Apraxia of speech: how reliable are speech and language therapists' diagnoses?

Katharyn Mumby Central Cheshire PCT, now at School of Education, Communication and Language Sciences, University of Newcastle upon Tyne, **Audrey Bowen** School of Psychological Sciences, University of Manchester and **Anne Hesketh** School of Psychological Sciences, University of Manchester, UK

Received 23rd June 2006; returned for revisions 8th November 2006; revised manuscript accepted 15th December 2006.

Objective: To discover how reliably speech and language therapists could diagnose apraxia of speech using their clinical judgement, by measuring whether they were consistent (intra-rater reliability), and whether their diagnoses agreed (inter-rater reliability).

Design: Video clips of people with communication difficulties following stroke were rated by four speech and language therapists who were given no definition of apraxia of speech, no training, and no opportunity for conferring.

Settings: Videos were made of people following stroke in their homes. Ratings of the videos were carried out in the university lab under controlled conditions. **Subjects**: Forty-two people with communication difficulties such as aphasia, apraxia of speech and dysarthria took part, and four specialist speech and language

therapists acted as raters.

Main measure: Speech and language therapists' ratings of the presence and severity of apraxia of speech using videos.

Results: Intra-rater reliability was high for diagnosing (1) the presence of apraxia of speech (Cohen's kappas ranging from 0.90 to 1.00; 0.93 overall), and (2) the severity of apraxia of speech (kappa 0.84 to 0.92; 0.90 overall). The inter-rater reliability was also high for both the presence of apraxia of speech (kappa 0.86) and severity of apraxia of speech (0.74).

Conclusion: Despite controversy over its nature and existence, specialist speech and language therapists show high levels of agreement on the diagnosis of apraxia of speech using their clinical judgement.

Introduction

Apraxia of speech is a communication impairment that can occur following stroke, but there is considerable controversy in the literature about how to define and assess it in order to tailor rehabilitation.

disorder of motor programming,¹ but there has been heated debate over the last 30 years about its true nature.^{2,3} Currently there is no universally agreed definition of apraxia of speech or 'gold standard' for assessment. Definitions of apraxia of speech have been evolving within different theoretical frameworks: behavioural, cognitive and neuroanatomical⁴

and also within acoustic and phonetic-perceptual

frameworks.⁵ However, it is unclear to what extent

In apraxia of speech, speech sounds are made in an effortful or erratic way in the absence of muscle weakness. Traditionally this has been thought of as a

Address for correspondence: Anne Hesketh, HCD, School of Psychological Sciences, Humanities Devas Street Building, University of Manchester, Oxford Road, Manchester M13 9PL, UK. e-mail: anne.hesketh@manchester.ac.uk

each framework is useful in differential diagnosis with other communication impairments such as aphasia and dysarthria.

Other recent major reviews⁶⁻⁸ have highlighted how difficult it is to interpret the existing studies in the field because of the failure to describe subject selection adequately. Apraxia of speech is known to co-occur with aphasia, which may also result in sound errors in speech, aphasic 'phonemic paraphasias', involving the selection and production of the wrong phoneme or speech sound (e.g. 't' in 'speat', when the target word was 'speak'). Their errors might be described as the wrong sound easily articulated. The sound errors characteristic of apraxia of speech on the other hand tend to involve articulatory struggle, mistiming and distortion of sounds, which may lead to a different sound being produced, for example the word 'speech' pronounced as: 'stree stree skee skeech'. (Both these examples are taken from the same person (AS) within the current study.)

Both types of error involve faulty sound production, but for different reasons, and the two can be difficult to distinguish. The most widely used test for apraxia of speech, the Apraxia Battery for Adults (II)⁹ cannot be regarded as a gold standard.⁸ It considers phonemic paraphasias to be an apraxic symptom, although they are characteristic of aphasia. In other words, people presenting with phonemic paraphasias are included within the criteria for diagnosis of apraxia of speech, and this is acknowledged to be a shortcoming of the Apraxia Battery for Adults (II).

