Reliability and Validity of Cervical Auscultation: A Controlled Comparison Using Videofluoroscopy

Paula Leslie, BSc,[1] Michael J. Drinnan, PhD,[2] Paul Finn, MSc,[3] Gary A. Ford, FRCP,[4] and Janet A. Wilson, FRCS[5]

[1]School of Surgical and Reproductive Sciences, The Medical School, University of Newcastle, Newcastle-upon-Tyne, NE2 4HH; [2]Regional Medical Physics, Freeman Hospital, Newcastle-upon-Tyne, NE7 7DN; [3]Medical Statistics, School of Health and Social Care, University of Teesside, Middlesbrough, TS1 3LF; [4]Institute of Ageing and Health, University of Newcastle, Newcastle-upon-Tyne, NE2 4HH; and [5]Department of Otolaryngology and Head and Neck Surgery, Freeman Hospital, Newcastle-upon-Tyne, NE7 7DN, United Kingdom

Abstract. Cervical auscultation is experiencing a renaissance as an adjunct to the clinical swallowing assessment. It is a controversial technique with a small evidence base. We have aimed to establish whether cervical auscultation interpretation is based on the actual sounds heard or, in practice, influenced by information gleaned from other aspects of the clinical assessment, medical notes, or previous knowledge. We sought to determine (a) rater reliability and its impact on the clinical value of cervical auscultation and (b) how judgments compare with the "gold standard": videofluoroscopy. Swallow sounds were computer recorded via a Littmann stethoscope. Sounds were sampled from 10 healthy control swallows with no aspiration/penetration and 10 patient swallows with aspiration/penetration, all recorded during simultaneous videofluoroscopy. The system generated sound quality similar to "live" bedside listening, a feature rarely seen in cervical auscultation studies. The 20 sound clips were classified as "normal" or "abnormal" by 19 volunteer speech–language pathologists with experience in cervical auscultation. After at least four weeks, 11 of these judges rated the sounds rerandomized on a new CD. Intrarater reliability kappa ranged from −0.12 to 0.71. Individual reliability did not correlate with years of experience, practice pattern, or frequency of use. Interrater reliability kappa = 0.17. Comparison with radiologically defined aspiration/penetration yielded 66% specificity, 62% sensitivity, and majority consensus gave 90% specificity, 80% sensitivity. There was a significant relationship between individual reliability and true positive rate (r_s = 0.623, p = 0.040). The reliability of individual judges varied widely and thus, inevitably, agreement between judges was poor. Validity is dependent upon reliability: Improving the poor raters would improve the overall accuracy of this technique in predicting abnormality in swallowing. The group consensus correctly identified 17 of the 20 clips so we may speculate that the swallow sound contains audible cues that should *in principle* permit reliable classification.

Key words: Cervical auscultation — Swallowing — Reliability — Validity — Videofluoroscopy — Dysphagia — Stroke — Deglutition — Deglutition disorders.

Chronic dysphagia resulting from a wide range of causes [1] contributes to morbidity and mortality [2] and has significant psychosocial effects [3]. Increasing referral numbers to dysphagia services compete for limited resources, with healthcare purchasers demanding clear evidence of beneficial outcomes. The assessment procedures used to diagnose and formulate a management plan and the resultant outcomes are appropriately coming under increasing scrutiny.

The reliability of any technique needs to be determined: A useful dysphagia assessment needs to be sensitive and reproducible within and between clinicians. There are limited data to support many of the assessment procedures used in dysphagia [4]. The bedside or clinical swallow assessment (CSA) varies in its reported reliability [5] and validity in terms of comparison to the "gold standard" videofluoroscopic swallow study (VFSS) [6–8]. This "gold standard" itself has poor reliability for intra- and interjudge rating, and even these studies are limited to between 3 and 10 judges [9–11]. Novel techniques are often supported enthusiastically but with little evidence base. New interventions should be evaluated carefully but comparisons with existing techniques must be tempered with the knowledge that prevalence of use is not evidence of superiority.

Cervical auscultation (CA) is increasingly being used to supplement the CSA. The sounds associated with swallowing have been investigated using accelerometers and microphones for acoustic properties [12] and prediction of aspiration [13]. There are few robust studies of assessment by CA and no consensus has been reached on its reliability or validity. Reliability refers to the trustworthiness of an instrument: Is it consistent in the answers it is giving? Validity asks if the instrument is measuring what we expect. One of the conditions for validity is that an instrument must be reliable [14]. The most recent work suggests that agreement between judges is poor with only a few people suitably consistent within themselves to be classified as reliable [15]. The technique is part of the battery of tools used in the CSA, but it is possible that judgments of the sounds are unduly influenced by what has been read in the notes or already observed at bedside. In other words do raters in fact anticipate acoustic abnormality rather than detect it?

