrouvé, Paris, France). All patients met the diagnostic and statistical manual of mental disorders (DSM IV) criteria for schizophrenia. The psychiatric diagnosis was established by a senior psychiatrist located at each of the recruiting institutions. Each patient also completed the positive and negative syndrome scale (PANSS) to evaluate the severity of each dimension of schizophrenia. Matched controls were recruited at NeuroSpin (Gif-sur-Yvette, France). Written informed consents were obtained from all participants in accordance with the Declaration of Helsinki and the Ethics Committee on Human Research at the Commissariat à l'Energie Atomique et aux Energies Alternatives (NeuroSpin, Gif-sur-Yvette, France). All participants were compensated for their participation in the study. Thirteen patients with schizophrenia with schizophrenia (10 men; mean age of 35+/−7.5 years) and their matched controls (10 men; mean age of 33 years+/−8.9 years) took part in Study 1. Thirteen different patients (7 men; mean age of 39+/−8.6 years) and their matched controls (7 men; mean age of 44.2+/−8.8 years) took part in Study 2. Table 2 reports a complete description of patients PANSS scores and antipsychotic treatments. The selection of participants for subgroup analysis is described in details where needed.

2.2. Experimental setup

Experiments were run in a quiet room of medium-intensity ambient light using an Intel(R)core^{TM}2duo PC (Windows XP). Experiments were designed with PsychToolbox (v3; Brainard, 1997) in Matlab (v7.9.0.959) with Quick time pluggin (version 7.1.6) to display movies. Visual stimuli were presented on a 15′ screen with a vertical refresh rate of 60 Hz. Auditory stimuli were delivered through headphones (Bayer dynamics, DT 880 Pro) at a comfortable loudness level. Participants sat at 80 cm from the presentation screen and gave their responses by one of the two or three response keys on the keyboard.

2.3. Stimuli

2.3.1. Video and audio processing

Three speakers (two women, one man) were digitally recorded with a digital camera (Sony Handycam DCR-DVD203E) while they were pronouncing the syllables [ba] and [ga]. Recordings were made at a rate of 25 fps (1 frame=40 ms) and at a sampling rate of 44.1 kHz for the sound. Digital outputs were MPEG files.

2.3.2. McGurk pairs

The MPEG videos were edited using Magix Video Deluxe 16 (v.9.0.0.55). Each video ([ga] henceforth referred to as V_g and [ba] henceforth referred to as V_b) was dubbed with the incongruent audio syllable ([ba] or A_b and [ga] or A_g, respectively) at the timing of the original congruent token. This processing provided McGurk stimuli namely, fusion pairs (A_bV_g: audio [ba] dubbed onto the visual place of articulation [ga]) and combination pairs (A_bV_g: audio [ga] dubbed onto the visual place of articulation [ba]). Three instances of audiovisual speech fusion and combination were thus obtained (two females, one male). All files were converted in AVI format. Pilot data were collected to test the robustness of the created stimuli with 10 naïve participants. In this pilot testing, the maximum fusion rate was obtained for the male token A_bV_g with a mean fusion rate of 55% ($SD=0.46$) and with a female token A_bV_g with a mean combination rate of 81% ($SD=0.29$). Those stimuli were chosen for the reported studies.

2.3.3. Audiovisual (AV) temporal alignment

AV asynchronies were realized by shifting the audio portion of the stimuli by 40 ms (1 frame) or multiple values of 40 ms (frame units) with respect to the original sound onset in the movie file. In the first study, stimuli ranged from 440 ms of auditory (A) lead (−440 ms, −11 frames) to 440 ms of auditory lag (+440 ms, +11 frames); in the second study, stimuli covered 960 ms of A lead (−960 ms, −24 frames) to 1440 of A lag/visual lead (+1200 ms, +36 frames). All AV timings were controlled with an oscilloscope, a microphone and a photocell.

2.4. Procedure

Participants answered by pressing the "*J*","*K*", "*L*" or the "*J*" and "*K*" keys as a function of the task. Stickers were placed on each key to provide the actual choice (e.g., BA, DA and GA for fusion blocks; BA, BGA, GA, for combination blocks). Responses were continuously recorded on line.

2.4.1. Identification tasks (ID)

2.4.1.1. Study 1. The identification task (ID) consisted of two blocks: one fusion (A_bV_g) and one combination (A_bV_g) block. Each block consisted of 10 trials per timing condition presented pseudo-randomly (17 AV asynchronies conditions × 10 trials per condition for a total of 170 trials per block). The tested AV asynchronies (SOA) ranged from −440 to +440 ms (+/−0, 80, 120, 200, 240, 280, 320, 360, 440 ms). In the fusion A_bV_g identification block, five trials of A_b, A_g, V_b and V_g were included in order to obtain an estimate of unisensory recognition rate. In each block, a 3-alternative-forced-choice (3-AFC) procedure was used: participants decided "what they heard while listening to and looking at the talking face". They were given three choices: BA (A component), DA (illusory fusion) or GA (V component) in the fusion (A_bV_g) block and BA (V component), BGA (illusory combination) or GA (A component) in the combination (A_gV_b) block. Noone reported experiencing any other alternatives.

2.4.1.2. Study 2. As in Study 1, the ID task consisted of a fusion (A_bV_g) and a combination (A_bV_g) blocks. Each block contained 8 trials per SOA (15) for a total of 125 trials per block. Study 2 extends the range of SOA tested in Study 1: SOAs ranged from −960 to 1440 ms (−960, −560, −240, −80, 0, +80, +160, +240, +320, +400, +480, +560, +720, +1040, +1440). Five unisensory trials were included in the fusion (A_bV_g) block; a 3-AFC procedure and the response key mapping were identical to Study 1.

2.4.2. Subjective simultaneity judgment tasks (SIM)

2.4.2.1. Study 1. The SIM task contained 10 repetitions of each SOA presented in pseudo-random order: 17 SOAs × 10 repetitions in A_bV_g (combination) and A_bV_g (fusion) blocks for a total of 170 trials per block. SOAs ranged from −440 ms to +440 ms (+/−0, 80, 120, 200, 240, 280, 320, 360, 440 ms). A 2-AFC procedure was used ("simultaneous" or "successive"). Participants were told that the congruency between *A* and *V* speech were not to be estimated, and that they should solely focus on the timing of AV events.

2.4.2.2. Study 2. The SIM task contained 8 repetitions of each SOA presented in pseudorandom order: 15 timing conditions × 8 repetitions in both A_bV_g and A_bV_g blocks for a total of 125 trials per block. A third block was added, containing only congruent pairs (A_bV_g) in order to control the effect of the incongruence on simultaneity estimation, as recommended by Vroomen and Keetels (2010). This block contained the same characteristics as other blocks namely, 17 SOAs × 10 repetitions. The SOAs ranged from −960 to 1440 ms (−960, −560, −240, −80, 0, +80, +160, +240, +320, +400, +480, +560, +720, +1040, +1440). A 2-AFC forced choice procedure was used ("simultaneous" or "successive").

2.5. Analysis

In all studies, responses were sorted out and averaged for each participant and each condition of interest. A grand average of each possible response per SOA was computed across participants for each population (patients, matched controls). Subgroup analyses focused on those participants (patients and matched controls) showing illusory reports in the ID task: this analysis specifically focuses on the direct comparison between the temporal granularity of AV speech integration (ID) and simultaneity judgments (SIM).

2.5.1. Measures of multisensory integration

Measures of AV speech integration (namely, rates of McGurk illusory fusion "da" and McGurk illusory combination "bga") were estimated for each individual as the maximum value irrespective of SOA ($_{\mathrm{MAX}}$) and for simultaneous presentation (SYNC). Individual values were averaged for each population and tested between populations.

2.5.2. Temporal windows of integration—ADS fits

Temporal profiles refer to the entire curve obtained for all SOAs values. The temporal window of integration (TWI) corresponds to the width of the perceptual window in ID and SIM. Parametrization of the TWI were accomplished using an asymmetric double sigmoid fit (ADS) (van Wassenhove et al., 2007) in order to derive the following parameters: (i) the just-noticeable-differences (jnds) taken at the 75% threshold (Vroomen & Keetels, 2010) on either side of the curve (audio lead and audio lag); (ii) the point of subjective equality or simultaneity (PSE, PSS, respectively) taken as the median of the two jnds and (iii) the width of the window (i.e., the temporal window of integration per se) simply defined as the duration between the two jnds. Data fitting were made using TableCurve 2D (SYSTAT software, v5.01).

2.5.3. Statistical analysis

All statistical analyses were performed using SPSS software (IBM, v19). Groups and subgroups considered in each statistical analysis is detailed for each set of results in (Section 3).

3. Results

3.1. AV speech integration: no impaired multisensory binding in patients with schizophrenia

Across both studies, 15 patients with schizophrenia and 14 controls showed some fusion; 18 participants in each group showed some combination. To address the strength of illusion in both populations, we first proceeded with an analysis including all participants. We then included those participants showing illusions only to check whether differences between patients and controls could be found in those individuals showing multisensory illusion (cf. subgroup analysis).

3.1.1. Maximal illusion rates irrespective of SOA ($_{\mathrm{MAX}}$)

AV speech illusion rates obtained in Study 1 and 2 were gathered to evaluate the hypothesized differences of AV integration deficit in all patients ($n=26$) compared to their matched controls ($n=26$). Fig. 1a provides a summary of the integration rates for fusion and combination tested for each population (patients in dark gray, matched controls in light gray). 2 × 2 repeated measures ANOVA were conducted with illusion rate as dependent variable with factors of population (2: patients, controls) and illusion type (2: fusion, combination). Contrary to our expectations, no significant differences were observed between the two populations ($F_{1,25}=0.478$, $p=0.496$). A significant effect of illusion was found ($F_{1,25}=6.291$, $p=0.019$) but the two-way interaction between population and illusion was not significant ($F_{1,25}=1.525$, $p=0.228$).