In the absence of a 'gold standard' for assessment, diagnosis tends to be made by clinical judgement, with reference to checklists of behaviours.^{7,8} For example, Wertz et al. 10 referred to:

- Effortful, trial and error, groping articulatory movements and attempts at self-correction.
- 2 Dysprosody unrelieved by extended periods of normal rhythm, stress, and intonation.
- 3 Articulatory inconsistency on repeated productions of the same utterance.
- Obvious difficulty initiating sentences.

The checklists, which have evolved in the light of recent research, still have to be interpreted through clinical judgement, so it is crucial to measure the reliability of this judgement. This is particularly

important given that some in the field have questioned the existence of apraxia of speech (see discussion of definitions by Croot⁴). The controversy may stem from the fact that aphasia may occur without apraxia of speech but it is rare for apraxia of speech to occur in the absence of aphasia, 11 reflecting patterns of lesion in stroke. Many of the studies in the field fail to describe the way in which aphasia and apraxia of speech coexist in the participants selected, making it difficult to draw conclusions. Measuring whether apraxia of speech can be reliably identified is the first step in addressing these problems.

The present study addressed the issue of whether clinical judgement about the diagnosis of apraxia of speech can be reliable without imposing a definition of apraxia of speech on clinicians. The study investigated whether speech and language therapists specializing in acquired neurological disorders could consistently diagnose apraxia of speech (intra-rater reliability) without additional training or conferring, and whether their diagnoses agreed (inter-rater reliability). The objectives were to determine the intra-rater reliability for each of four therapist raters about the presence and severity of apraxia of speech in people with communication problems following stroke, and to measure the inter-rater reliability amongst the four therapist raters. This information is important in providing a basis for the process of developing a 'gold standard' for assessment, for understanding the nature of apraxia of speech further, and ultimately for developing and evaluating rehabilitation interventions.

Methods

Participants with communication difficulties were recruited from three UK hospitals. They were referred to the study by their speech and language therapists on the basis that they had communication problems after stroke, potentially including aphasia, dysarthria and apraxia of speech. Participants were at least four months post onset (so that they could cope with being interviewed about their stroke). They were also selected on the basis that, in the opinion of the referring therapist, they could converse enough for an interview, and would agree to two home visits for assessment and videoing.

The 43 referrals were all appropriate, but one was excluded from the final sample due to the poor quality of the video. The participants were visited at home on two occasions by the research speech and language therapist (KM) to gain their consent for the process, and to complete an assessment of their communication including a semi-structured interview which was videoed. The videos were used to prepare 42 video clips which would later be rated by four therapists.

The assessment was designed to ensure as far as possible that the sample contained a balance of people with and without apraxia of speech, and that the latter had other communication problems. The researcher characterized the nature of the participants' communication problems in terms of apraxia of speech, aphasia and dysarthria. Categorization was based on observations during the semi-structured interviews, and observations of behaviours elicited by several widely used clinical assessments. Participants were asked to produce a range of non-speech oral and laryngeal movements taken from the Frenchay Dysarthria Assessment¹² to help detect any muscle weakness or abnormality which could be attributed to dysarthria. Selected subsections of the Boston Diagnostic Aphasia Examination¹³ were also used to give an indication of the presence and severity of aphasia. These subtests (Commands, Complex ideational material, Responsive naming, and Grammatical rating) were deliberately non-contentious in terms of the debate over apraxia of speech and phonemic paraphasias, and are reported to be particularly sensitive to the presence of aphasia. 13 Some of the test materials from the Apraxia Battery for Adults (II)⁹ were also used to elicit speech.

Criteria based on the checklists of Wertz et al. 10 and McNeil et al. 7 for the purpose of detecting apraxia of speech, were interpreted by the researcher (KM) who had 20 years of experience as a speech and language therapist. Importantly, the checklists were not used to exclude people with aphasia from the sample, because it is common for aphasia and apraxia of speech to cooccur. The intention was to identify those with apraxia of speech and those who showed no evidence of apraxia of speech in the opinion of the researcher. This identification was not as a standard for comparison with the therapist raters, but to try to ensure that the sample of people with communication problems would include a balance of people with and without apraxia.

The researcher interviewed each participant at home about their stroke and communication, using a series of general questions to encourage them to talk freely. The videos lasted between 20–30 minutes depending on how talkative people were, and their rate of speaking. These videos were too long to be viewed in their entirety by raters. Therefore clips were selected by the researcher according to prescribed criteria, which ensured that each clip was 2–3 minutes in length, and included both spontaneous speech and repetition of polysyllabic words (thought to be a good task for highlighting apraxia of speech⁶). Clips were also selected to minimize the number of edits.

