Self-monitoring in speech production: effects of verbal hallucinations and negative symptoms

I. LEUDAR, P. THOMAS AND M. JOHNSTON

From the Department of Psychology and Department of Education, University of Manchester; and the Department of Psychological Medicine, University of Wales College of Medicine

SYNOPSIS This paper reports results of a study on self-monitoring in speech production. Thirty schizophrenics, varying in verbal hallucination and in negative symptoms status, and 17 controls were tested on the reporter test. The position of interruptions of the speech-flow to repair errors was used to indicate whether the detection of the errors was through monitoring of internal phonetic plans or through external acoustic feedback. We have found that the internal error detection was twice as frequent in controls as in schizophrenics. The relevance of this finding to Frith's (1992) model of schizophrenia is discussed. Our conclusion is that the problem with internal monitoring of phonetic plans is common to all schizophrenics, and not just to those with verbal hallucinations.

INTRODUCTION
Planning, self-monitoring and verbal hallucinations

Planning is one of the well researched areas in cognitive science (Fikes & Nilsson, 1971; Sacerdotti, 1977; Wilensky, 1983; Appelt, 1985; Wilkins, 1988; Allen, 1989), and the discussion of the relationship between planned and situated actions is one of the more interesting recent controversies (Winograd & Flores, 1986; Suchman, 1987). The research on planning activities has been recently made relevant to symptoms of schizophrenia, or if Bentall (1990) is correct, the phenomena which are interpreted mistakenly as symptoms of schizophrenia.

Hoffman (1986a, b) explains verbal hallucinations in terms of planning process disruptions. He argues that verbal hallucinations are produced in a two-stage process. The first necessary condition is the occurrence of fragments of plans, or more precisely plan bodies, unrelated to agents' current goals. Hoffman refers to these as verbal images. According to him these are produced especially in difficult planning situations, but on the whole the variables, which specifically cause plan fragmentation, are not clear.

The second necessary condition is that these verbal images are interpreted intentionally, and if attributed to an outside agency, may be experienced as voices. In other words, hallucinators misattribute their own thoughts to external sources. In this respect, Hoffman's explanation is like that of Bentall & Slade's (1985) in terms of 'reality monitoring'. Agent's thoughts are externalized and interpreted as belonging to another agent. Indeed, there is some measure of agreement in the literature that voices are not interpretations of external stimuli, but are identified by an experience of unintendedness (cf. Kass, 1968; Horowitz, 1975; Frith, 1987; Frith & Done, 1988).

The process of misattribution is, however, not clarified by Hoffman's and Frith's models. It is likely to be a complex process. This is because the misattribution does not simply involve treating one's own thoughts as belonging to another agent (the voice) but it also involves the construction of the voice's social identity. Frith (1992) explains 'signs and symptoms of schizophrenia' in terms of deficiencies in producing actions. According to him positive symptoms (e.g. verbal hallucinations, thought insertion, delusions of control) can be accounted for in terms of deficiencies in internal monitoring of intentions, and negative symptoms (e.g. avolition, poverty of thought) are related to deficiencies in goal selection, and plan and action initiation. Following Goldberg (1985), Frith...
argues that there are two neuro-cognitive systems, one for producing actions in response to environmental stimuli, the other for willed actions. The function of monitoring is to detect errors of planning and execution rapidly, and to keep track whether an activity arose as a reaction to a stimulus or as a result of an intention (Frith, 1987, 1992; Frith & Done, 1988).

It is, however, not certain that there is a psychologically unitary mechanism of internal monitoring. In models of planning it is usual to distinguish goal selection, plan formulation and plan execution. All these processes and their products may be monitored. Planner SIPE (Wilkins, 1988) for example, may temporarily produce plans which are not valid. It employs plan critics, who check the correctness of proposed plans on a variety of criteria (e.g. interactions between unordered actions, resource conflicts, etc.). SIPE also employs a monitor of effects of actions in the environment. On close reading, Frith (1987) in fact postulates more than one kind of monitoring: of intentions, of action attempts and of environmental effects of actions. Only the former two are said to be deficient in schizophrenia.

Frith (1992) suggests that in positive symptoms there is a dissociation between intentions and action – patients perform actions, but do not experience themselves as being the intentional agents of those actions. According to Frith this is due to deficient monitoring of intentions. In ‘delusions of control’ there is a failure to monitor the intention to act (the intention is lost to the patient, who subsequently infers falsely that the action was caused by some other agent). In thought insertion, patients think, but do not experience their thoughts as their own and they may interpret their thoughts as having been inserted by another agent.