3.1.2. Illusion rates at AV synchrony (SYNC)

To ensure that selecting maximal integration irrespective of SOA did not bias the estimate of illusion rates across populations, the same analysis was conducted using the illusion rates obtained in synchronous AV presentations (as is typically the case in McGurk empirical work, Fig. 1a). 2 × 2 repeated measures ANOVA with illusion rate as dependent variable and with factors of population (2: patients, controls) and illusion type (2: fusion and combination) was conducted revealing no significant differences between populations ($F_{1,24}=0.089$, $p=0.767$). A slight effect of illusion was found ($F_{1,24}=4.524$, $p=0.044$) but no interaction between population and illusion was observed ($F_{1,24}=2.713$, $p=0.113$).

To further insure no bias in our measure, a 2 × 2 × 2 statistical design with illusion rate as dependent variable and factors of population (2), illusion type (2) and parameterization method (2: MAX, SYNC) showed a significant effect of method ($F_{1,24}=44.602$, $p\leq 0.0001$) but no interaction of population with method ($F_{1,24}=0.601$, $p=0.446$).

Paired-*t* tests within population for each pair (MAX, SYNC) showed a significant and consistent overestimation of illusion rate when disregarding the asynchrony value (i.e., in MAX, cf. Fig. 1a.). This suggests that natural synchrony does not necessarily lead to maximal AV integration. This is typically found in AV integration.

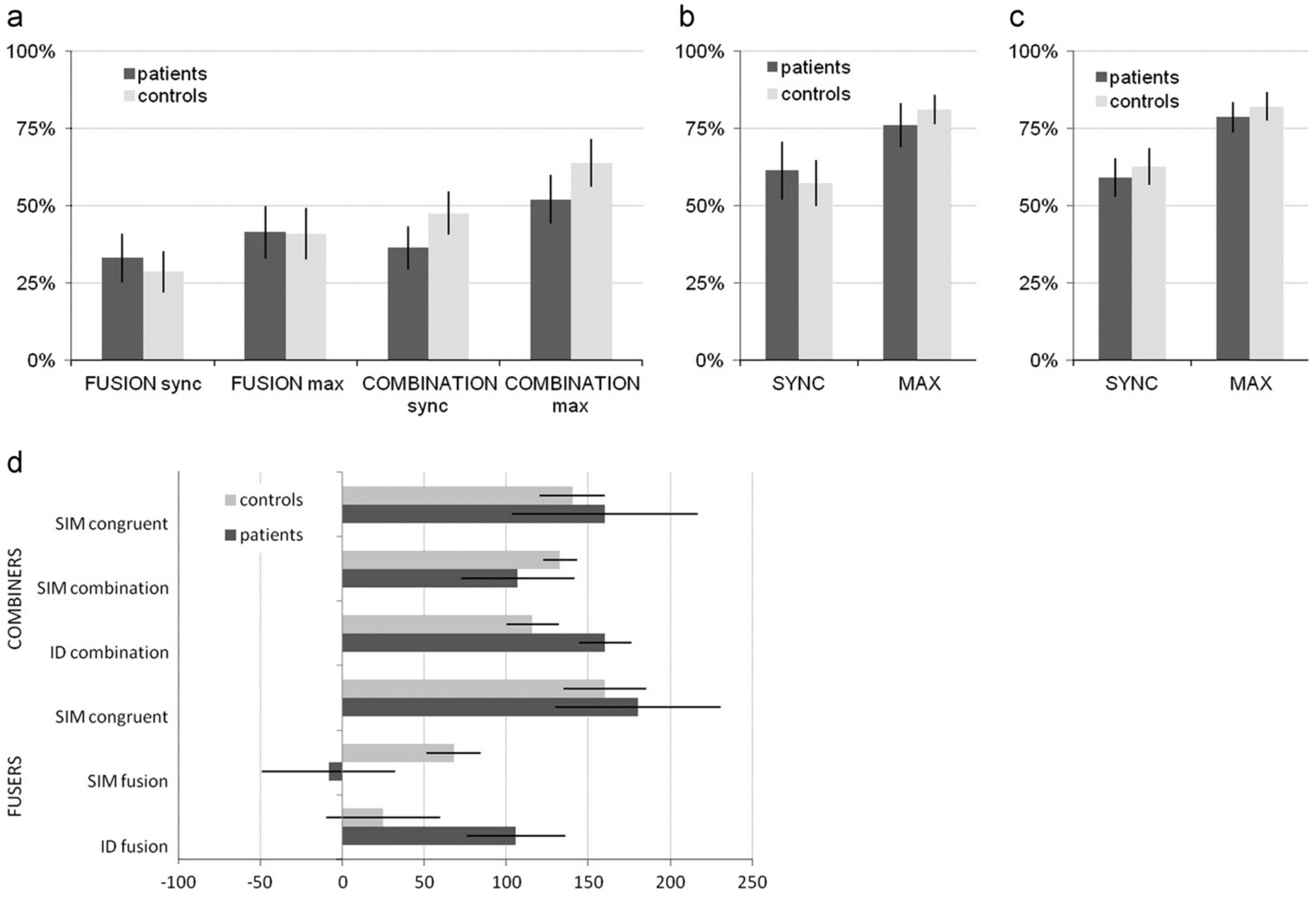


Fig. 1. Illusion rates. Illusion rates were quantified at AV synchrony (SYNC) or irrespective of SOA ($_{\mathrm{MAX}}$). Participants in both studies were pooled. (a) All participants are considered (26 patients and their 26 matched controls). No significant differences in fusion or combination rates were found between patients with schizophrenia and controls, (b) participants showing fusion in the schizophrenia group ($n=13$) and the control group ($n=14$) were pooled together. No significant differences were found between the two groups, (c) participants showing combination in the schizophrenia group ($n=16$) and the control group ($n=19$) were pooled together. No significant differences were found between the two groups, (d) SOA at which fusion, combination and simultaneity ratings were found to be maximal in Study 1 and Study 2 combined. Error bars are two standard errors of the mean.

3.1.3. *Subgroup analysis for participants showing McGurk illusions*

We further checked whether differences in illusory rates between patients and controls could be found by selecting only those participants who showed illusory reports. In Study 1, 10 patients and 8 controls showed fusion, 11 patients and 11 controls showed combination; in Study 2, 4 patients and 5 controls showed fusion, and 5 patients and 8 controls showed combination. When comparing fusion ($n_p=14$, $n_c=13$) and combination ($n_p=16$, $n_c=19$) rates between patients and controls, no significant differences were observed whether considering illusion scores at synchrony or individuals' maximal illusory scores (Fig. 1b and Fig. 1c). Additionally, no significant differences were found within these subgroups regarding the recognition scores for auditory speech or visual speech stimuli used to create the fusion and combination (Supp. Mat. Fig. 1b).

All together this first set of analysis does not provide evidence for a profound deficit of AV speech integration in schizophrenia. This point will be critically assessed in Section 4.

3.2. *Temporal constraints on multisensory integration—ID task (implicit timing)*

The second question addressed in this study was whether the temporal constraints on AV speech integration differed in patients with schizophrenia and controls namely, whether AV speech integration operates with similar time scales as those previously reported in healthy population (Conrey & Pisoni, 2006; Maier et al., 2011; Munhall et al., 1996; van Wassenhove et al., 2007). Considering that patients have previously been reported to tolerate larger asynchronies in their simultaneity ratings (Foucher et al., 2007; Giersch et al., 2009; Schwartz, Winstead, & Walker, 1984), if explicit temporal judgments reflect implicit temporal processes implicated in binding operations, one prediction was that AV speech integration in the ID task should also tolerate larger asynchronies in patients.

Analyses were separately done for Study 1 and Study 2 as different SOAs were used. Despite 400 ms ms of asynchrony in Study 1, the illusion and the asynchrony judgments did not fully reach zero; hence, a larger set of SOAs was used in Study 2. Only those participants showing illusory reports were kept in the reported analyses, namely, 10 patients and 8 controls in Study 1 and 4 patients and 5 controls in Study 2 for the analyses pertaining to fusion (the "fusers" subgroups) and 11 patients and controls in Study 1 and 5 patients and 8 controls in Study 2 for the analyses pertaining to combiners ("combiners" subgroups). Data across studies were gathered together when relevant for the question of interest.

3.2.1. *Identification task (ID)*

Fig. 2 shows the percentage of illusory fusion "da" (Fig. 2a and c) and combination "bga" (Fig. 2b and d) as a function of SOA in

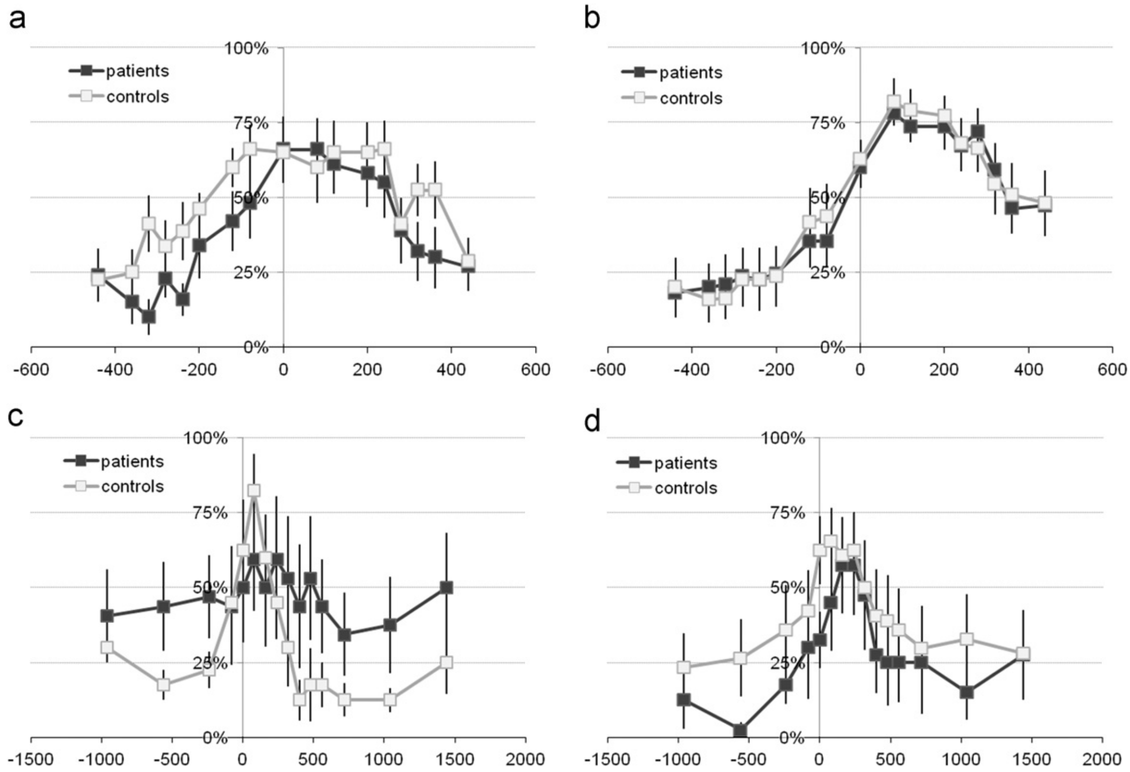


Fig. 2. Identification task (ID). Two groups of patients (dark gray) and their matched controls (light gray) were tested on their rate of fusion and combination with AV speech stimuli presented asynchronously. McGurk fusion (a), (c) and combination (b), (d) were obtained with two sets of asynchronies. Only participants showing the illusory effects are reported in theses curves. (a) Fusers subgroup of Study 1, (b) combiners subgroup of Study 1, (c) fusers subgroup of Study 2, (d) combiners subgroup of Study 2. SOA significantly affected the rate of illusions in all cases. Patients and controls only differed in the ≪combiners≫ subgroup of Study 2, with controls showing a significantly extended tolerance window compared to patients (panel c). Error bars are two standard errors from the mean.