Four speech and language therapists specializing in adult neurological disorders from three different hospitals in the UK participated as raters. A therapist from one other hospital declined to participate, so two of the four therapists worked within the same locality. The departments were selected on the basis of geographical proximity to the researcher's base. None of the therapists worked within the same team as the researcher. All four therapists had trained at different establishments, graduating between 7 and 19 years ago, mean (SD) 11.25 (5.4) years. They were all female, and aged 28–40 years, mean (SD) 34 (5.7) years, and were in current practice.

To prevent each therapist having to rate all 42 clips, an overlapping design of paired ratings was used (Table 1). This was an economical and effective use of therapists' time which maximised the number of participant video clips being rated within the study, and avoided therapists rating videos of people they knew. Using this design, every clip was viewed by a pair of therapists, and each therapist was compared with all three others. Each therapist rated 21 clips in two separate sessions under controlled conditions which avoided conferring.

The use of 21 clips reduced the likelihood of therapists remembering diagnoses from one session to the next, and a break with a distracter task was used between the sessions to minimize this further. The clips of people who were thought by the researcher to have apraxia of speech (see Results) were distributed equally across all the therapists, so that they were each allocated 12 people who potentially showed apraxia of speech. Therapists were not told how many people had apraxia of speech, and were not forewarned that they would be rating the same clips twice. They made their diagnoses on the basis of the clips alone, without any access to data from the other assessments.

Table 1 Overlapping allocation of 42 video clips, 21 clips per therapist rater

Clip code	e 21 clips allocated to each therapist rater			er
	Rater1	Rater2	Rater3	Rater4
TKNSSGCNLCSENT BELTSENTESDSSREYRJSGDSGGNRRTLCR	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21	1 2 3 4 5 6 7 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21	1 2 3 4 5 6 7 8 9 10 11 12 13 14	1 2 3 4 5 6 7 10 11 12 13 14 15 16 17 18 19 20 21

No training was given about the nature and definition of apraxia of speech, in order to measure how therapists approached the issues from their own expertise. They rated the severity of apraxia of speech by selecting one of the following terms: 'none', 'mild' or 'moderate/severe'. These categories were chosen because of the nature of the sample. People with particularly severe apraxia of speech were not included because they would not have been able to complete the interview.

The consistency of individual therapists (intra-rater reliability) and the agreement between the therapists (inter-rater reliability) were calculated using Cohen's kappa which takes into account the amount of agreement there would be by chance. Confidence intervals for the kappa scores have not been stated because such calculations are not reliable for this sample size. Where no confidence intervals exist, kappa scores should be regarded as estimates.

Results

Categorization of participants by the researcher

The sample of 42 people with communication problems included 15 women (36%) and 27 men (64%), and their ages ranged from 39 to 88 years, mean (SD) age 67 (11.3) years. The participants were between four months and just over eight years post stroke, mean (SD) time 30 (28.8) months, and their occupations or previous occupations ranged from manual workers to professional workers.

According to the researcher's observations, the sample of 42 participants contained a balance of those with and without apraxia of speech (19 people with no apraxia of speech (45%) and 23 people with apraxia of speech, of whom 15 had a mild and eight had a moderate or severe apraxia of speech) (Figure 1). The sample included people with a range of aphasic impairment and all those with apraxia of speech showed some evidence of aphasia (see Discussion). Some who showed no apraxia of speech still had

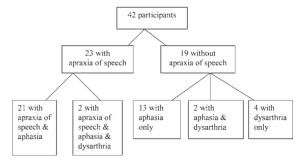


Figure 1 Categorization of communication impairments by researcher.

aphasic sound production errors in the form of phonemic paraphasias. In addition, the researcher judged that six of the 42 were predominantly dysarthric without apraxia of speech. Several of the participants appeared to have a combination of communication problems: for example aphasia and dysarthria; apraxia of speech, aphasia and dysarthria.