The objective of this study was to identify if clinicians experienced in CA could identify normal/abnormal swallow sounds from listening alone. The aims were to establish in a representative sample of judges:

1. The range of intrarater reliability: *Is an individual consistent?*
2. Their interrater reliability: *Do colleagues agree?*
3. The overall validity of CA against the "gold standard": videofluoroscopy: *Does CA get it "right"?*
4. The association between intrarater reliability and validity of CA judgment: *Do features such as experience or work pattern make an individual more reliable or more right?*

Participants and Methods

Controls

Ten healthy volunteers were recruited to act as the control sample (median age = 72 years, range = 24–78 years). Exclusion criteria were previous history of dysphagia or eating/drinking difficulties, neurological impairment, cardiorespiratory disease, current medical conditions requiring medication, or structural abnormalities that could affect the swallowing or respiratory systems.

Dysphagic Stroke Patients

Over a 6-month period 20 consecutive dysphagic stroke patients (median age = 78 years, range = 65–90 years) who failed the CSA (showed clinical signs of dysphagia and to be at risk of aspiration) [16] were approached. Exclusion criteria were general medical unfitness (the consultant in charge deemed the patient too ill to participate), neurological condition other than stroke, methicillin-resistant *Staphylococcus aureus* (MRSA) as advised by infection control because of the use of noncleanable equipment, previous history of dysphagia or involvement in other studies, presence of tracheostomy tube because of interference with respiration and swallow sounds, or transfer or discharge before they could participate. Of the initial group 14 were recruited (1 transferred, 1 refused, 2 condition worsened, 1 had no next of kin to give assent, 1 MRSA late swab). The patients were studied a minimum of 48 hours poststroke. This allowed the physical system to stabilize and aimed to reduce the anxiety that the patient might experience poststroke. See Table 1 for characteristics of the 14 participating patients.

Swallow Sound Raters

Speech-language pathologists (SLPs) with experience in dysphagia and CA were notified of the study and asked to consider participation. Thirty-one SLPs from regional and national special interest groups and local hospitals agreed to participate in the study. Dysphagia experience ranged from 1 to 13 years (median = 6 years) and CA experience ranged from 1 to 6 years (median = 5 years). Each rater completed a detailed questionnaire, (Appendix 1, SLT = SLP). Participants varied in all aspects addressed in the questionnaire (see Figs. 1–6).

Ethical Approval and Consent

Written informed consent or assent was obtained for all participants in the study. The Newcastle and North Tyneside Joint Ethics Committee granted ethical approval for the study.

Equipment

This easily portable and noninvasive system is a development of an earlier one [16]. The sounds were recorded onto a notebook (Toshiba, Tokyo, Japan) computer hard drive via a Littmann Cardio III stethoscope (3M, Loughborough, UK), with a BL 1994 microphone (Knowles Acoustics, Burgess Hill, UK) inserted into the tubing at the bifurcation. The recording quality of the system was optimized to match what clinicians actually hear at bedside. Tube length and recording quality were modified iteratively until the consensus of two medical physicists (one with perfect musical pitch) and an experienced clinician agreed the sound was as close as

Table 1. Patient characteristics

Patient	Age	Bamford classification	VFSS days postonset
1	90	LPACS	17
2	80	LTACS	12
3	72	LPACS	11
4	74	LPACS	9
5	71	LTACS	5
6	78	LTACS	9
7	66	RPACS	5
8	83	RTACS	10
9	65	LTACS	8
10	89	LPACS	15
11	84	LTACS	20
12	72	LTACS	11
13	80	RTACS	11
14	78	RPACS	12

Bamford classification [22]: L = left, R = right, PACS = Partial Anterior Circulation Stroke, TACS = Total Anterior Circulation Stroke.

possible to the sound heard via an identical unmodified stethoscope at bedside. Three SLPs were blindfolded and asked to listen to live and prerecorded swallow sounds and to comment on the quality. All of the sounds were from one healthy, nondysphagic person swallowing 5 ml water. There was no discernable difference between the live and prerecorded sounds. Interestingly, all of the SLPs reported hearing swallow sounds they thought abnormal and from stroke patients. Many studies have used accelerometers or even microphones but the recordings do not sound like those a clinician actually hears.

The stethoscope head is radiopaque which can cause interference with the image obtained during the VFSS. One option is to use a stethoscope with a radiotransluscent head [15]. We chose to use the radiopaque Littmann Cardio stethoscope as used at bedside to give realistic sounds and optimize recording quality for future acoustic analysis [17]. The stethoscope head was positioned on the neck over the lateral aspect of the thyroid cartilage, just off-center [18] and held in place by an elasticated Velcro band. A preliminary position screening was performed to check the image.