How can hearing voices be explained in this model? The tacit assumption is that thoughts are like speech acts, and have representational content and psychological mode (i.e. they are intentions, beliefs, desires, cf. Searle, 1983). If the psychological mode is lost (maybe due to defective internal monitoring), only the representational content remains. This content may be experienced as a voice if, as Frith puts it, ‘the process took the form of inner speech’. The difference, thus, between Hoffman’s and Frith’s models is that in the former account it is fragments of plans that are interpreted as voices, in the latter account, it is ‘representational content’ stripped of its ‘psychological mode’ that is so represented.

According to Frith, in subjects with negative symptoms, the will fails to be translated into action. Responses, i.e. stimulus-elicited actions are, however, still possible. This may result in perseverations and poverty of action. It’s not, however, clear what happens in negative symptoms on occasions when a desire becomes an intention and is translated into an action, as clearly occurs sometimes. Is there an additional problem with internal monitoring?

There is some empirical evidence to support Hoffman’s and Frith’s models. Chapman (1966) has reported that schizophrenics frequently experience their speech as poorly matched to what they intended to say. Frith & Done (1989) and Malenka et al. (1982) report problems with internal monitoring of non-verbal actions in schizophrenics. Hoffman (1986b) and Andreasen et al. (1986) found that severity of discourse disorganization correlates with the incidence of verbal hallucinations.

Leudar et al. (1992) studied self-repair in speech. They found a relatively high frequency of errors in schizophrenics, possibly indicating problems with internal monitoring of semantic intentions. This finding was partly explained in terms of limitations of working memory. The putative semantic monitoring problem was, however, common to schizophrenics with verbal hallucinations and negative symptoms and not specific to the former group, as Frith’s model implies. Leudar et al. (1992) also found, however, that providing that a word, or a fragment of it, has been articulated in error, schizophrenics and controls detected and repaired these errors equally frequently. This at first sight seems again to contradict Frith’s model. It is possible, however, that even though control subjects and schizophrenic subjects with different symptomalogies detected speech errors equally frequently, they used different monitoring systems in error detection. Errors can be detected on the basis of internal phonetic plan monitoring or through monitoring of acoustic information (Levelt, 1989). The present paper addresses this possibility. We consider whether there is a problem with the monitoring of internal speech plans specific to schizophrenics with verbal
hallucinations, but which is not found in controls or in schizophrenics with negative symptoms. To make this possibility more precise we need to consider some available models of language production.

**Models of monitoring and editing in speech production**

There are several views of monitoring in language production. Most models are modular, and that of Levelt (1989) includes a conceptualizer, formulator and speech comprehension system. The conceptualizer constructs ‘preverbal messages’, ‘conceives intentions, selects relevant information, orders it for expression’ and produces representations of the message, acceptably by the language planning agent. The formulator transforms them into phonetic plans.

Monitoring and editing of message plans may take place at different stages of utterance production, but at present there is no agreement on the precise cognitive architecture of the speech processing system (compare for example Laver, 1980; Butterworth, 1981; Motley et al. 1982; Levelt, 1989). Levelt (1989) allows three loci of monitoring, two internal and one external: monitoring and editing of the output of the conceptualizer; pre-articulatory editing of phonetic plans; and post-articulatory editing, which monitors the actual acoustic output.

Levelt argues that there is no dedicated language monitoring system for checking the output of the formulator. Pre- and post-articulatory monitoring is done by the agent’s comprehension system. In other models of speech production, as for example that of Motley et al. (1982), there is a dedicated module for the monitoring of the ‘formulator’.

The output of the conceptualizer has a dedicated monitor, which is distinct from the speech comprehension system (Levelt, 1989, p. 9). Levelt (1989) however, does not specify this monitor and how this differs from the comprehension system. This is unfortunate, since the monitoring of conceptualizer output is clearly relevant to Frith’s (1992) model. The presence of one or many monitors is clearly relevant to understanding of the problems of self-monitoring in schizophrenia. It is at present, however, difficult to decide between the different views, except on weak grounds such as parsimony.

Levelt (1989) also describes several rules governing the structure of repair as showing in the following examples. According to him speech is interrupted as soon as an error is detected, that is within the error word as in (1), or immediately following it as in (2), rather than later, as in (3). The transcription conventions are given in the Appendix.