Findings from rating sessions

Therapists were extremely consistent in their diagnoses between rating sessions 1 and 2, giving almost perfect intra-rater reliability for the presence/absence of apraxia of speech (kappas between 0.9 and 1.00; kappa of 0.93 overall, Table 2). Intra-rater reliability was also almost perfect for the severity of apraxia of speech (kappas between 0.84 and 0.93), and expressed across all four therapists, the kappa score was 0.90. Each therapist was consistent within themselves in 20 or 21 of the clips (98–100%) about whether apraxia of speech was present, and about its severity in 19 or 20 of the clips (95–98%).

In terms of inter-rater reliability, when therapists' diagnoses were compared, the kappa scores demonstrated substantial agreement (kappa 0.86 for presence/absence and 0.74 for severity when all the severity ratings are combined, shown in Table 2) even though no definition of apraxia of speech or criteria for labelling severity had been given. In fact they were unanimous about the presence or absence of apraxia of speech in 38 (90%) of the clips, and unanimous about severity in 34 (81%) of the clips. The

measure of inter-rater reliability represents the most conservative estimate, because kappas were used from session one to avoid any possibility of the results being influenced by practice or conferring (however limited) during the break.

The researcher and the therapist raters were in agreement about the presence/absence of apraxia of speech in 37 of the 42 participants. Where there was lack of unanimity between the four therapists and the researcher, the discrepancies related to judgements about whether there was mild apraxia of speech or none at all (Table 3).

Discussion

The results gave a clear indication that the therapists were consistent within themselves and agreed with each other about both the diagnosis and the severity of apraxia of speech. The possibility that they remembered all their ratings from session 1 into session 2 was reduced by separating the sessions with a distractor task, by not forewarning the therapists that they would see the clips again, and by using a high number of clips (21 per therapist in each session). In fact, therapists commented after the sessions that it was very hard to remember what they had put from one session to the next.

Confidence intervals cannot be stated for the kappa scores, so the reliability measures in this study should be regarded as estimates, but because they are

Table 2	Results of			

Intra-rater reliability: Kappa scores within raters	Apraxia of speech presence/absence	Apraxia of speech severity
Therapist 1 Therapist 2 Therapist 3 Therapist 4 Overall (across four therapists) % consistently rated for each therapist	0.90 0.90 0.90 1.00 0.93 98–100%	0.92 0.92 0.84 0.92 0.90 95–98%
Inter-rater reliability: Kappa scores shown between therapists rating for apraxia of speech in session 1		
Apraxia of speech – presence/absence Apraxia of speech – severity ('none', 'mild', and 'moderate/severe' combined) % of total clips unanimously rated	0.86 0.74 81%	

Case	Therapist rating	Researcher	Scenario
DN	Mild	Not proven apraxia of speech	Severe fluent aphasia (jargon). Unable to complete repetition tasks. Receptive deficits make it difficult to assess
BJ	None	Mild	Mild anomic with some phonological deficits (shown on in-depth testing) and very mild apraxia of speech
BD	None	Mild	Mild anomic, tends to use circumlocution to avoid certain words. Subtle non-verbal behaviours due to rapid rate of speech delivery
DR	None or mild	Mild	Conduction aphasic, with additional apraxia of speech shown on in-depth testing
LG	Mild	Not apraxia of speech – may have some other form of apraxia	Presents like neurogenic dysfluency: some features in common with apraxia of speech, but not the same profile. Initiation problems and syllable repetitions predominate: fluent when whispering

Table 3 Cases involving discrepancies over diagnosis between therapists and researcher (using ratings from session 1)

consistently high it is likely that they reflect a good level of agreement. The therapists all worked in one area of the UK, and it could be argued that the findings do not represent specialist speech and language therapists in general, but they had all trained at different establishments, and were from three different hospitals, which increases the generalizability.

When interpreting the results, it is important to consider that the therapists were making their diagnosis on the basis of a short video clip rather than the more lengthy face-to-face assessment typical of clinical practice. However, the diagnoses of the therapists agreed very closely with those of the researcher, even though the therapists had access to less information in the 2–3 minute clips than the researcher who undertook two home visits and completed the interview and assessment. Limiting the length of each clip to 2–3 minutes could have made it less likely for therapists to detect very mild apraxia of speech compared with the researcher who carried out a longer assessment.