Test Bolus Materials

All studies of control and patient group participants were performed with simultaneous VFSS in the X-ray department. Three boluses each of 5 ml thin barium, 20 ml thin barium, and 5 ml yogurt were presented. E-Z-Paque Barium Sulphate Ph Eur (E-Z-EM Ltd, Bicester, UK) was used as the contrast material. Patients received fewer boluses if it was deemed clinically inappropriate to continue with the VFSS. The "thin barium" was a standardized runny barium sulfate contrast liquid of 52% weight/volume. This was the most dilute material that could be imaged clearly on VFSS but is still not rheologically identical to water. The liquid was stirred frequently to keep the barium in suspension since it settles out quickly, thus affecting the contrast.

The liquids were measured by graduated syringe into a small plastic cup; the participant was asked to drink the entire contents in one swallow in order to mimic real drinking as closely as possible. Injecting materials into the mouth may affect the normal swallow process, whether or not a person is then allowed to swallow at will.
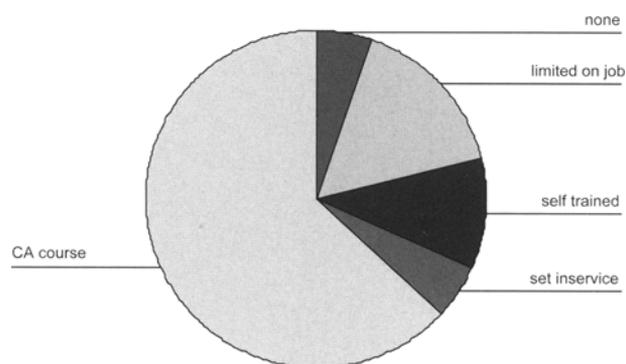
Fig. 1. CA training level (Q3).

Yogurt was measured using an accurate 5-ml medicine spoon. Wherever possible, participants fed themselves. If needed, the SLP supported the cup/spoon or helped to feed while letting the participant guide the process.

The boluses were presented in the same order to all participants. To avoid "learner" effects or fatiguing, the boluses would ideally have been presented in a random order. This was not possible as the materials were presented in the standard order for a VFSS. Analysis of data on the development of the system indicated that there were no learner or fatiguing effects.

Swallow Sound Compact Disk

Ten sound clips were taken from healthy control group swallows (with no aspiration/penetration) and ten from patient group swallows (with radiologically defined aspiration/penetration) giving a total of 20 examples. All patient swallows were aspiration or significant penetration (which is an indication of an "at-risk patient" [19]). Current clinical practice in CA is to classify aspirators and penetrators as "abnormal."

Swallow clips were first identified in the patient group where clear aspiration/penetration had occurred. These clips were then paired with control swallows with no aspiration/penetration. They were matched for bolus consistency and volume, gender, and, as close as possible, the age of the person. This reduced the clips to those from 7 control participants (median age = 73 years, range = 61–78 years) and 7 patient participants (median age = 78 years, range = 65–90 years). Swallow clips that were affected by voicing or coughing were excluded. The microphones were more sensitive than expected and occasionally the clinician could be heard commenting on the swallow. Such examples were also excluded. Where there was a choice of clip the first recording was chosen. This procedure was carried out to minimize the effect of researcher bias in choosing very different or very similar sound clips.

Sound clips were split randomized and recorded onto compact disk (CD). The CDs were sent to the 31 volunteer SLPs with written instructions (Appendix 2), a response form, the questionnaire, and a return post-paid envelope. The SLPs were asked to rate "normal or abnormal" swallow, to say if it was probable or definite, and then give any other qualitative comments. No definitions of "normal" or "abnormal" were given since no standard definitions exist in clinical practice. Responses were re-
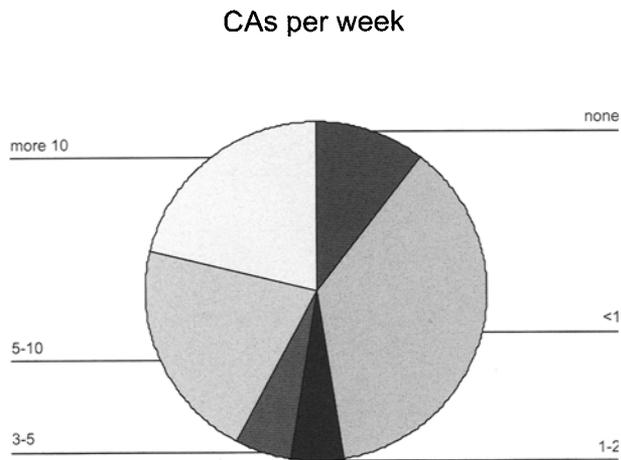
CAs per week