1. **S41/12**
   *Interviewer*
   1. twelve [points to white and red circle]
   *Subject*
   2. Twelve
   3. white {sk 0·145}
   4. 0·13 white circle
   5. 1·09 and red circle
   (Here {sk} on the line 3. is the initial segment of ‘square’; transcribed quasi-phonetically. It is 0·145 s in duration)

2. **S41/3**
   *Interviewer*
   1. three [points to blue circle]
   *Subject*
   2. (Black 0·270)
   3. 0·17 blue circle

3. **S058/10**
   *Experimenter*
   1. [points to large brown and yellow squares]
   *Subject*
   2. She’s got both her hands on a large brown circle 0·490
   3. and a large yellow sk
   4. 0·39 large brown square
   5. 0·37 and a large 1·12 yellow 0·47 square

The detections in example (1) are particularly interesting. The duration of the pronounced fragments of error words is typically between 100 to 300 ms. Assuming a 250 ms reaction time, we can reason that the speaker has actually detected the error between 150 ms before the beginning of the articulation and 50 ms into it. This means that there was no, or not enough acoustic information available when the error was detected. According to Levelt (1989) the detection must have depended on the internal monitoring of a phonetic plan. Of course, normally, the internal/external monitoring feedback systems are coordinated and both involved in speech monitoring (cf. Lackner & Tuller,
1979). Nevertheless, one can use the speech flow interruptions as in example (1) to be indications that an error has been detected internally.

The aim of the present study is to investigate self-repair in controls, and schizophrenics with negative symptoms and verbal hallucinations, using the within error word interruptions as indicators of internal monitoring of phonetic plans. The prediction is that the repairs with speech flow interruptions within the error words, as in example (1), will be less frequent in subjects with verbal hallucinations than in those with negative symptoms and in controls, because of the problems with the monitoring of phonetic plans.

**METHOD**

The results reported in the present study were obtained from a sub-population of subjects who participated in a larger longitudinal study of the relationship between language, discourse and symptoms of mental illness. Some different results of the project have been previously reported in Montague *et al.* (1989), Leudar *et al.* (1992) and Thomas *et al.* (1993).

**Subjects**

**Clinical subjects**

There were altogether 30 subjects diagnosed as schizophrenics and 17 controls. The subjects with schizophrenia were drawn from admissions to two hospitals in Manchester. All subjects met the following inclusion criteria:

- first language English;
- age 16–50 years inclusive;
- not suffering from an organic psychosyndrome;
- not fulfilling Research Diagnostic Criteria (RDC) (Spitzer *et al.* 1975) for depressive psychosis;
- no history of alcohol misuse.

The subjects with schizophrenia were within 2 years of first onset of psychosis.

Initial clinical assessments were made within 72 h of admission and included the Present State Examination (PSE) (Wing *et al.* 1974), the Thought Language and Communication disorder (Andreasen, 1979) and the Schedule for the Assessment of Negative Symptoms (SANS) (Andreasen, 1981). An RDC diagnosis was derived for each clinical subject, based on the PSE ratings. Only those subjects were included who met RDC criteria for ‘definite’ or ‘probable’ schizophrenia. Clinical subjects were placed into hallucinating or non-hallucinating categories, based on the sum of ratings of PSE items 60, 62, 63, 64 and 65. The SANS summary score (the sum of scores for five global ratings) was used to establish the patients' negative symptom status, as described by Montague *et al.* (1989) on almost the same patient sample. A severity of illness score was derived by summing the PSE ratings for items 55, 56, 58, 59, and 66–99. This was used to compare the severity of illness between the patient groups.

**Controls**

Seventeen psychiatrically healthy controls were selected from orthopaedic admissions to Manchester Royal Infirmary. They were chosen to match psychiatric subjects as closely as possible for sex, age, educational attainment, years of full-time education and parental social class. All were screened by the project psychiatrist (P. T.) for psychosis, and although nobody was excluded on account of this, two were excluded because of heavy alcohol consumption. Orthopaedic admissions were selected in an attempt to control for the possible effects that contextual factors such as being a hospital patient might have on tasks involving communication.

**Design**

The subjects with schizophrenia were classified into the following three groups according to their hallucinations and negative symptoms status (see Fig. 1).

*Group 1* contained 11 subjects with high verbal hallucinations and minimal negative symptoms.

*Group 2* contained 9 schizophrenics with high negative symptoms and no verbal hallucinations.

*Group 3* contained 11 subjects with both verbal hallucinations and negative symptoms.

Together with the controls (Group 0), there were thus four independent groups.

**Materials**

**Reporter Test**

The reporter test (De Renzi & Vignolo, 1962) uses a referential communication task. Subject and experimenter sit at a table, with a display of
coloured tokens between them. The experimenter carries out simple actions (e.g. she touches a token or moves it). The subject is asked to imagine that she is on a telephone to somebody who has the same display, and to give them instructions so that they can duplicate the action. In our version of the reporter test there were 36 items.