Categorization of the sample of participants was hampered by the lack of a gold standard for assessing apraxia of speech but, indeed, this is the issue at the heart of the study. The studies in the field rely for their diagnosis of apraxia of speech upon clinical judgement, usually from two therapists. 14-20 One study used consensus by discussion between three therapists.²¹ Most studies also refer to a series of checklists, dating from Kent and Rosenbek's four-point list

of characteristics of apraxia of speech22 through Wertz's version¹⁰ to the more comprehensive list from McNeil⁷ which has been adopted in some more recent studies.^{20,23} Both the checklists and 'clinical diagnosis' are subjective and depend on interpretation, but there has been no study until now to determine whether the therapist diagnoses are reliable. The findings of the current study are therefore crucial for underpinning further work.

The first edition of the Apraxia Battery for Adults (subsequently revised⁹) has been used in some studies in diagnosing apraxia of speech.^{24–27} However, the Apraxia Battery for Adults does not discriminate between aphasic (phonological) impairments and apraxia of speech, and its standardization is limited. Moreover, published tests for dysarthria and aphasia are not specific about differential diagnosis of these disorders from apraxia of speech. For these reasons it is likely that research will continue to rely on clinical judgements and detailed single-case expositions to clarify the nature of apraxia of speech and its possible treatment.

In the light of these difficulties, the categorization of the sample of 42 participants in this study has limitations, but it represents a good range of communication disorders, particularly as far as aphasia is concerned. There were fewer people with dysarthria, and none with very severe dysarthria because of the nature of the interview task, but the sample did include people with more than one communication disorder at a time (see Table 1). Therefore a good range of participants has been rated for the potential diagnosis of apraxia of speech. However it needs to be borne in mind that this study relates only to stroke, and different factors could come into play when diagnosing apraxia of speech occurring as a result of head injury or progressive neurological disease.

Identifying apraxia of speech as being distinct from certain types of aphasia is recognized to be a very thorny issue.⁸ All 23 people in this sample who were identified by the researcher as having apraxia of speech, and all 23 identified by the therapists also had some form of aphasia on assessment. There was a large range of severity of aphasia. Therapists were not just assigning a diagnosis of apraxia of speech to those with more pronounced aphasia, because when the group with apraxia of speech was compared with the group with no apraxia of speech, there was considerable overlap in the range of aphasia severity scores obtained from the Boston Diagnostic Aphasia Examination (percentile score for those with apraxia of speech: range 6-87; for those with no apraxia of speech: range 29–100). There was no direct relation between the identification of apraxia of speech and the severity of aphasia. The therapists also consistently diagnosed the absence of apraxia of speech in six people who produced aphasic sounds errors (phonemic paraphasias). Further investigation will be needed to determine how apraxia of speech and aphasia are interconnected. The study looked at fairly crude categorical judgements about apraxia of speech in order to establish a basis for further work. A further study has also been undertaken to explore actual instances of apraxic errors, whether therapists consistently identify such errors and where they occur.

In summary, the results indicated that specialist speech and language therapists' clinical judgement about the diagnosis of apraxia of speech is highly consistent. There was a high level of agreement amongst speech and language therapists diagnosing the presence and severity of apraxia of speech. Despite the theoretical debate about the existence and nature of apraxia of speech these results indicate that therapists are identifying a clinically meaningful population. These results form a foundation for further research into the nature of apraxia of speech, which is essential for tailoring rehabilitation for this extremely frustrating disorder of communication.

Acknowledgements

The authors wish to express their gratitude to Andy Vail, Biostatistics Group, University of Manchester, for statistical advice, and to all the participants and therapist raters for their contribution.

Clinical messages

- Speech and language therapists assessing the presence and severity of apraxia of speech following stroke using clinical judgement were highly reliable.
- Future research needs to evolve standardised assessments of apraxia of speech and compare these assessments with clinical judgement (shown here to be reliable).