patients (dark gray) and controls (light gray). As predicted, as the SOA between the AV speech syllables increased, the illusory responses decreased. 2 × 17 repeated measures ANOVA for fusion responses with factors of population (2) and SOA (17) revealed a significant main effect of SOA (Study 1: $F_{16,\ 112}=11.18$, $p \leq 0.0001$; Study 2: $F_{14,\ 42}=3.95$, $p \leq 0.0001$). In Study 1 and 2, repeated measures ANOVA for combination responses with factors of population (2) and SOA (17 or 15, respectively) also revealed a significant main effect of SOA (Study 1: $F_{16,\ 160}=49.195$, $p \leq 0.0001$; Study 2: $F_{16,\ 144}=30.189$, $p \leq 0.0001$). A two-way interaction between SOA and population was found for combination in Study 2 ($F_{16,\ 144}=2.655$, $p \leq 0.\ 001$) in which the temporal profile of patients in the combiners subgroup was surprisingly narrower than that of controls (Fig. 2d).

In the ID task, post-hoc *t*-tests revealed significant differences between patients and controls in fusion (Study 1: $t_{1,16}=-4.341$, $p \leq 0.001$; Study 2: $t_{1,14}=3.153$, $p \leq 0.007$) and in combination (Study 1: n.s; Study 2: $t_{1,16}=-5.7$, $p \leq 0.0001$). As revealed in Fig. 2, a narrower temporal profile for patients than controls was observed in fusers of Study 1 (Fig. 2a) and combiners of Study 2 (Fig. 2d); the temporal profile was identical for patients and controls in the combiners of Study 1 (Fig. 2b) but larger for patients than controls in the fusers of Study 2 (Fig. 2c).

In both studies, clear differences between the two populations could be observed with respect to the temporal constraints of AV speech integration when engaged in an identification task. These results hold when considering the whole population.

3.2.2. *Optimal SOA for AV fusion and combination*

In both studies, the SOAs for maximal fusion and combination rates were found to be positive, thereby indicating a preference for visual leads. The MAX asynchrony approximated 150–200 ms in both patients and controls. 2 × 2 × 2 repeated measures ANOVA with MAX SOA and factors of study (2), speech stimuli (2) and population (2) showed no significant effect of Study. For this reason, data across both studies were pooled together (Fig. 1d). 2 × 2 repeated measures ANOVA with factor of speech stimulus (2) and population (2) showed a main effect of stimulus ($F_{1,\ 12}=9.317$, $p \leq 0.01$): the SOA at which fusion was maximal was on average shorter than that found for combination for both groups (Fig. 1d). Although a trend towards a longer SOA of maximal fusion in patients can be seen, the large variance prevented any significant difference with controls from being seen.

3.3. *Access to time events in multisensory integration—SIM task (explicit timing)*

We now assess explicit simultaneity judgments using the same set of asynchronies. Fig. 3 shows the percentage of simultaneity responses for incongruent stimuli in Study 1 and 2 for the fusers (3a and 3c) and the combiners subgroups (3b and 3d). The same analysis was conducted for congruent AV speech stimuli [ba].

Repeated measures ANOVA for simultaneity responses with factors of population (2) and SOA (17 or 15, respectively) were performed for each study and type of stimuli. Significant main effects of SOA were obtained in all cases. Additionally, a two-way interaction of population with stimulus was obtained for combiners of Study 2 ($F_{14,\ 56}=2.016$, $p \leq 0.033$). This is consistent with the observation that patients' temporal profiles for combination stimuli and for congruent stimuli are larger than those of controls. Post-hoc *t* tests showed significant differences in the temporal profiles of patients and controls in fusers (Study 1: $t_{1,16}=-2.897$, $p \leq 0.011$;

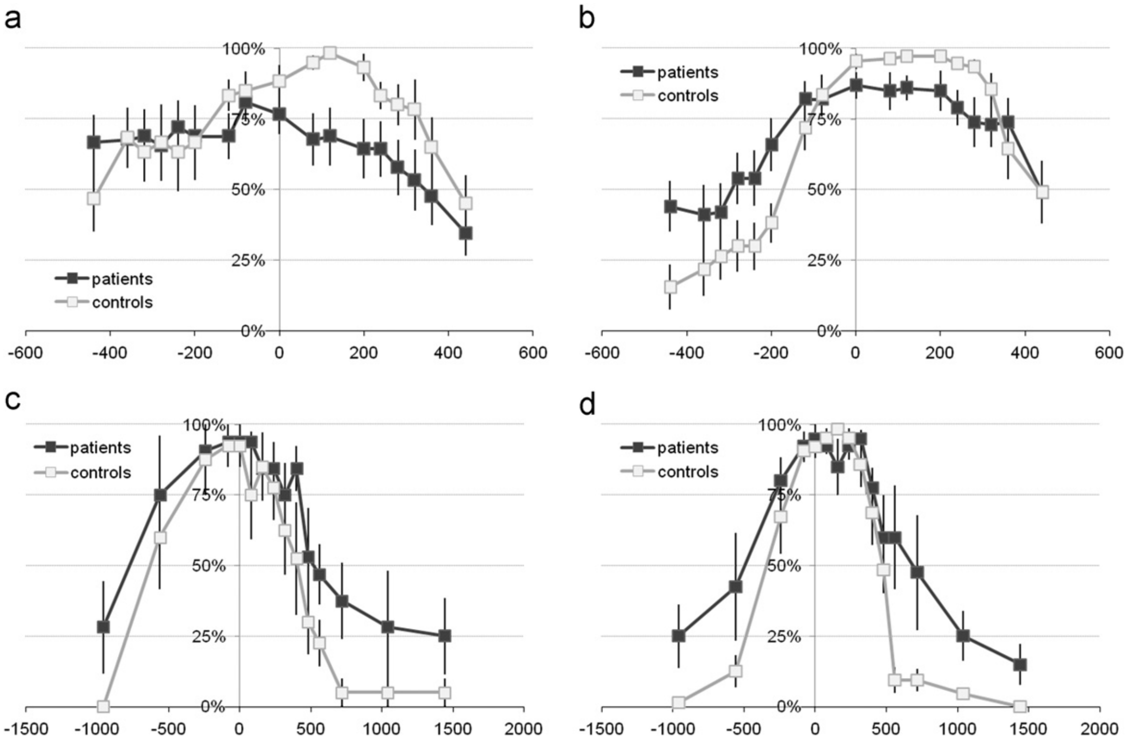


Fig. 3. Simultaneity judgment task (SIM) of incongruent AV speech. Two groups of patients (dark gray) and their matched controls (light gray) judged simultaneity of desynchronized fusion and combination illusions. Only participants showing the illusory effects are reported in theses curves. (a) Fusers subgroup of Study 1, (b) combiners subgroup of Study 1, (c) fusers subgroup of Study 2, (d) combiners subgroup of Study 2. SOA significantly affected the rate of simultaneity in all cases.

Study 2: $t_{1,14}=5.433$, $p\leq 0.0001$) and in combiners (Study 1: $t_{1,16}=-2.16$, $p\leq 0.046$; Study 2: $t_{1,14}=3.153$, $p\leq 0.007$).

As seen in Fig. 3, larger temporal profiles for patients compared to controls are ubiquitous except for fusers of Study 1 (Fig. 3a). In the profiles obtained for the congruent speech condition, patients of the fusers subgroup showed a significant widening of their temporal profiles (fusers: $t_{1,14}=3.968$, $p\leq 0.0001$); although a similar trend for a widening of the temporal profile was observed in the patients of the combiners subgroup, no significant effect was found compared to controls.

3.4. Implicit and explicit timing: Temporal windows of integration

The main goal of this study was to explore the relationship between the temporal constraints of AV speech integration (implicit timing) and simultaneity judgments (explicit timing). If profiles obtained in ID tasks provide information on the temporal resolution of the integration process, it is unclear whether it also provides relevant cues on the availability of temporal information for explicit temporal judgments. We thus proceeded in specifically comparing the results between SIM and ID tasks.