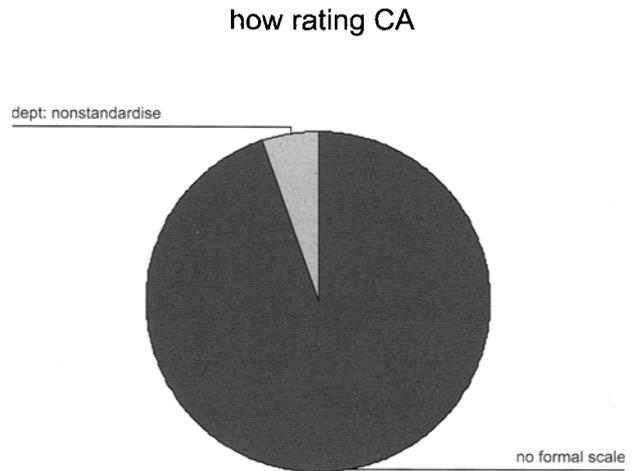**Fig. 2.** CAs performed per week (Q4).

how rating CA

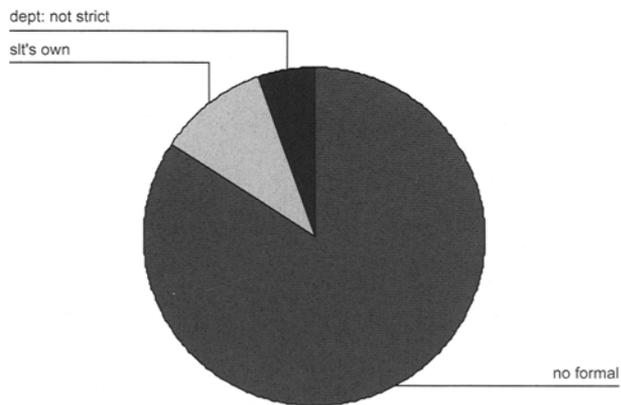**Fig. 4.** How CA is rated (Q7).

Fig. 3. CA procedure (Q6).

ceived from 19 of the 31 SLPs, of whom 15 said they would be prepared to rerate the sounds for intrarater reliability. The sound clips were rerandomized and recorded onto a second CD. After at least 4 weeks the 15 volunteers were sent the new CD and responses were received from 11 of these.

Data Collection/Statistical Analysis

Data were analyzed with SPSS for Windows (Release 11 .0, SPSS Inc, Chicago, IL) and Stata (SE 7, Stata Corporation, College Station, TX) packages. Observed agreement results are quoted together with the kappa values for predicted agreement. Kappa allows for the effect of chance and bias. Correlation of rater characteristics with reliability and validity was analyzed using Spearman's coefficient for ranked non parametric data. For the purposes of statistical analysis of rater reliability it was deemed appropriate to collapse the data into dichotomous variables, hence the division of probable and definite was removed. Results were accepted as statistically significant at the 5% level.

Results

Individual Test Retest Reliability

Are we able to make a consistent judgment on the same sound from one hearing to another?

Figure 7 shows the 11 rerater responses:

- The observed self-agreement, i.e., the number of times an individual's ratings agreed when comparing the same sound on CD1 with CD2
- Whether this self-agreement was on "normal" or "abnormal" sounds
- The derived kappa statistic
- The number of swallows the rater "correctly" identified (as defined radiologically)

Clinicians varied in their self-agreement from 9/20 to 17/20. That is, some individuals could rate the same sounds no better than chance (10/20) but some were much more consistent. To allow for chance and personal rating bias, the kappa statistic was calculated.

$$kappa = \frac{(observed\ agreement - agreement\ by\ chance)}{(perfect\ agreement - agreement\ by\ chance)}$$

perfect agreement = 1

Chance agreement is 0.5 for an unbiased observer. Chance agreement becomes more likely when an observer has an overall bias toward one or another rating. For example, rater 2 self-agreed on 13 of 19 swallows. Rater 2 (with 22 normal, 16 abnormal ratings) had a kappa of 0.38. Rater 17 also self-agreed on 13 of 19 swallows but was rather more biased (12 normal, 26 abnormal) and had a correspondingly lower kappa of 0.27.