**Reverse Digit Span Test (RDS1)**

In the reverse digit span test, the experimenter speaks sequences of digits to subjects who then repeat the sequence of digits backwards. In our test, we used a minimum sequence of two digits, and a maximum one of five. For each length of a digit sequence there were six test items. The total number of completely correct repetitions was taken as a score. The minimum possible score was thus 0, the maximum 24.

**Shape Cancelling Test (SCT)**

In the shape cancelling test, subjects were presented with an A4 sheet with 891 randomly printed non-alphabetic and non-numeric characters, there were 33 columns and 27 rows of the following characters @ < > & % ! $ ? | *.

The task was to cancel all the right or left pointing arrowheads, of which there were 102. The time taken to do this task was measured and the numbers of correctly cancelled, omitted and incorrectly cancelled characters were counted. The following three indices were calculated: the relative frequency of characters incorrectly cancelled ('errors'); the frequency of characters missed ('omissions'), (these two indices are assumed to be measures of attention); and 'work rate', which was calculated as time in ms per character inspected (i.e. the total time divided by 891).

The purpose of taking these working memory and attention indices was to match the groups on at least some aspects of information processing, as well as to determine the information processing co-variants of discourse disturbances.

**Procedure**

The three tasks were presented to each subject in a pre-determined random order together with three other tasks not relevant here. Each session lasted for up to 45 min, with a break between tasks 2 and 3, filled with informal conversation. The sessions took place in hospitals, usually in quiet rooms, with only the tester and the subject present.

The testers were in most cases voluntary research assistants (all third year psychology students), who were naive with regard to our experimental hypotheses. In a minority of cases (when the voluntary research assistants were not available), the tester was either the project secretary or the research associate. None of the testers had psychiatric training and they were naive to the subjects' diagnoses. Psychiatically naive testers were used to avoid the experimenter bias effect. They were, however, all given thorough training by the authors in administering the tasks used in the study.

**Transcription of speech samples**

The audio-recordings of each reporter test session were transcribed by the project secretary and the accuracy of transcripts checked by the first and the third authors. The transcriptions followed the conversation analytical system (CA) (cf. Sacks et al. 1974; Schegloff et al. 1977). Each tape was digitized using Audiomedia on Macintosh IIci. The duration of hesitation pauses and the durations of relevant speech segments of speech were measured, and included in transcripts as is indicated in the examples below.

**Locus of speech flow interruptions**

The speech flow interruptions were classified as follows.

*Inside the error word, as in (1) above and (4)*

(4) (S16/27)

**Interviewer**

1. [touches large red square]

**Subject**

2. Touch the large red {se 0:191}

3. 0:00 eh square

(In this reporter test item no. 27 the interviewer touched the large red square. Subject 16, however, started saying 'circle' in error ( {se} above is quasi-phonetically transcribed initial segment of 'circle', and it is 191 ms long.) Subject stopped within the error word 'circle', and repaired the error by saying 'square'. This repair was marked by a pragmatic particle 'eh' and the correct noun 'square' was stressed.)

**Immediately after the incorrect word, as in (2) above and (5)**

(5) S077/19
Interviewer

1. and this is number nineteen [moves the blue square]
2. ehm 1·58 put the (black 0·412)
3. 0·27 the blue square behind 0·22
(Here subject 77 erroneously identified the square being moved as ‘black’ (line 2) when in fact it was blue. She actually pronounced the whole error word, and interrupted the speech flow just after it. The error word ‘black’ was 412 ms long).

Later in the speech flow, as in (3) above, (6) and (7)

(6) (S611/17)

Interviewer

1. seventeen [moves the large red circle onto the large blue square]
Subject
2. the large red circle
3. on top of a large red 0·30 skv
4. 0·11 eh blue 0·22 square
(In this case, subject erroneously identified the square being moved as ‘red’ (line 3). He pronounced all of this error word, carried on speaking and only interrupted the speech flow in the middle of the next word ‘square’, which was actually referentially correct. {skv}, the pronounced initial segment of ‘square’ is an overshoot.)

(7) S132/10

Interviewer

1. this is number 10 [points to large black square and a large yellow circle]
Subject
2. you pointed to the large 0·55 (yellow 0·432)
3. and a yellow
4. 0·00 large black square
5. and a yellow circle
(Here subject 132 erroneously identified the square being moved as ‘yellow’. He pronounced all of this error word, carried on speaking, and only interrupted the speech flow in the next clause. Speech segment ‘square, and a yellow’, is the overshoot in this case.)

For each subject the following proportions were calculated, using as the base the number of repairs of errors attempted by a subject: the proportion of interruptions within error words; the proportion of interruption immediately after error word; and the proportion of speech flow interruption later in the report.