3.4.1. Identification vs. simultaneity judgment in speech illusions

When considering the fusers and combiners subgroups, a main effect of task and SOA were found irrespective of populations. Specifically, $2\times 2\times 15$ or 17 (Study 1 or 2, respectively) repeated measures ANOVA consistently showed a main effect of task (Study 1, fusers: $F_{1,\ 4}=29.85$, $p\leq 0.005$; Study 2, fusers: n.s.; Study 1, combiners: $F_{1,\ 9}=20.132$, $p\leq 0.002$ and Study 2, combiners: $F_{1,\ 4}=9.608$, $p\leq 0.036$) and a main effect of SOA (Study 1, fusers: $F_{16,\ 64}=11.452$, $p\leq 0.0001$; Study 2, fusers: $F_{14,\ 42}=13.354$, $p\leq 0.0001$; Study 1, combiners: $F_{16,\ 144}=77.637$, $p\leq 0.0001$ and Study 2, combiners: $F_{14,\ 56}=46.649$, $p\leq 0.0001$). Additionally, two-way interactions between task and SOA were often found significant (Study 1, fusers: n.s.; Study 2, fusers: $F_{14,\ 42}=9.688$, $p\leq 0.0001$; Study 1, combiners: $F_{16,\ 144}=2.948$, $p\leq 0.0001$ and Study 2, combiners: $F_{14,\ 56}=9.245$, $p\leq 0.0001$). These results suggest dissimilar temporal profiles in the ID and SIM tasks. Importantly, post-hoc *t*-tests revealed that nearly all temporal profiles between tasks (ID vs. SIM) within patients and controls differed except for the controls in the combiners subgroup of Study 2. Comparisons of profiles per subgroup are provided in Figs. 4 and 5.

All together, these results suggest that temporal constraints on integrating AV speech (implicit timing) cannot be straightforwardly equated to perceived simultaneity (explicit timing) in either controls or patients. Consistent with those results, when considering all participants irrespective of their illusory reports, post-hoc *t*-tests revealed that implicit (ID) and explicit (SIM) temporal profiles differed significantly ($p\leq 0.0001$) except for patients' fusion and controls' combination temporal profiles in Study 2. These results are congruent with the subgroup analysis and suggest that although conditions and stimuli are identical in both tasks, access to task-relevant information entails different operations.

3.4.2. Predicting audiovisual integration (ID) on the basis of simultaneity judgments (SIM)

Taken together, results suggest that task requirements affect patients and controls' temporal profiles in both experiments. Specifically, the temporal profiles observed in the SIM task capture the pattern of the ID profiles well but tended to be wider. These results are consistent with the unity assumption: auditory and visual information perceived as being simultaneous is more likely to be bound together. However, this constraint alone is insufficient to explain integration: the temporal profiles in the SIM task are often seen as slightly more tolerant than those observed in the ID task—more so in patients than in controls.

Hence, an additional analysis was carried out to evaluate the extent to which the temporal profile observed in SIM task can predict the temporal profile in an ID task—in other words, to which

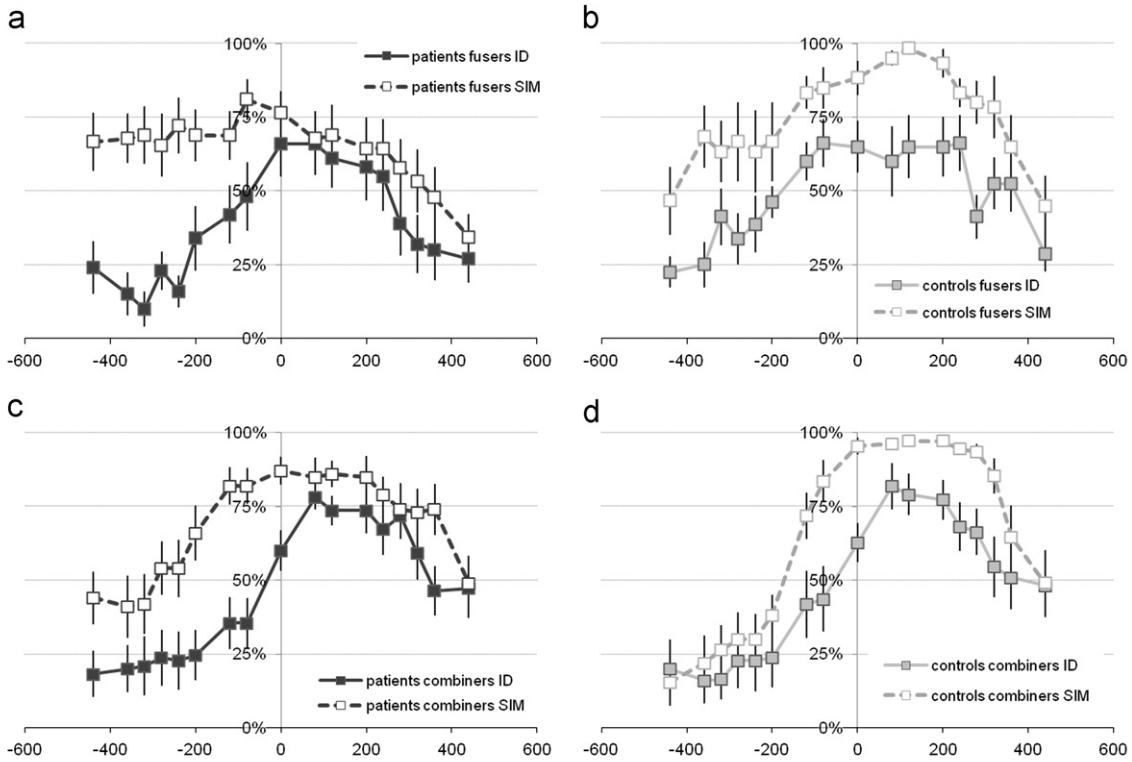


Fig. 4. Comparison of temporal profiles in ID and SIM task Study 1. Only participants showing the illusory effects are reported in theses curves. SOA significantly affected the rate of simultaneity in all cases. (a) Patients fusers temporal profiles in ID (black filled) and SIM (unfilled), (b) controls fusers temporal profiles in ID (gray filled) and SIM (unfilled), (c) patients combiners temporal profiles in ID (black filled) and SIM (unfilled), (d) control combiners temporal profiles in ID (gray filled) and SIM (unfilled). Error bars are two standard errors from the mean. Temporal profiles in the SIM task are systematically larger than in the ID task irrespective of the population. Nevertheless, patients show a pronounced tolerance for A leads in the SIM task.

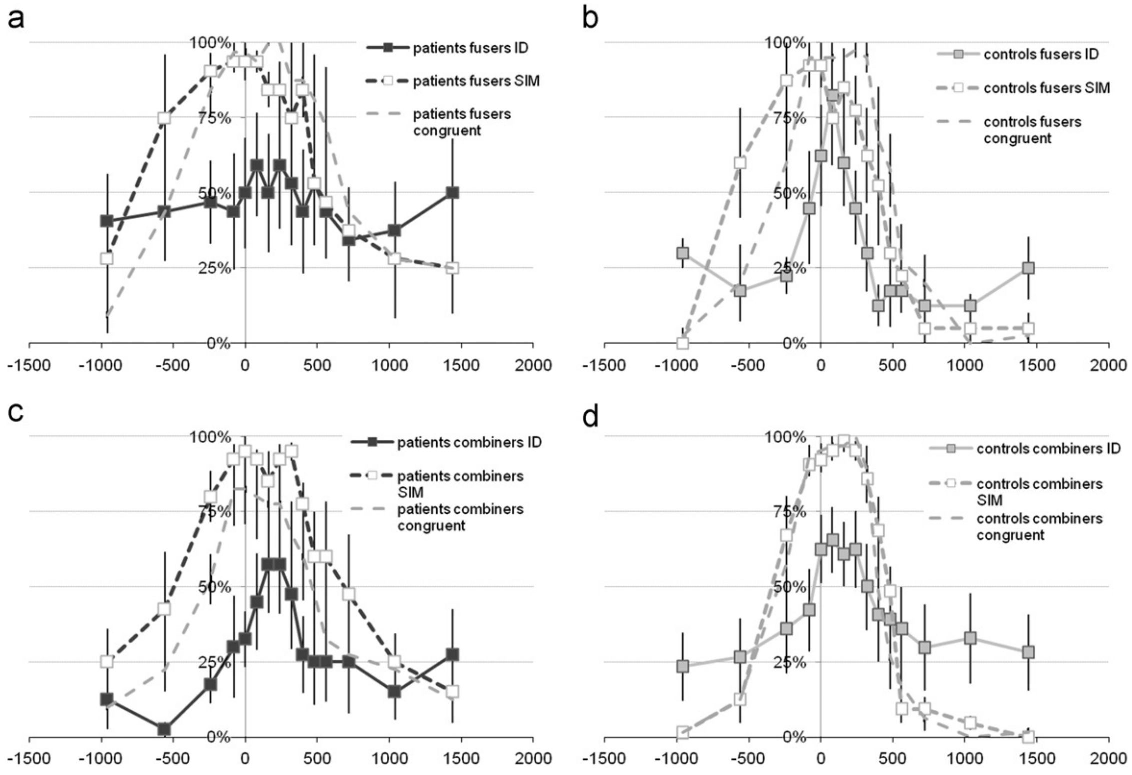


Fig. 5. Comparison of temporal profiles in ID and SIM task Study 2. Study 2 included a set of congruent audiovisual [ba] in the ID task. Only participants showing the illusory effects are reported in theses curves. SOA significantly affected the rate of simultaneity in all cases. (a) Patients fusers temporal profiles in ID (black filled) and SIM (gray filled), (b) controls fusers temporal profiles in ID (gray filled) and SIM (unfilled), (c) patients combiners temporal profiles in ID (black filled) and SIM (gray filled), (d) control combiners temporal profiles in ID (gray filled) and SIM (unfilled). Error bars are two standard errors from the mean. As in Study 1, temporal profiles in the SIM task are systematically larger than in the ID task irrespective of the population. Patients show a pronounced tolerance for A leads in the SIM task.

extent an explicit judgment task can predict the integrative properties of the perceptual system. Results and associated correlation coefficients are reported in Fig. 6: overall, controls' temporal profile in the SIM task was more predictive of their temporal profile in the ID task than in patients. This suggests that at least one additional operation specifically implicated in the access to temporal information differs in patients compared to controls.