Intrarater individual kappas ranged from −0.12 to 0.71 with a mean of 0.35. Of these 11 SLPs, 7
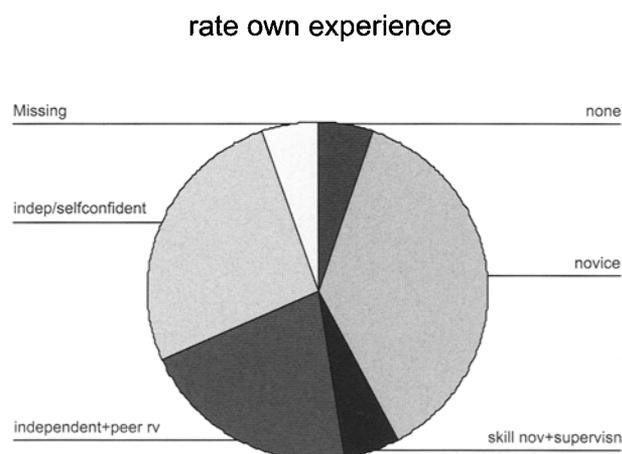
rate own experience

Fig. 5. Self-rating of experience level (Q8).
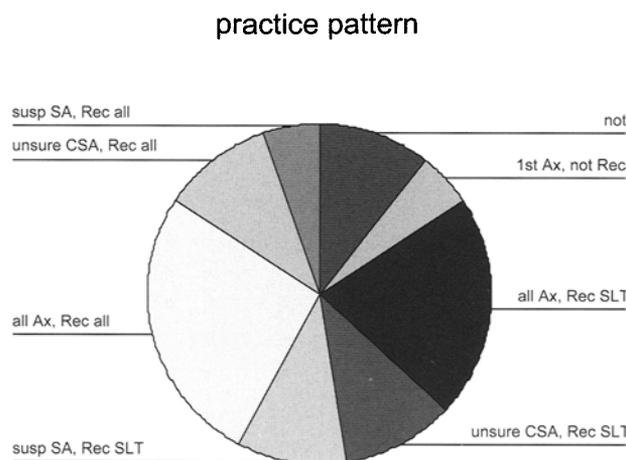
practice pattern

Fig. 6. CA practice pattern: When is CA used and where are the findings recorded (Q9)?

judges rated "fair" or better according to the Landis and Koch guidelines [20]:

kappa value <0.20 poor agreement
    0.21–0.40 fair
    0.41–0.60 moderate
    0.61–0.80 good
    0.81–1.00 very good
    0 = chance, <0 = worse than chance

Predictors of intrarater reliability

In total 148 swallows were classified the same on both ratings (Fig. 7). Of these, 74 (50%) were classified as normal. This finding implies that across the group, as a whole, self-agreement was not linked to whether a rater was classifying a swallow as normal or abnormal.

Of the 148 self-agreed swallows, 102 agreed with VFSS, i.e., were "correct." Of these, 50 (49%) swallows were classifyed as normal, i.e., self-agreement classification was not influenced by whether the rating was VFSS "correct." There was no correlation between an individual's reliability (kappa) and "correctness" of self-agreement.

There was no correlation between an individual's reliability and his/her behavior/practice pattern/ experience nor self-acclaimed expertise level.

Group Agreement

Do we hear what our colleagues hear?

The interrater reliability of the 19 clinicians (based only on the first reading for the reraters) gave kappa of 0.17. For radiologically normal swallows in control participants (VFSS NO aspiration/penetration), kappa = 0.02. For radiologically abnormal

swallows in poststroke participants (VFSS aspiration/penetration), kappa = 0.18. The lower kappa value could be interpreted as people being even more unreliable when rating normal swallows but both conditions were "poor." Indeed, for 10 of the 20 swallows (Nos. 3, 4, 5, 7, 9, 10, 12, 17, 18, 20), there were raters who classified the swallow as "definitely normal" *and* raters who classified it "definitely abnormal."

The interrater kappa of the 11 reraters (first reading) was 0.13, indicating that they were representative of the original 19 judges. We know from the intrarater results that there is a range of individual reliability, i.e., individuals will be affecting the overall reliability of the group. Similarly, some sound clips were more consistently rated than others, e.g., sound No. 1 was overwhelmingly scored as normal when aspiration/penetration did occur. This may have been clinically silent.

Figure 8 shows the total ratings for normal or abnormal for each of the 20 sounds rated by the whole group of 19 raters. How do we decide where the cutoff is for a significant vote one way or the other? Using the Landis and Koch guidelines:

For "fair" agreement, kappa ≥ 0.21, so

$$0.21 = (X - 0.5)/(1 - 0.5), \text{ where}$$

  $X =$ observed agreement

$X = \{0.21(1 - 0.5)\} + 0.5,$

$X = 0.605,$ i.e., 60.5% of raters

  must vote in one direction,

$X = 11.5$ of the 19 raters.