The absolute frequency of speech errors – and hence of their repairs – is typically higher in schizophrenics than in controls (cf. Leudar et al. 1992). Proportions of repairs of a particular type, rather than their absolute frequencies were used in our analyses in order to correct for this. The average number of error repairs used as the base was, however, not different across the four groups, $F(3, 44) = 1.85$, NS (The respective means for the four groups were 5.0 (0.5), 6.0 (1.1), 6.7 (1.3) and 8.0 (1.5), the figures in brackets being the standard errors.)

Duration of speech segments

The following two average speech durations were calculated: the average duration of error word fragments, such as e.g. {se} in (4); and the average duration of the error words, which were followed by immediate speech flow interruptions, such as ‘black’ in (5).

Analysis of pauses

For each subject the following average hesitation pauses were also calculated: hesitations preceding the error fragments, error words or error words with overshoots; and hesitations separating the reparandum and the repair.

Scoring reliability

Each reporter test transcript was scored for: ‘error presence’ (did a reporter item contain a
The results were analysed using analysis of variance (or co-variance where appropriate) with the following planned comparisons.

According to diagnosis (schizophrenia diagnosed or not, i.e. Groups 1, 2 and 3 v. Group 0).

According to presence/absence of verbal hallucinations (Groups 1 and 3 v. Groups 2 and 0).

According to presence/absence of negative symptoms (Groups 2, 3 v. Groups 1 and 0).

According to schizophrenia subtype (Group 1 v. Group 2).

The three schizophrenic groups were compared for overall severity of symptoms. The difference was not statistically significant, \(F(2, 28) = 2.064, \text{NS}\). Fig. 1 shows that the severity was slightly lower in the clinical group of subjects with both verbal hallucinations and negative symptoms. There was, however, no difference in severity between the groups with

**RESULTS AND DISCUSSION**

The results were analysed using analysis of variance (or co-variance where appropriate) with the following planned comparisons.

According to presence/absence of verbal hallucinations (Groups 1 and 3 v. Groups 2 and 0).

According to presence/absence of negative symptoms (Groups 2, 3 v. Groups 1 and 0).

According to schizophrenia subtype (Group 1 v. Group 2).

The three schizophrenic groups were compared for overall severity of symptoms. The difference was not statistically significant, \(F(2, 28) = 2.064, \text{NS}\). Fig. 1 shows that the severity was slightly lower in the clinical group of subjects with both verbal hallucinations and negative symptoms. There was, however, no difference in severity between the groups with only the verbal hallucinations and only the negative symptoms. The severity was hence used as a co-variate, but it had no significant effect in any analysis.

All the four groups were compared for distribution of sex, age, education achievement, age on leaving education and social class (see Table 1). None of the differences on these measures were statistically significant, with the exception of social class. This was marginally significant \(F(3, 41) = 2.225, 0.05 < P < 0.10\). Table 1 shows that the social class of clinical groups was, on average, slightly lower than that of controls. The social class was used as a co-variate, but again its effect was not significant in any analysis.

### Table 1. Demographic matching of the experimental groups (means and standard errors)

<table>
<thead>
<tr>
<th>Group</th>
<th>Sex (% men)</th>
<th>Age (years)</th>
<th>Qualification (no. of GCSE)</th>
<th>School leaving age (years)</th>
<th>Social class</th>
</tr>
</thead>
<tbody>
<tr>
<td>Controls</td>
<td>50 (22)</td>
<td>27.7 (0.4)</td>
<td>3.3 (0.5)</td>
<td>17.9 (0.5)</td>
<td>3.8 (0.4)</td>
</tr>
<tr>
<td>Schizophrenics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>With verbal</td>
<td>58 (27)</td>
<td>26.4 (0.5)</td>
<td>2.4 (0.5)</td>
<td>16.3 (0.5)</td>
<td>4.1 (0.5)</td>
</tr>
<tr>
<td>hallucinations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>With negative</td>
<td>58 (29)</td>
<td>26.0 (0.6)</td>
<td>2.5 (0.7)</td>
<td>16.7 (0.7)</td>
<td>4.9 (0.4)</td>
</tr>
<tr>
<td>symptoms</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>With verbal</td>
<td>55 (27)</td>
<td>24.3 (0.6)</td>
<td>2.4 (0.7)</td>
<td>17.3 (0.7)</td>
<td>5.3 (0.4)</td>
</tr>
<tr>
<td>hallucinations and</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>negative symptoms</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 2. Group matching on information processing measures (shape cancelling: \(\Box\), errors; \(\Box\), omissions).
Reverse Digit Span Test (RDST)

Fig. 2a indicates that the controls did better on this test than the schizophrenics and that there were no differences between the schizophrenic groups. The overall difference between the four groups was however, not significant, \( F(3,44) = 1.897 \).