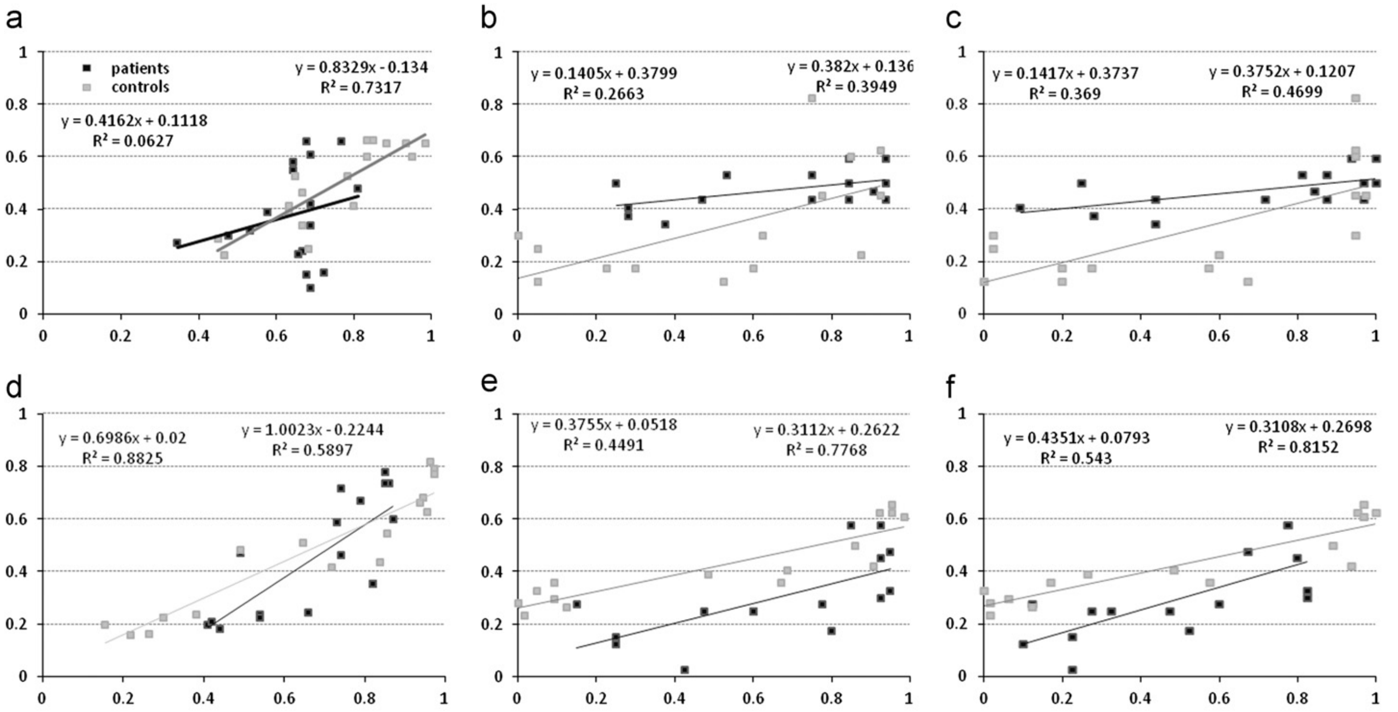
To further quantify this aspect, we turn to the temporal window of integration (TWI) per se. Note that up until now, the entire temporal profile was considered (i.e., the full range of tested SOA). Here, the temporal window of integration (TWI) specifically refers to SOA values at which optimal integration (ID) or perceived simultaneity (SIM) do not significantly differ from one another. As such, temporal profiles and TWI offer different insights on processes engaged in AV speech integration and simultaneity estimation.

3.4.3. *Temporal windows of integration (TWI)*

To specifically address potential differences between patients with schizophrenia and their controls in ID and SIM tasks, we derived each individual's temporal windows with limits defined as the just-noticeable-differences ("jnds" or 75% threshold observed on each side of the curve, Vroomen & Keetels, 2010). Four parameters for the fitted curves were drawn from the fits: the minimal (75% threshold for audio leads) and maximal (75% threshold for audio lags) thresholds, the PSE or PSS taken as the median point between the two thresholds and the width of the window thus defined (cf. Section 2.6.2). Jnds and PSS capture different aspects of behavior (Vroomen & Keetels, 2010): jnds specify the smallest AV speech asynchronies participants can detect whereas PSS provide insights on the SOA at which stimuli are maximally integrated (ID) or considered to be maximally temporally aligned (SIM). These parameters were gathered across both studies but independently so for the fusers (Fig. 7a) and the combiners (Fig. 7b) subgroups.

When considering the fusers subgroup, the width of the window significantly differed between patients and controls in the ID task ($t_{1,10}=-3.532$, $p \leq 0.005$), the TWI being significantly less tolerant to audio leads in patients than in controls (audio lead jnd: $t_{1,10}=4.683$, $p \leq 0.001$; cf. Fig. 7a). In the SIM task, no significant differences between patients and controls were observed for the width of the window in fusion due to the large variability across participants; the windows obtained for the congruent AV speech condition was however significantly larger for patients than for controls ($t_{1,5}=-3.243$, $p \leq 0.023$; Fig. 7a). When comparing the TWI obtained in the ID and SIM tasks, controls showed a significant difference of width ($t_{1,8}=-2.376$, $p \leq 0.045$; Fig. 7a) but patients did not.

In the combiners subgroup, the width of the window in the ID task significantly differed between patients and controls ($t_{1,22}=3.017$, $p \leq 0.006$; Fig. 7c) but not in the SIM task, again due to large variability. In both tasks, the PSS significantly differed



FUSION						COMBINATION					
a. study 1 ID x SIM		b. study 2 ID x SIM		c. study 2 ID x congruent SIM		d. study 1 ID x SIM		e. study 2 ID x SIM		f. study 2 ID x congruent SIM	
patients	controls	patients	controls	patients	controls	patients	controls	patients	controls	patients	controls
0.25	**0.85**	0.52	**0.62**	0.61	**0.69**	0.77	**0.94**	0.67	**0.88**	0.74	**0.91**

Fig. 6. Fusion or combination responses in ID as a function of simultaneity responses in SIM. Reported data only include those participants (patients in dark gray, controls in light gray) with illusory reports gathered across both studies. (a) Rate of fusion as a function of simultaneity responses to fusion stimuli in fusers of Study 1, (b) rate of fusion as a function of simultaneity responses to fusion stimuli in fusers of Study 2, (c) rate of fusion as a function of simultaneity responses to congruent speech in fusers of Study 2, (d) rate of combination as a function of simultaneity responses to combination stimuli in combiners of Study 1, (e) rate of combination as a function of simultaneity responses to combination stimuli in combiners of Study 2, (f) rate of combination as a function of simultaneity responses to congruent speech in combiners of Study 2. Linear fits equations (y) and goodness of fits ($r2$) are provided on the graphs: patients on the left, controls on the right. Temporal profiles in SIM are systematically better predictors of fusion and combination profiles in controls compared to patients: correlation coefficients are provided for each population in the table.

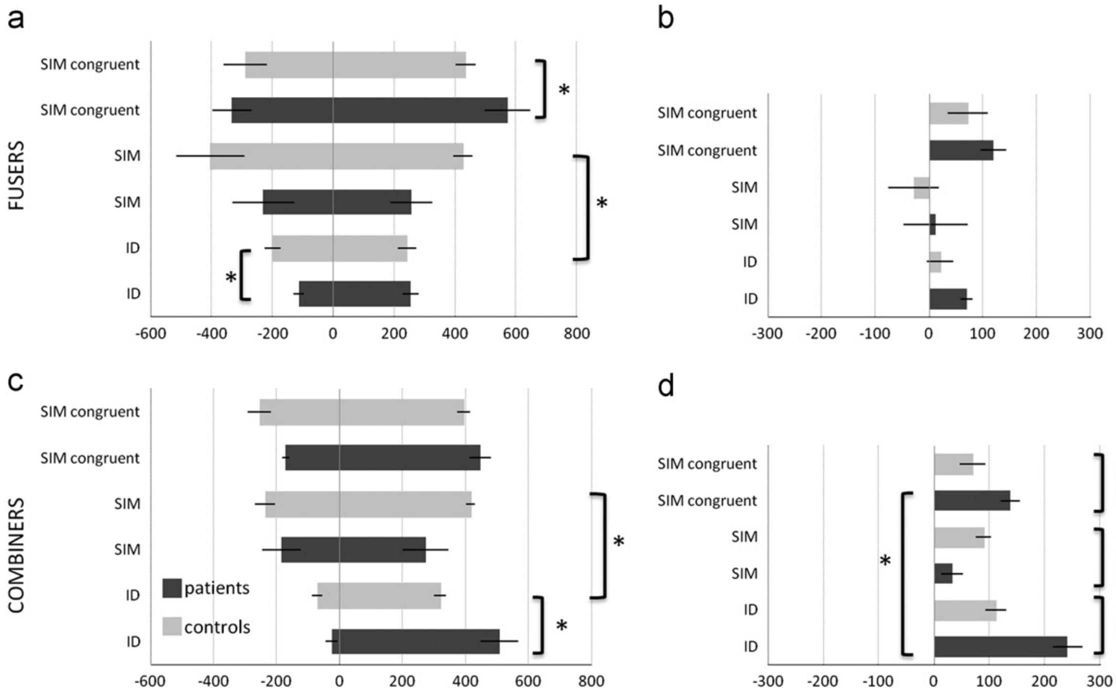


Fig. 7. Temporal windows of integration derived from ADS fits. Reported data only include those participants (patients in dark gray, controls in light gray) with illusory reports gathered across both studies. (a) Width of temporal window of integration as defined by jnds for the fusers subgroup of Study 1 and 2 combined for the ID and SIM tasks with fusion stimuli and SIM task with congruent AV speech, (b) assymetric point for the Fusers subgroup for the same tasks, (c) width of the temporal windows of integration in the combiners subgroup of Study 1 and 2 combined for the ID and SIM tasks with combination stimuli and SIM task with congruent AV speech, (d) assymetric point for the combiners subgroup in the same tasks. Error bars are two standard errors from the mean.

Table 3
Summary of correlations between PANSS scores, medication and behavioral indices.