References

- 1 Darley FL, Aronson EA, Brown JR. *Motor speech disorders*. Saunders, 1975: 304.
- 2 Miller N. Changing ideas in apraxia of speech. In Papathanasiou I ed. Acquired neurogenic communication disorders: a clinical perspective. Whurr, 2000: 173–202.
- 3 Code C. Models, theories and heuristics in apraxia of speech. *Clin Linguist Phon* 1998; **12**: 47–65.
- 4 Croot K. Diagnosis of AOS: definition and criteria. Semin Speech Lang 2002; 23: 267–79.
- 5 Ballard KJ, Granier JP, Robin DA. Understanding the nature of apraxia of speech: theory, analysis, and treatment. *Aphasiology* 2000; **14**: 969–95.
- 6 Kent RD. Research on speech motor control and its disorders: a review and prospective. *J Commun Disord* 2000; 33: 391–428.
- 7 McNeil MR, Robin D, Schmidt D. Apraxia of speech: definition, differentiation and treatment. In McNeil MR ed. Clinical management of sensorimotor speech disorders. Thieme, 1997: 311–44.
- 8 McNeil MR, Pratt SR, Fossett T. The differential diagnosis of apraxia of speech. In Maassen B *et al.* eds. *Speech motor control in normal and disordered speech.* Oxford University Press, 2004: 389–413.
- Dabul B. Apraxia Battery for Adults, second edition. Pro-ed Inc., 2000.
- Wertz RT, LaPointe LL, Rosenbek JC. Apraxia of speech in adults: the disorder and its management. Grune and Stratton, 1984: 318.

- 11 Duffy JR. Motor speech disorders: substrates, differential diagnosis and management. Mosby, 1995.
- 12 Enderby P. Frenchay dysarthria assessment. Pro-Ed. Inc., 1983.
- 13 Goodglass H, Kaplan E, Barresi B. Boston Diagnostic Aphasia Examination, third edition. Lippincott Williams and Wilkins, 2001.
- 14 Kent RD, McNeil M. Relative timing of sentence repetition in apraxia of speech and conduction aphasia. In Ryalls J ed. Phonetic approaches to speech production in aphasia. College-Hill Press,
- 15 Ziegler W. Phonetic realization of phonological contrast in aphasic patients. In Ryalls J ed. *Phonetic* approaches to speech production in aphasia and related disorders. College-Hill Press, 1987: 163-79.
- 16 Odell K, McNeil MR, Rosenbek JC, Hunter L. Perceptual characteristics of consonant production by apraxic speakers. J Speech Hear Disord 1990; 55: 345-59.
- 17 Aichert I, Ziegler W. Syllable frequency and syllable structure in apraxia of speech. Brain Lang 2004; 88:
- 18 Canter GJ. Apraxia of speech and phonemic paraphasia. Aphasiology 1988; 2: 251-53.
- 19 Canter GJ, Trost JE, Burns MS. Contrasting speech patterns in apraxia of speech and phonemic paraphasia. Brain Lang 1985; 24: 204-22.

- 20 Rogers MA, Storkel HL, Planning speech one syllable at a time: the reduced buffer capacity hypothesis in apraxia of speech. Aphasiology 1999; **13**: 793–805.
- 21 Haley KL, Wertz RT, Ohde RN. Single word intelligibility in aphasia and apraxia of speech. Aphasiology 1998: **12**: 715-30.
- Kent RD, Rosenbek JC. Acoustic patterns of apraxia of speech. J Speech Hear Res 1983; 26: 231-49.
- Shuster L.I. Wambaugh JL. Perceptual and acoustic analyses of speech sound errors in apraxia of speech accompanied by aphasia. Aphasiology 2000; 14: 635-51.
- O'Connell PF, Barber KL, O'Connell EJ. Phonological and syntactic effects in silent reading in aphasics with apraxia of speech. Folia Phoniatrica 1985; **37**: 265–70.
- Waters GS, Rochon E, Caplan D. The role of highlevel speech planning in rehearsal – evidence from patients with apraxia of speech. J Memory Lang 1992: **31**: 54–73.
- Towne RL, Crary MA. Verbal reaction-time patterns in aphasic adults - consideration for apraxia of speech. Brain Lang 1988; 35: 138-53.
- Rogers MA, Redmond JJ, Alarcon NB. Parameters of semantic and phonologic activation in speakers with aphasia with and without apraxia of speech. Aphasiology 1999; 13: 871-86.

Copyright of Clinical Rehabilitation is the property of Sage Publications, Ltd. and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.