Shape Cancelling Test (SCT)

The relative frequencies of errors and omissions on this test were calculated as proportions of possible errors and omissions. They are given in Fig. 2b. It can be seen that controls made fewer errors than the schizophrenics, but the overall difference was not statistically significant, \( F(3,44) = 1.99, \) NS. The rate of omissions was again lower in controls but the overall difference was not significant, \( F(3,44) = 1.54, \) NS. These variables (i.e. RDST scores and the SCT omissions) were subsequently used as co-variates in the statistical analyses of repair.

Position of speech flow interruptions in error repair

Immediate, within word repair

These were the interruptions such as (1) and (4) above. The durations of the error word fragments (e.g. 'se' in the example (4) above) are shown in Fig. 3. On average, for all the groups combined, the duration is 257 ms. There were no significant differences between the groups, \( F(3,32) = 1.023, \) NS. (Some subjects were not included in this analysis, because if a subject produced no within error word repairs, then obviously the average durations of error word fragments was not available.) The average duration of these error word fragments is such that we can assume that their detection cannot have used acoustic information but must have been internal, presumably based on monitoring of phonetic plans.

Fig. 4 shows that the within-word interruptions of errors were more frequent in the control group than in the three schizophrenic groups. The difference was significant overall, \( F(3,44) = 5.463, \) \( P < 0.003 \). The planned comparison shows that the difference between the controls and the three schizophrenic groups was significant \( T(44) = 4.04, \) \( P < 0.001, \) but none of the other comparisons were. (This significant difference between controls and the three schizophrenic groups for proportion data was also found when the data for absolute number of within error repairs was analysed.) An additional analysis of co-variance was conducted, with the two attention and one working memory variables used as co-variates. None of the co-variates was significantly correlated to the frequency of the within-error-word interruptions of speech flow. This indicates that the difference between controls and schizophrenics is not caused by working memory or attention deficits.

Repair immediately after the error words

These were the interruptions such as (2) and (5) above. The duration of the error words (e.g. 'blue' in both (2) and (5)) was on average 400 ms. There was no significant difference between the four groups, \( F(3,38) = 0.321, \) NS. The detection of the error words with the average duration of 400 ms could have used acoustic information, unlike in the case of error word fragments, which were on average 257 ms in duration.

There was a significant difference between the four groups in the relative frequency of these error repairs, \( F(3,44) = 4.111, \) \( P < 0.025 \). Fig. 4 shows that these speech flow interruptions were more frequent in the schizophrenic groups, \( T(44) = 2.932, \) \( P < 0.005 \). The difference between the Groups 2 and 3, that is those with only verbal hallucinations or only negative symptoms, was, however, not significant. The difference between groups was again independent of the attention and working memory variables.
If the cumulative frequency of repair locus is examined, on average 75% of speech flow interruptions occurred by the end of the incorrect word. There was no significant difference between the four groups in this respect, $F(3, 44) = 1.654$, NS, and as would be expected there is no significant difference in the relative frequency of repairs with overshoots in the four groups, $F(3, 44) = 1.654$, NS. This result is consistent with a rule ‘interrupt speech flow as soon as you detect an error’ proposed by Levelt (1989).

According to Levelt (1989) in the case of the within-error-word breaks, the speaker cannot depend on acoustic information, and must have detected the error on the basis of monitoring an internal phonetic plan. So our results support Frith’s (1992) model, in that schizophrenics with verbal hallucinations do indeed produce fewer speech repairs indicative of internal monitoring (of phonetic plans) than the controls but have no problem with using external feedback. More precisely, our results imply the internal monitoring of phonetic plans is not altogether absent, but is less effective in schizophrenics than in controls — they seem to detect only a half of these errors internally. The main difficulty for Frith’s model is however, that the problem with the efficiency of internal monitoring does not seem to be unique to schizophrenics who hear voices, but was also obtained for schizophrenics with negative symptoms, who do not. This is in fact the main finding of this study.

Before using our results to suggest changes to Frith’s model, it is prudent to examine two other possible explanations of the difference observed between controls and subjects with schizophrenia in the frequency of error word interruptions within, and after these words.

The first explanation is that the subjects in one or more schizophrenic groups found it difficult to interrupt speech within a word. There is a datum in our experiment relevant to this possibility. Some repairs, such as (6) and (7) above involved the interruptions of speech flow within words other than the error words. We calculated their frequency relative to the total number of repairs and compared them for the four groups. There was an overall significant difference between the four groups, $F(3, 44) = 4.930, P < 0.005$. Fig. 5 shows that these interruptions were more frequent in schizophrenic subjects, $T(44) = 2.345, P < 0.05$, and especially those with verbal hallucinations, $T(44) = 2.904, P < 0.01$. So, it seems unlikely that the relatively low frequency of fast within the error-word repairs was due to problems with stopping within words in general.