		Fusion (*r*2)			Combination (*r*2)		
		Illusion	TWI (ID)	TWI (SIM)	Illusion	TWI (ID)	TWI (SIM)
PANSS scores	Positive	0.0003	0.095	0.275	0.05	0.082	0.093
	Negative	0.28	0.0000003	0.015	0.002	0.092	0.0005
	Global	0.00002	0.0054	0.046	0.14	0.004	0.001
	Total	0.0002	0.0091	0.066	0.2	0.011	0.002
Chlorpromazin-equivalent		0.047	0.135	0.112	0.003	0.001	0.004

between controls and patients (ID: $t_{1,22}=3.982$, $p \leq 0.001$; SIM: $t_{1,22}=-2.189$, $p \leq 0.039$; Fig. 7d). This was also observed for congruent speech ($t_{1,9}=-11.97$, $p \leq 0.0001$). Surprisingly, patients did not show differences in the width of their TWI between the ID and SIM tasks yet showed a significant difference in their PSS values between the two tasks ($t_{1,22}=5.936$, $p \leq 0.0001$) which were much larger in the combination than in the fusion task and larger than those observed in controls.

Overall, in both fusion and combination, controls tended to systematically show a larger TWI in SIM than in ID tasks, suggesting that only when AV speech stimuli were perceived as simultaneous (SIM) would AV integration (ID) take place. In patients, this pattern was not as robust: for instance, the boundaries of the TWI in ID were highly dependent on the type of stimulus (congruent, fusion or combination) such that more integration were seen for audio lead in fusion and for audio lag in combination.

3.5. *Extrinsic factors: Medication and PANNS scores*

To ensure that the reported results were not confounded by the medical treatments undertaken by the patients, we looked at potential correlations between equivalent-chlorpromazin and behavioral indices of interest. For these correlations, we used the rate of illusory reports (MAX) and the width of the temporal windows of integration in ID and SIM tasks. None of the results were significantly correlated with medication (Table 3). We were also interested in checking whether any correlation would be found between the same indices and with the screened PANSS scores. However, none of the PANSS scores were good predictors on the behavioral measures (Table 3).

4. Discussion

4.1. *Summary of findings*

Contrary to our expectations and previous reports, no significant differences were found in the rates of McGurk fusion or combination between patients with schizophrenia and their matched controls, neither at synchrony, nor when considering the maximal illusion rate irrespective of asynchrony value. Additionally, no major differences in recognition scores for auditory or visual speech alone were found. This suggests that there is no major impairment of AV speech integration in schizophrenia. Differences observed in the temporal constraints of AV speech integration (ID) between patients and controls affected the width of the temporal window: in patients, the temporal windows of integration were found to be smaller on the auditory lead side for fusion stimuli and larger on the auditory lag side for combination stimuli compared to controls under the same conditions. This would suggest that integration of AV speech is less tolerant in patients than in controls. However, this is in contrast with the differences of temporal profiles observed in the subjective simultaneity task (SIM): patients showed larger temporal profiles than those of controls although their temporal windows of integration did not significantly differ. This suggests impairments in patients are seen most easily at higher AV asynchrony values. As such, and importantly, patients' temporal profiles obtained in the SIM task did not robustly predict those observed in the ID task: using tasks of explicit temporal judgments (SIM) in patients captures only partially the integrative properties of perceptual systems (ID). All together, these results suggest a generic impairment in patients with schizophrenia that is not reducible to a deficit in perceptual binding but rather to the temporal structuring of events in time that may impair typical binding mechanisms.

4.2. *Structuring events in time: Distinction between implicit and explicit event structuring*

Overall, patients showed a larger tolerance profile to AV speech asynchronies (SIM) in incongruent (three out of four cases, Fig. 3) and congruent speech. These results are in line with prior findings on simultaneity judgments in multisensory context (Foucher et al., 2007) and within sensory modalities (Giersch et al., 2009; Lalanne et al., 2012; Schwartz et al., 1984) namely, patients consider AV information to be simultaneous for a larger range of asynchronies than controls, notably beyond the temporal window of integration (as defined here by jnds). One strong a priori under the unity assumption (e.g., Vatakis and Spence (2007)) is that events ought to be perceived as simultaneous (or emanating from the same cause) for them to be integrated. However, it appears that the unity assumption is necessary but clearly not sufficient for integration especially for patients with schizophrenia: for instance, at large SOA values, both patients and controls show integration despite judging stimuli to be at nearly 100% desynchronized (Fig. 5b and d).

In AV speech and in ecological stimuli (naturally complex and evolving over time), some of the basic operations that needs to be solved are the parsing of sensory information and the structuring of this information in time in order to bind or segregate information within and across sensory modalities. One proposed view of event segmentation (e.g., Zacks & Tversky, 2001) suggests three major properties for these operations: they are predictive, recurrent and cyclical. In this view, cognitive operations can naturally apply attentional parsing to incoming sensory events (cf. for instance Jones (1976)). Recent neurophysiological work provides a mechanistic implementation of such parsing mechanisms notably by ways of neural oscillations (Schroeder, Lakatos, Kajikawa, Partan, & Puce, 2008). The internal set of temporal parsing mechanisms (neural oscillations in different frequency bands) provides the logistical platform for automatic temporal structuring in the brain (Pöppel, 2009; van Wassenhove, 2009). Patients' wider temporal profiles in SIM suggest two possible interpretations: (i) parsing mechanisms in patients are slower with the implication that speech integration and temporal perception relies directly on the parsing or (ii) specific operations engaged in the parsing and/or segregration/binding of temporal features are noisier and lack reliability (specifically, finer levels of parsing are impaired).

Against the first hypothesis, ID profiles in schizophrenia tended to be narrower despite identical levels of speech integration compared to controls: this suggests that irrespective of potential parsing problems, the informational content for speech processing is sufficiently well encoded to permit AV speech integration – fusion or combination. Thus, differences between patients and controls are more subtle: if speech processing is not impaired, it is the encoding of temporal information which may be so. This is supported by the difficulties for patients with schizophrenia to judge simultaneity. Additionally, controls and patients showed a larger permissible temporal window of integration in the SIM task than in the ID task: this suggests the existence of an additional operation for the extraction of temporal features to serve a simultaneity task (temporal awareness). It is, we argue, those specific operations pertaining to accessing temporal content (not speech content) that are impaired in schizophrenia.

When task demands require an explicit comparison of temporal features (SIM), the most parsimonious approach would be to compare an auditory and a visual cue in the dynamic stream of events. Such process is not required in the ID task (by virtue of integration). AV simultaneity judgments rely both on the comparison and the reliability of the temporal parsing mechanisms in each sensory modality. As such, if one of the two sensory modalities is sluggish, it would be reflected in the temporal profile in SIM. An additional possibility is that the coordination of temporal parsing mechanisms between auditory and visual sensory modalities is impaired. In the context of speech, two natural time scales have been argued to be necessary: the subphonetic features (a few tens of milliseconds) and syllables (a couple hundred of milliseconds) scales (Poeppel, 2003; Poeppel, Idsardi, & van Wassenhove, 2008; Giraud & Poeppel, 2012). These two time-scales are considered necessary for the discretization of (AV) speech information eventually interfacing with the demands of the linguistic system (Giraud & Poeppel, 2012). If those remain functional in schizophrenia, we suggest that it is the coordination of these two parsing mechanisms that may be impaired. The changes of PSS in the ID task (reflecting the dependency on visual encoding) and the enlargement of the temporal profiles in SIM observed in patients are consistent with this hypothesis.

It is noteworthy that previous studies found enlarged windows of simultaneity within a single sensory modality (Giersch et al., 2009; Lalanne, van Assche, & Wang Giersch, in press; Schwartz et al., 1984). Yet, and similarly to the present study (ID), patients with schizophrenia displayed a high sensitivity to short asynchronies at an implicit level (Lalanne et al., in press). The weakened link found between the predictability of SIM profile with the ID profile in patients with schizophrenia further support a partial dissociation between implicit temporal processing and explicit access to time. Further support for the second hypothesis-impaired explicit access to temporal features-can be found in the enlarged SIM window that may reflect an impaired temporal processing and temporal binding sensitivity. Neural synchronization problems between temporal parsers may provide a generic basis for this impairment and could be observable at multiple levels of cognitive operations in patients with schizophrenia, in line with recent neurophysiological hypotheses (Uhlhaas & Singer, 2010).

4.3. *Hypotheses on the neural bases of a deficit in temporal event structuring in patients with schizophrenia*

With regards to more classic models of temporal processing, it is important to note that the tasks that have been used here do not rely on duration estimation but on synchrony judgments. Additionally, they do not cover a range of timing classically supported by interval timing mechanisms reaching the second range (Buhusi & Meck, 2005) and, as such, cannot straightforwardly be interpreted within a perturbed dopaminergic system. Very little is known about the neurophysiology of short timing mechanisms below the second range and even less when a simultaneity judgment (as opposed to duration estimation) is required. Nevertheless, as suggested above, the range of frequency regions in cortical oscillations impaired in patients with schizophrenia suggests a generic impairment of temporal processes in support of cognition (Uhlhaas & Singer, 2010).

The working hypothesis that we develop here straightforwardly maps onto the notion that the coordination of particular oscillatory mechanisms is impaired in patients with schizophrenia: the relevance for behavior of this posited intrinsic neural noise is not trivial to tease apart without fine-tuned psychophysical paradigms. Recent computational models have started to address this issue by testing biologically-plausible neural networks and showing how alterations in synaptic function can lead to systematic oscillatory disruptions (Rolls & Deco, 2011). Cortical oscillations are largely influenced by different neuromodulators: glutamate is implicated in nearly all cortical oscillations whereas the implication of dopamine has been essentially tested in the beta (13–30 Hz) and somewhat in the gamma (40 Hz) bands (Uhlhaas et al., 2008). Synaptic alterations will affect the time constant of the oscillatory activity in a particular frequency band, its duration, and importantly, the strength of synchronization in short- or long-range connectivity throughout the network. At this stage, and as recently concluded (Uhlhaas & Singer, 2010), a major effort is needed to bridge neurosciences with the phenomenological specificities of schizophrenia. Here, we suggest that the temporal noise and weakened oscillatory connectivity observed in patients with schizophrenia is captured in the enlarged temporal windows of integration and their variable boundaries. Future work using magneto- or electro-encephalography (MEG or EEG, respectively) would allow this hypothesis to be tested directly.