A score of greater than 11.5, i.e., 12, in one direction would show "fair" agreement between the

Swallow no

Rater no	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Kappa
2	o	o	?	o	o	o	o	?	o	?	o	•	?	•	?	X	•	•	•	?	0.38
3	o	o	?	o	o	?	o	o	?	?	o	?	?	•	•	o	•	•	•	?	0.30
4	o	o	?	o	o	?	o	•	?	o	?	?	?	?	?	?	?	?	•	•	-0.12
5	o	o	•	o	o	o	o	o	o	?	?	•	•	•	o	•	•	?	•	•	0.71
6	o	o	•	o	?	o	o	o	o	o	o	•	o	•	?	•	•	?	•	?	0.58
9	o	o	•	o	•	?	?	?	o	o	o	•	•	•	o	?	•	•	•	•	0.61
11	o	o	o	?	o	o	•	o	o	?	o	•	•	?	?	•	•	•	•	•	0.60
12	o	o	o	o	•	?	•	?	?	o	•	?	•	•	?	?	?	•	?	•	0.20
14	?	?	?	•	?	o	•	?	•	•	•	?	?	o	o	o	o	•	?	•	0.20
15	o	?	o	o	o	?	?	?	?	•	?	o	•	?	?	•	•	?	•	•	0.12
17	X	?	?	?	o	?	?	?	•	•	•	•	o	•	o	•	•	•	•	•	0.27

■ Abnormal by VFSS 'gold standard'
o Two 'normal' ratings
• Two 'abnormal' ratings
? Inconsistent ratings
X Not rated on two occasions

Data are presented for the eleven raters who agreed to re-rate the swallows

Fig. 7. The first and second ratings of the same 20 swallows by 11 SLPs. Note that circles indicate swallows that were rated the same on both occasions. Whether the rating agreed with VFSS can be judged from the background shading; open circles should lie on a white background and vice versa.

raters. Similarly, or "moderate" agreement 14 raters and for "good" agreement 16 raters must vote in one direction.

- 16 sounds were rated with fair agreement, 8 of which were "abnormal" on VFSS
- 11 of these 16 sounds were rated with moderate agreement, 6 of which were "abnormal" on VFSS
- 3 of the 11 sounds were rated with good agreement, all 3 of which were "abnormal" on VFSS.

A major concern for the researchers was that raters would guess that there were 10 normal and 10 abnormal sounds. Abnormal ratings of exactly 50% were obtained from only four raters in the interrater group. Abnormal ratings ranged from 6 to 13 of the 20 sound clips.

Validity

Do our judgments agree with VFSS?

Using VFSS as the "gold standard," 125/190 normal ratings matched no aspiration/penetration, i.e., specificity of 66%. Similarly, 117/190 abnormal ratings matched aspiration/penetration, i.e., sensitivity of 62%. Ratings can be used to predict the results of VFSS (via positive and negative predictive values); however, as this study is concerned with assessing raters' reliability and ability to match the gold standard, sensitivity and specificity values have been calculated. It is inappropriate to calculate predictive values where the prevalence has been artificially controlled.

Sensitivity and specificity are based on individuals' ratings. If we look at majority consensus (Fig. 8, bottom row) the *group* correctly identified 9/10 normal and 8/10 abnormal swallows (90% specificity and 80% sensitivity).

There was a significant relationship between an individual's reliability and true positive rate ($r_s = 0.623$, $p = 0.040$).

Discussion

This is the largest study to date of rater reliability in CA, using 19 initial raters, 11 reraters, and 20 swallow sounds. There is a dearth of robust studies in the field of CA. Studies involving reasonable numbers of swallow sounds have often had few raters, lack of a healthy, asymptomatic control group, or lack of simultaneous VFSS for a gold standard comparator [13,15]. This is one of the few studies to have recorded swallow sounds that match what a clinician really hears at bedside. This study isolated sounds to reduce the influence of previous knowledge.

Reliability of CA

The data clearly demonstrate that few people are reliable in CA. This unreliability in an individual's judgments will affect the usefulness of acoustic characterization of swallow sounds as reported by Cichero and Murdoch [12]. The clinical applicability of

Swallow no

Rater no	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	o	o	o	o	o	o	o	o	o	o	o	o	o	o	•	•	•	•	•	•
2	o	o	o	o	o	o	o	o	o	o	o	•	o	•	•	o	•	•	•	o
3	o	o	o	o	o	•	o	o	o	•	o	o	o	•	•	o	•	•	•	o
4	o	o	o	o	o	o	o	•	o	o	o	o	•	•	•	•	o	•	•	•
5	o	o	•	o	o	o	o	o	o	o	o	•	•	•	o	•	•	o	•	o
6	o	o	•	o	•	o	o	o	o	o	o	•	o	•	•	•	o	•	•	o
7	o	o	o	o	o	•	X	•	•	o	•	o	o	o	o	•	•	•	•	•
8	o	o	o	•	o	o	o	o	o	•	•	o	•	•	o	•	o	•	•	•
9	o	o	•	o	•	o	o	o	o	o	o	•	•	•	o	•	•	•	•	•
10	o	•	o	•	o	o	o	o	•	o	o	•	•	o	•	o	•	•	o	•
11	o	o	o	o	o	o	•	o	o	•	o	•	•	o	•	•	•	•	•	•
12	o	o	o	o	•	o	•	o	•	o	•	o	•	•	•	•	o	•	o	•
13	o	o	•	o	o	•	o	o	o	o	o	•	o	•	•	•	•	•	•	•
14	•	•	o	•	o	o	•	•	•	•	•	o	o	o	o	o	o	•	o	•
15	o	o	o	o	o	•	•	•	o	•	•	o	•	•	•	•	o	o	•	•
16	o	o	o	•	•	o	o	•	o	o	•	o	•	•	•	•	•	•	•	•
17	X	•	o	o	o	•	o	•	•	•	•	•	o	•	o	•	•	•	•	•
18	o	•	•	•	o	o	•	o	•	•	•	•	•	o	•	•	o	•	•	•
19	•	o	o	o	•	o	•	o	o	•	•	•	•	o	•	•	•	•	•	•
Consensus	o	o	o	o	o	o	o	o	o	o	o	•	•	•	•	•	•	•	•	•