The second alternative explanation is as follows. The presence of the fast, within error speech flow interruptions does indicate internal monitoring of phonetic plans, but their absence does not necessarily indicate the lack of it. This is so for the following reason. The speaker may detect an error internally but be relatively slow to react, the result being that a larger proportion of the word, or in fact the whole of the word is...
pronounced, rather than just a fragment. It is therefore possible, that in some individuals the relatively low proportion of speech flow interruptions within the error words reflects their slow reaction times, rather than problems with internal monitoring. This could be so for all the schizophrenics or just for one of the subgroups.

One aspect of the definition of negative schizophrenia is the slow reaction to speech stimuli. Frith & Done (1988), however, did not find a difference between negative and positive schizophrenics and controls in auditory RTs. There is some evidence that chronic schizophrenics have relatively slow reaction times. Our clinical subjects, however, were not chronic. What is needed to test this possibility is a temporal co-variante which would provide an estimate of reaction time. There are three types of relevant data which may provide such an estimate.

1. The duration of pauses separating errors and their repair  Fig. 6 compares hesitation pauses (cf. Goldman-Eisler, 1968), separating error words or error word fragments on the one hand and their repairs on the other. These pauses are twice as long when the whole error word has been articulated. The pause seems to mark the onset of repair for the listener, as Howel & Young (1991) argue. There were however no significant differences in this pattern of pause duration between the four groups.

2. The duration of pauses preceding the errors  For the pauses immediately preceding the errors, the results were different depending on whether the whole word or only a fragment of it was articulated. In the former case, the average duration of pauses was 138 ms and there were no significant differences between the groups, $F(3,40) = 1.45$. The duration of pauses preceding partially articulated error words was 93 ms. The overall difference was just not significant, $F(3,31) = 2.40$, $P < 0.08$. These pauses were shorter in schizophrenics with verbal hallucinations than in controls and schizophrenics without, $T(31) = 2.47$, $P < 0.02$.

3. The work-rate index  The work-rate index expresses the time in milliseconds per character inspected in the shape cancelling test. This can be used as an estimate of choice-reaction time. For all the groups combined, the work-rate was 255 ms. The means for the four groups are given in Fig. 7. The difference between the four groups was not statistically significant, $F(3,44) = 1.506$, NS. We have also correlated the work-rate index to the frequency of speech interruptions in the error words. If this frequency is in general affected by a slow reaction to internal error detections, a significant negative correlation would be expected. The correlation was negative but not significant ($r = -0.08$). Finally, we have introduced the work-rate index as a co-variante in the analysis of variance of the frequency of speech interruptions in the error words, as we did working memory and attention indices previously. The analysis revealed that the work-rate index was not a significant co-variante.

So the analysis of speech pauses and the work-rate index do not support the possibility that the relatively low frequency of interruptions of speech flow within error words in schizophrenic groups is due to a non-specific slow reaction to internal detection of errors in phonetic plans. It is of course the case, that these variables were only estimates of the speed of reaction to internal error detection, and the hypothesis should be further investigated in subsequent studies.

As one of the anonymous referees of this paper pointed out, it is possible, that the difference in the locus of speech monitoring between schizophrenics and controls observed
in our experiment is not actually a difference between schizophrenics and controls but due to some common non-criterial factor, such as systematic difference in attention, reaction time, working memory or maybe some other unspecified variable. We have in fact used indices of working memory, attention, and estimates of choice reaction time as co-variates to control for this possibility. We have found that indeed schizophrenics did worse on working memory and attention tests, but these and the choice reaction time estimates did not co-vary with the measure of locus of speech monitoring. It is of course possible that there is another non-criterial variable which would differentiate our orthopaedic controls and schizophrenics, and be at the same time correlated to the locus of speech monitoring. This possibility should be addressed by further research, possibly by using an additional relevant control group, such as manic depressive patients.

The main findings of the present study are as follows: (a) in all the subjects diagnosed as schizophrenic, irrespective of their negative symptoms and verbal hallucinations status there seems to be a deficit with internal monitoring of internal phonetic plans; and (b) this deficit is independent of working memory, attention variables and choice reaction time.