If generic and large-scale neural perturbations are central to the effects reported here, an additional line of research on the neural bases of AV speech integration has also permitted refined advances in the understanding of binding mechanisms across sensory modalities. Of particular interest for this study, recent fMRI findings suggest specialized neural populations in an area of cortex well known for its multisensory properties, namely the superior temporal sulcus (STS in monkey) or superior temporal cortex (STC, human homolog). The organization of this multisensory region is particularly difficult to unravel (Beauchamp, Argall, Bodurka, Duyn, & Martin, 2004) yet recognized to be an essential part of the AV speech integration network (Arnal et al., 2009; Beauchamp, Nath, & Pasalar, 2010). The middle STC (mSTC) is a prime area for the detection of asynchrony and the integration of AV speech information (Bushara, Grafman, & Hallett, 2001; Miller & D'Esposito, 2005; Stevenson, Altieri, Kim, Pisoni, & James, 2010; Stevenson, VanDerKlok, Pisoni, & James, 2011). Recent investigations suggest that at least two neural subpopulations coexist in this region: the synchrony population tagged S-mSTC showing increased activation to AV speech stimuli when the auditory and visual streams are in synchrony and the bimodal population tagged B-mSTC showing the opposite pattern, namely a decrease of activation with the presentation of synchronized audiovisual speech streams (Stevenson et al., 2010, 2011). These results may help disambiguate the role that some neural subpopulations in mSTC may play i.e., pass speech or time relevant information to higher processing stages in cortex.

Interestingly, patients with schizophrenia show some functional impairment in these regions of the STS (see discussion in Stevenson et al., 2011) but it is unclear to which extent this could support our current findings.

One hypothesis then is that a thorough description of neural oscillations implicated in the tasks used here could help making a link between the putative role of neural populations in STC and the coordination of neural oscillations between brain regions (notably auditory and visual cortices, here). Neuroimaging techniques such as MEG and EEG can help disentangle whether the dissociation between accessing speech (ID) or time (SIM) information, and Maier et al., 2011).

4.4. No impairment of AV speech integration in patients with schizophrenia?

Our results do not concur with prior reports showing a specific impairment of AV speech integration in patients with schizophrenia (de Gelder et al., 2003; Ross et al., 2007) but are in line with some other studies (Myslobodsky, Goldberg, Johnson, Hicks, & Weinberger, 1992; Surguladze et al., 2001). It is noteworthy that no major differences between patients and controls were found despite our strict criterion for fusion reports, namely fusion was only considered to have taken place when participants reported “da” (i.e., neither the auditory or visual percept). “ga” reports (visually-driven responses) were not considered a case of AV speech integration (van Wassenhove et al., 2007). Our quantification of AV speech integration was thus more conservative and constitutes one major difference with prior studies (de Gelder et al., 2003). A second possible reason for the discrepancies with earlier findings is that previous reports used a smaller sample of patients with larger variability of age (de Gelder et al., 2003). Additionally, one study by Pearl and colleagues (2009) showed that AV speech impairments in schizophrenia may be confined to younger patients (children and adolescents). A third important difference with prior reports showing AV speech impairments in schizophrenia (Ross et al., 2007) is the signal-to-noise ratio of the auditory speech signal. In AV speech studies, it is well-known that visual speech benefits auditory comprehension mostly under noisy conditions (Grant & Seitz, 2000; Ross et al., 2007). In the study of Ross et al., (2007), patients with schizophrenia showed AV speech impairments under low SNR levels (e.g., -12 dB); in our study, SNR was not manipulated and we thus cannot conclude on the possibility that in noisy environments, patients may show less multisensory integration than controls. A fourth intriguing possibility suggested by our data is that maximal illusory rates are not necessarily observed at natural synchrony in both patients and controls; in both fusion and combination, patients tended to show higher integration for larger visual leads than controls (cf. Fig. 1d and Fig. 7a and c). Hence, one possibility is that the optimal delay between AV speech information may differ for patients and controls. Under these circumstances, prior studies may have underestimated patients' ability to integrate AV speech information since “natural synchrony” was used (note that for dubbed AV speech stimuli such as McGurk fusion and combination, natural synchrony is meaningless). It would be informative to test a wider set of AV speech stimuli and see whether the degree of asynchrony for maximal illusory rates systematically differs for patients and controls. This could indicate different temporal characteristics in the binding operations of AV speech and provide a refined insight on possible AV speech integration impairments in schizophrenia.

4.5. AV speech integration in time

In patients, AV speech integration showed a trend towards being less tolerant to AV asynchronies as compared to controls. In fusion, the auditory lead tolerance was significantly shortened; in combination, tolerance to auditory lags was significantly lengthened. Although patients were not primarily impaired in the strength of AV speech integration compared to controls, the integration process appears to operate under different temporal constraints.

Recent predictive models of AV speech integration emphasize the predictive role of visual speech information on auditory speech categorization. Visual speech often precedes the auditory utterance (Chandrasekaran, Trubanova, Stillittano, Caplier, & Ghazanfar, 2009): this natural delay enables the visual system to extract information relevant to speech which can in turn predict the impending auditory utterance (Arnal et al., 2009, 2011; Poeppel et al., 2008; van Wassenhove et al., 2005). In combination, patients tolerate more auditory lag suggesting that they may be less visually-driven than controls: specifically, the speech system would require more evidence than that solely provided by visual speech for the categorization of a given speech token and hence rely on auditory information more. Alternatively, the phonological categorization of visual speech may be slower in patients (in line with the visual leads side of the curve taken as the time needed to encode visual speech information and the longer PSS values; cf. Supp Mat. Fig. 2). Consistent with this interpretation, patients' temporal window in fusion is shortened for audio leads compared to controls: auditory information provides sufficient evidence to allow speech categorization thereby preventing visual speech to modify the perceptual outcome.

Hence, two important observations emerge. First, although patients did not show impaired fusion or combination, the temporal constraints with which AV speech binding occurs slightly differ from controls. Differences observed in the temporal profiles and in the temporal windows of integration in ID suggest that the extraction of visual speech information in patients may be less robust and perhaps slowed down compared to that of controls. This is in line with prior findings showing impaired AV speech integration in patients with schizophrenia when auditory speech is presented with low SNR (Ross et al., 2007). Second, this is also in agreement with visual speech deficit in schizophrenia (de Gelder et al., 2003): within a predictive coding framework of speech processing, we suggest that it is the strength of the visual prediction in time which may be impaired in patients, not visual speech processing per se.

To substantiate this working hypothesis, the impaired temporal organization of neural activity in patients with schizophrenia (Uhlhaas & Singer, 2010) is crucial to consider within the recent predictive models of speech processing: the coordination across neural frequency bands have been shown to be crucial (Arnal et al., 2011) in AV speech binding and it would thus be interesting to see whether the variability observed at different SOAs can be captured by these neurophysiological indices. To our knowledge, no data currently exist on this topic in patients with schizophrenia.

5. Conclusions

In these two studies, we have shown that while patients with schizophrenia do not show major deficits in AV speech integration, some differences in the temporal handling of AV speech information persist which are particularly salient when participants are asked to report the explicit timing of events. Based on the pattern of results, we suggest that the temporal structuring of events in patients is fundamentally difficult to access and likely to impair processes requiring explicit access to temporal information. We emit the hypothesis that subtle temporal features in the binding operations of sensory events and that generic neural oscillatory dysfunctions are at the core of timing deficit in schizophrenia.

6. Funding

This work was supported by the Fondation pour la Recherche Médicale "FRM Mastere Recherche" awarded to B.M. and IRG-249222, ERC-StG-263584 and ANRJCJC-1904 granted to VvW. This work was conducted under the European project COST ISCH Action TD0904 "Time In MEntaL activitY: theoretical, behavioral, bioimaging and clinical perspectives (TIMELY; www.timely-cost.eu).

Acknowledgements

We would like to thank the Psychiatry Department of the University of Strasbourg, the Association Elan Retrouvé (Paris, France) and the NeuroSpin medical staff for their dedication and their precious help with the recruitment of patients with schizophrenia and their matched controls. We also thank two anonymous reviewers for their critical comments on an earlier version of this manuscript. The authors report no competing interest.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.neuropsychologia.2012.07.002.

References

Andreasen, N. C. (1999). A unitary model of schizophrenia: Bleuler's fragmented phrene as schizencephaly. *Archives of General Psychiatry*, *56*, 781–787.

Arnal, L. H., Morillon, B., Kell, C. A., & Giraud, A. L. (2009). Dual neural routing of visual facilitation in speech processing. *Journal of Neuroscience*, *29*(43), 13445–13453, http://dx.doi.org/10.1523/JNEUROSCI.3194-09.2009.

Arnal, L. H., Wyart, V., & Giraud, A. L. (2011). Transitions in neural oscillations reflect prediction errors generated in audiovisual speech. *Nature Neuroscience*, *14*(6), 797–801, http://dx.doi.org/10.1038/nn.2810.

Beauchamp, M. S., Argall, B. D., Bodurka, J., Duyn, J. H., & Martin, A. (2004). Unraveling multisensory integration: patchy organization within human STS multisensory cortex. *Nature Neuroscience*, *7*, 1190–1192, http://dx.doi.org/10.1038/nn1333.

Beauchamp, M. S., Nath, A. R., & Pasalar, S. (2010). fMRI-guided transcranial magnetic stimulation reveals that the superior temporal sulcus is a cortical locus of the McGurk effect. *Journal of Neuroscience*, *30*, 2414–2417, http://dx.doi.org/10.1523/JNEUROSCI.4865-09.2010.

Brainard, D. H. (1997). The psychophysics toolbox. *Spat Vis*, *10*, 433–436.

Buhusi, C. V., & Meck, W. H. (2005). What makes us tick? Functional and neural mechanisms of interval timing. *Nature Reviews Neuroscience*, *6*(10), 755–765.

Bushara, K. O., Grafman, J., & Hallett., M (2001). Neural correlates of auditory–visual stimulus onset asynchrony detection. *Journal of Neuroscience*, *21*, 300–304.

Buzsáki, G. (2006). *Rhythms of the Brain*. Oxford, UK: Oxford University Press.

Buzsáki, G. (2010). Neural syntax: cell assemblies, synapsembles, and readers. *Neuron*, *68*, 362–385.

Chandrasekaran, C., Trubanova, A., Stillittano, S., Caplier, A., & Ghazanfar, A. A. (2009). The natural statistics of audiovisual speech. *PLoS Computational Biology*, *5*, e1000436, http://dx.doi.org/10.1371/journal.pcbi.1000436.