▪ Abnormal by VFSS 'gold standard'
o Rated 'normal'
• Rated 'abnormal'
X Not rated

Rater numbers and swallow numbers are as for Figure 7
The consensus is the rating made by the majority (10 or more) raters

Fig. 8. The first ratings of 20 swallows by 19 SLPs. Whether the rating agreed with VFSS can be judged from the background shading; open circles should lie on a white background and vice versa.

such detailed analysis will be hampered until we can improve intrarater reliability.

Demographic data analysis revealed that reliability of a clinician is independent of factors historically presumed to improve skills, such as years of experience (Figs. 1–6). The definitions of aspects such as practice and experience were left deliberately vague. We do not know if any person can be trained to improve his/her listening skills, or are there some whose background or innate ability predisposes them to be more reliable auscultators? For example, is someone with a "musical ear" at an advantage when listening and characterizing sounds of any type? Clinicians using this technique do not get their hearing tested routinely: this could be a significant factor affecting an individual's performance.

More detail is required before we could draw definite conclusions on, say, how much and what type of training a clinician should have, and continuing peer review. Training and discussion have been shown to improve the reliability in VFSS judgments [21]. A system such as we describe here will be in-

valuable in future studies to address the manifestly unstandardized current approaches to CA. We would then be in a position to quantify the "added value" of this technique to the CSA when investigating the effect of previous knowledge of patient history for example.

The agreement among the group of raters was poor. This matches established techniques such as VFSS and laryngoscopic swallow studies.

Validity

The overall accuracy of the technique in identifying aspiration/penetration was limited because of individual variation in reliability. The validity of a technique is dependent upon reliability and this is borne out by the results of this study. Since the group consensus correctly classified 17 of the 20 swallows, we may speculate that the swallow sound contains audible cues that should *in principle* permit reliable classification. If we could improve the poor raters, we would improve the overall accuracy of the technique

in detecting abnormality in swallowing. The question of auscultation improving the accuracy of the CSA has yet to be answered but these issues will affect it.

The Future

What is the physiology behind swallow sounds? Future work should expand the limited evidence base of synchronized sounds and images from both VFSS and laryngoscopy. Analysis of simultaneous sound and image data will contribute to the continuing debate: What can cervical auscultation detect and what does it contribute to the clinical swallow assessment?

Appendix 1

Questions Relating to Rater Experience

Q1. How many years have you been working in the field of dysphagia _____ years

Q2. How many years have you been performing Cervical Auscultation?
☐ Not performing Cervical Auscultation
☐ Training course experience only
☐ Less than 1 year
☐ 1-2 years
☐ 2-3 years
☐ 3-5 years
☐ 5-10 years
☐ 10+ years

Q3. What has been your level of training?
☐ No training whatsoever
☐ Limited informal "on the job" training
☐ Self education
☐ Established work based "in-service" training program
☐ Attended Cervical Auscultation training course
Please specify _____
☐ Other
Please specify _____

Q4. On average, how many examinations with Cervical Auscultation do you perform per week?
☐ I do not perform Cervical Auscultation at my facility
☐ Less than 1
☐ 1-2
☐ 2-3
☐ 3-5
☐ 5-10
☐ 10+

Q5. How are you using Cervical Auscultation at your facility? (Tick all applicable)
☐ Ad-hoc service/nothing formal in place
☐ Occasional assessments only
☐ Cervical Auscultation assessment and diagnostic clinic for inpatients and outpatients
☐ Therapy and biofeedback
☐ On hospital wards, for example ITU
☐ Other
Please specify _____