**GENERAL DISCUSSION**

Leudar et al. (1992) reported that schizophrenics with verbal hallucinations, those with negative symptoms and the controls failed to detect errors in speech equally frequently. The findings of the present study, however, imply that the schizophrenics depend less than controls on monitoring of internal phonetic plans and more on external, acoustic feedback. This supports Frith's (1992) claim that in schizophrenia, there is a problem with internal self-monitoring of action plans. Our data however, implies that the problem is common to all the schizophrenics irrespective of their verbal hallucinations and negative symptoms status. This of course cannot be taken to disprove Frith's (1992) explanation of verbal hallucinations in terms of a specific problem with internal monitoring. As we have seen above, the planning and production of utterances is monitored and edited at several stages of the process. Our data are only pertinent to the perceptual monitoring at pre- and post-articulatory levels. Levelt (1989) suggests that the output of the conceptualizer is also monitored. It may be at this level of speech production monitoring, that there is a deficit specific to schizophrenics with verbal hallucinations, and future investigations may compare the relative efficacy of the monitoring in individuals who do and do not hear voices.

A further feature of our results is that the monitoring of phonetic plans is not altogether absent in schizophrenia, but seems to be less efficient. This raises an interesting possibility in that the problem may not just involve specific aspects of intention or plan monitoring, but also the coordination of the different types of feedback information. Unfortunately, most of the research available focuses on isolating the different feedback systems, rather than on their coordination or on what happens when they are out of synchrony. But as the phenomenon of delayed auditory feedback indicates, it is precisely their coordination in action which is important (see MacKay, 1987 for a review). There does not seem to be at present any work available on the effects of delayed auditory feedback on speech of subjects with schizophrenia.

Further research could thus investigate the efficacy of the monitoring of the conceptualizer output in subjects with verbal hallucinations, and address the problem of coordinating different feedback information, when one or more feedback loops are less than efficient.

There is one residual problem we would like
to address. Searle (1983) distinguishes intention-prior-to-action from intention-in-action. Intentions-prior-to-actions clearly have a place in Frith's model as they do in most cognitive models of planned action. It is not clear, however, that intentions-in-action do. Simplifying Searle's (1983) account, he identifies intention-in-action with the experience of acting. According to him, intentions-in-action are caused by intentions-prior-to-action and they in turn cause behaviour events, of which they are presentations. He argues persuasively that all actions have intentions-in-action, but not necessarily intentions-prior-to-action. Imagine that while I am writing this paper, I remember that I was supposed to make a phonecall before one o'clock. I reach for the phone and make the call. This action has both an intention-prior-to-action – I decided to make the phonecall before I reached for the receiver – and an intention-in-actions of moving my arm slowly to the phone. This was an intentional, conscious activity, not a reflex. Now imagine that the telephone rings, and I reach for it and answer. In this case, reaching for the telephone is again an intentional action. It is, however arguably not caused by an intention-prior-to-action but by the bell. (I could of course have decided not to answer, but to answer is a default, at least for myself.) Finally, I could have just got up and paced awhile thinking. All three actions have intentions-in-action, but only the first one has a prior intention. It seems to us, therefore, that intentions-in-action are not the same as stimulus intentions (cf. Frith, 1992). Further, the absence of prior intentions does not guarantee that an action is without awareness, the absence of intention-in-action does. We suggest that verbal hallucinations must, and delusions of control may, involve stripping intentions-in-actions from intentional states involved in acting and speaking. We suspect that our results are relevant to this process.

APPENDIX

Transcription conventions used in all the examples of speech are as follows.

The headings identify subjects by number and reporter test items.
Speaker's identity is reported in italics.
Lines are numbered for reference.

The numbers in the speech examples are hesitation pauses in seconds.
The indentations in the examples indicates repairs.
Relevant non-verbal actions are reported in square brackets.
Speech errors are enclosed in curly brackets, with their durations following. For example, ‘circle’ in example (3), line 2 is 0·49 s in duration, ‘Sk’ in example (1), line 3, is a beginning of ‘square’ and it is 0·145 s in duration.
Words fragments are transcribed quasi-phonetically (e.g. ‘skv’ is an initial segment of ‘square’, ‘ce’ is an initial segment of ‘circle’).
Stressed speech is underlined,: refers to lengthening of the preceding sound.
Segments such as ‘and a large yellow sk’ in the example (3), line 3 are referred to below as ‘overshoots’.

We owe thanks to Chris Frith, Graham Hitch, Peter Meudell and Denis McKeown for useful comments on an earlier draft of this paper. We also wish to express our gratitude to the voluntary research assistants who helped to collect the data and Jean Picker who made the research practically possible. Finally, we wish to thank the following psychiatrists, who provided psychiatric assessment: Dawn Black, Linda Montague, David Newby and David Row.

REFERENCES