Conrey, B., & Pisoni, D. B. (2006). Auditory–visual speech perception and synchrony detection for speech and nonspeech signals. *Journal of the Acoustical Society of America*, *119*, 4065–4073.

Davalos, D. B., Kisley, M. A., & Freedman, R. (2005). Behavioral and electrophysiological indices of temporal processing dysfunction in schizophrenia. *Journal of Neuropsychiatry and Clinical Neurosciences*, *17*, 517–525.

de Gelder, B., Vroomen, J., Annen, L., Masthof, E., & Hodiamont, P. (2003). Audio–visual integration in schizophrenia. *Schizophrenia Research*, *59*, 211–218.

de Gelder, B., Vroomen, J., de Jong, S. J., Masthoff, E. D., Trompenaars, F. J., & Hodiamont, P. (2005). Multisensory integration of emotional faces and voices in schizophrenics. *Schizophrenia Research*, *72*, 195–203, http://dx.doi.org/10.1016/j.schres.2004.02.013.

Dennett, D., & Kinsbourne, M. (1992). Time and the observer. *Behavioral and Brain Sciences*, *15*, 183–247.

Elvevåg, B., McCormack, T., Gilbert, A., Brown, G. D., Weinberger, D. R., & Goldberg, T. E. (2003). Duration judgements in patients with schizophrenia. *Psychol. Med.*, *33*(7), 1249–1261.

Foucher, J. R., Lacambre, M., Pham, B.-T., Giersch, A., & Elliott, M. A. (2007). Low time resolution in schizophrenia lengthened windows of simultaneity for visual, auditory and bimodal stimuli. *Schizophrenia Research*, *97*, 118–127, http://dx.doi.org/10.1016/j.schres.2007.08.013.

Fuchs, T. (2007). The temporal structure of intentionality and its disturbance in schizophrenia. *Psychopathology*, *40*, 229–235.

Giersch, A., Lalanne, L., Corves, C., Seubert, J., Shi, Z., Foucher, J., et al. (2009). Extended visual simultaneity thresholds in patients with schizophrenia. *Schizophrenia Bulletin*, *35*, 816–825.

Giraud, A. L., & Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nature Neuroscience*, *15*, 511–517, http://dx.doi.org/10.1038/nn.3063.

Grant, K. W., & Seitz, P. F. (2000). The recognition of isolated words and words in sentences: individual variability in the use of sentence context. *The Journal of the Acoustical Society of America*, *107*, 1000–1011.

Jones, M. R. (1976). Time, our lost dimension: toward a new theory of perception, attention, and memory. *Psychological Review*, *83*(5), 323–355, http://dx.doi.org/10.1037/0033-295X.83.5.323.

Kimura, B. (1994). Psychopathologie der Zufalligkeit. *Daseinsanalyse*, *11*, 192–204.

Lalanne, L., van Assche, M., & Giersch, A. (2012). When predictive mechanisms go wrong: disordered visual synchrony thresholds in schizophrenia. *Schizophrenia Bulletin*, *38*, 506–513, http://dx.doi.org/10.1093/schbul/sbq107.

Lalanne, L., van Assche, M., Wang, W., & Giersch, A. (under review). Fragmented perception selectively related to time in patients with Schizophrenia. *Neuropsychologia*.

Maier, J. X., Di Luca, M., & Noppeney, U. (2011). Audiovisual asynchrony detection in human speech. *Journal of Experimental Psychology: Human Perception and Performance*, *37*, 245–256, http://dx.doi.org/10.1037/a0019952.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*, 746–748.

Miller, L. M., & D'Esposito, M. (2005). Perceptual fusion and stimulus coincidence in the cross-modal integration of speech. *Journal of Neuroscience*, *25*, 5884–5893, http://dx.doi.org/10.1523/JNEUROSCI.0896-05.2005.

Minkowski, E. (1933). *Le temps vécu*. Paris: Presses Universitaires de France.

Munhall, K., Gribble, P., Sacco, L., & Ward, M. (1996). Temporal constraints on the McGurk effect. *Perception and Psychophysics*, *58*, 351–362.

Myslobodsky, M. S., Goldberg, T., Johnson, F., Hicks, L., & Weinberger, D. R. (1992). Lipreading in patients with schizophrenia. *Journal of Nervous and Mental Disease*, *180*, 168–171.

Pearl, D., Yodashkin-Porat, D., Katz, N., Valevski, A., Aizenberg, D., Sigler, M., et al. (2009). Differences in audiovisual integration, as measured by McGurk phenomenon, among adult and adolescent patients with schizophrenia and age-matched healthy control groups. *Comprehensive Psychiatry*, *50*, 186–192.

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: cerebral lateralization as asymmetric sampling in time. *Speech Communication*, *41*, 245–255.

Poeppel, D., Idsardi, W. J., & van Wassenhove, V. (2008). Speech perception at the interface of neurobiology and linguistics. *Philosophical Transactions of the Royal Society of London B*, *363*, 1071–1086.

Pöppel, E. (2009). Pre-semantically defined temporal windows for cognitive processing. *Philosophical Transactions of the Royal Society of London B*, *364*, 1887–1896.

Rolls, E. T., & Deco, G. (2011). A computational neuroscience approach to schizophrenia and its onset. *Neuroscience & Biobehavioral Reviews*, *35*(8), 1644–1653.

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

Sams, M., Aulanko, R., Hämäläinen, M., Hari, R., Lounasmaa, O. V., Lu, S.-T., et al. (1991). Seeing speech: visual information from lip movements modifies activity in the human auditory cortex. *Neuroscience Letters*, *127*, 141–145, http://dx.doi.org/10.1016/0304-3940(91)90914-F.

Schmidt, H., McFarland, J., Ahmed, M., McDonald, C., & Elliott, M. A. (2011). Low-level temporal coding impairments in psychosis: preliminary findings and recommendations for further studies. *Journal of Abnormal Psychology*, *120*, 476–482.

Schroeder, C. E., Lakatos, P., Kajikawa, Y., Partan, S., & Puce, A. (2008). Neuronal oscillations and visual amplification of speech. *Trends in Cognitive Sciences*, *12*, 106–113, http://dx.doi.org/10.1016/j.tics.2008.01.002.

Schwartz, B. D., Winstead, D. K., & Walker, W. G. (1984). A corpus callosal deficit in sequential analysis by schizophrenics. *Biological Psychiatry*, *19*, 1667–1676.

Stevenson, R. A., Altieri, N. A., Kim, S., Pisoni, D. B., & James, T. W. (2010). Neural processing of asynchronous audiovisual speech perception. *NeuroImage*, *49*, 3308–3318, http://dx.doi.org/10.1016/j.neuroimage.2009.12.001.

Stevenson, R. A., VanDerKlok, R. M., Pisoni, D. B., & James, T. W. (2011). Discrete neural substrates underlie complementary audiovisual speech integration processes. *NeuroImage*, *55*, 1339–1345, http://dx.doi.org/10.1016/j.neuroimage.2010.12.063.

Surguladze, S. A., Calvert, G. A., Brammer, M. J., Campbell, R., Bullmore, E. T., Giampietro, V., et al. (2001). Audio–visual speech perception in schizophrenia: an fMRI Study. *Psychiatry Research*, *106*, 1–14.

Theunissen, F., & Miller, J. P. (1995). Temporal encoding in nervous systems: a rigorous definition. *Journal of Computational Neuroscience*, *2*, 149–162, http://dx.doi.org/10.1007/bf00961885.

Uhlhaas, P. J., & Mishara, A. L. (2007). Perceptual anomalies in schizophrenia: integrating phenomenology and cognitive neuroscience. *Schizophrenia Bulletin*, *33*(1), 142–156.

Uhlhaas, P. J., Haenschel, C., Nikolić, D., & Singer, W. (2008). The role of oscillations and synchrony in cortical networks and their putative relevance for the pathophysiology of schizophrenia. *Schizophrenia Bulletin*, *34*(5), 927–943.

Uhlhaas, P. J., & Singer, W. (2010). Abnormal neural oscillations and synchrony in schizophrenia. *Nature Reviews Neuroscience*, *11*, 100–113, http://dx.doi.org/10.1038/nrn2774.

van Wassenhove, V. (2009). Minding time in an amodal representational space. *Philosophical Transactions of the Royal Society of London B*, *364*, 1815–1830, http://dx.doi.org/10.1098/rstb.2009.0023.

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National academy of Sciences of the United States of America*, *102*, 1181–1186, http://dx.doi.org/10.1073/pnas.0408949102.

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2007). Temporal window of integration in auditory–visual speech perception. *Neuropsychologia*, *45*, 598–607, http://dx.doi.org/10.1016/j.neuropsychologia.2006.01.001.

Vatakis, A., & Spence, C. (2007). Crossmodal binfing: evaluating the "unity assumption" using audiovisual speech stimuli. *Percept Psychophys*, *69*(5), 744–756.

Vogeley, K., & Kupke, C. (2007). Disturbances of time consciousness from a phenomenological and neuroscientific perspective. *Schizophrenia Bulletin*, *33*, 142–156.

Volz, H. P., Nenadic, I., Gaser, C., Rammsayer, T., Häger, F., & Sauer, H. (2001). Time estimation in schizophrenia: an fMRI study at adjusted levels of difficulty. *Neuroreport*, *12*, 313–316.

Vroomen, J., & Keetels, M. (2010). Perception of intersensory synchrony: a tutorial review. *Attention, Perception & Psychophysics*, *72*, 871–884.

Wang, XJ. (2010). Neurophysiological and computational principles of cortical rhythms in cognition. Physiological Reviews, 90, 1195-1268. 10.11552/physrev.00035.2008.

Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychology Bulletin*, *88*, 638–667.

Wittmann, M. (2011). Moments in time. *Front Integrative Neuroscience*, *5*, 1–9, http://dx.doi.org/10.3389/fnint.2011.00066.

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: a mind-brain perspective. *Psychological Bulletin*, *133*, 273–293, http://dx.doi.org/10.1037/0033-2909.133.2.273.

Zacks, J. M., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin*, *127*, 3–21.