Q6. What is the procedure of Cervical Auscultation administration?
☐ No formal protocol/administration procedure
☐ Therapists involved have their own protocol/set of procedures
☐ Department staff has developed some protocol/set of procedures, although this is not rigidly adhered to
☐ Department staff has developed a protocol with strict guidelines for administration procedures, consistencies and volumes of bolus
☐ Protocol guidelines have been taken from literature and modified to suit the departmental requirements
☐ Other
Please specify _____

Q7. How are you rating Cervical Auscultation at present?
☐ No formal rating scale in place/therapist's own narrative comments
☐ Department rating scale (non-standardized)
☐ Department rating scale (standardized/some validity and reliability measures taken)
☐ Published rating scale from literature, for example:
Please specify _____
☐ Modified published rating scale
Please specify _____
☐ Other
Please specify _____

Q8. How would you rate your level of experience?
☐ No experience
☐ Novice/just starting
☐ Skilled novice/still need supervision
☐ Skilled clinician ready for independence
☐ Working independently but with frequent peer review by senior staff
☐ Independent and confident in own abilities
☐ Specialist in this field

Q9. What is the practice pattern for Cervical Auscultation?
☐ Perform Cervical Auscultation at all swallow examinations but do not record findings
☐ Perform Cervical Auscultation at all swallow examinations and record in SLT notes
☐ Perform Cervical Auscultation at all swallow examinations and record in all notes
☐ Perform Cervical Auscultation only at initial swallow examination but do not record findings
☐ Perform Cervical Auscultation only at initial swallow examinations and record in SLT notes
☐ Perform Cervical Auscultation only at initial swallow examinations and record in all notes
☐ Perform Cervical Auscultation only when unsure of clinical findings but do not record findings
☐ Perform Cervical Auscultation only when unsure of clinical findings and record in SLT notes
☐ Perform Cervical Auscultation only when unsure of clinical findings and record in all notes
☐ Perform Cervical Auscultation only when suspect silent aspiration but do not record findings
☐ Perform Cervical Auscultation only when suspect silent aspiration and record in SLT notes
☐ Perform Cervical Auscultation only when suspect silent aspiration and record in all notes
☐ Not performing Cervical Auscultation
☐ Other
Please specify _____

Q10. Who participates in the Cervical Auscultation assessment?
☐ 1 SLT experienced in Cervical Auscultation
☐ 2 SLTs experienced in Cervical Auscultation
☐ 1 SLT experienced and 1 SLT training in Cervical Auscultation
☐ 1 SLT and 1 student
☐ Other Please specify _____

Q11. Aetiology of patient(s) receiving Cervical Auscultation? (Tick all applicable)
☐ Head and neck cancer/ENT
☐ Stroke
☐ Pulmonary/COPD
☐ General medical
☐ Closed head injury
☐ Dementia/Alzheimer's
☐ Neurosurgery
☐ Progressive neurological disorder Please specify _____
☐ Neurological developmental, i.e., adults with learning disability
☐ Psychological component
☐ Other, nonspecific swallow complaint Please specify _____

Appendix 2

Instructions for Raters

1. Before listening to the CD, could you please fill out the demographics questionnaire relating to relevant experience and current practice in your facility. Please add any additional information you feel is relevant when asked to specify, and if you select the category *Other*, please clarify with a short comment.

2. The sound quality is best if you listen through headphones. You can use a PC or a standard CD player but the speakers are often not too good and we have optimized the sound quality to try to replicate the use of a standard stethoscope. Make sure the shuffle switch is not on,
    so the swallows are played in order 1–20. Each track is approximately 8 seconds long with a silence of 12 seconds to space the tracks. The swallow may occur anywhere in the 8 seconds. The microphone was very sensitive and so you will often hear voices; please try to ignore these and concentrate on the swallow sounds.
3. Try not to spend too much time deliberating over answers as in real life you would only hear a swallow once. Listen to each track as often as you feel is necessary. I must highlight that the completion of the questionnaire and the rating is an individual task, please **do not confer** with colleagues. The project involves interrater reliability not group consensus.
4. Tick or cross the box to rate each swallow as definitely normal **or** probably normal **or** probably abnormal **or** definitely abnormal. Please write down any comments that you feel are relevant, e.g., judgments you would make at bedside based on what you heard, bolus type, aspiration/penetration/pooling/residue. etc.
5. **Please attempt every question, do not leave any blank.**
    If a question states *tick all applicable*, please select more than one answer if you feel it is relevant.
6. Although the questionnaire is based on ticking a box, please do quality judgments or list alternatives that are not provided.
7. Please, return your questionnaire and rating form, using the envelope provided, to Newcastle University. The due date for returning them is **Friday 31 May**. If you have any queries, please do not hesitate to contact me at Newcastle University by telephoning 0191 222 6279 or via e-mail: paula.leslie@ncl.ac.uk. If you do not want to give your details that is fine. All data is confidential and will be made anonymous for analysis and publication.
8. Ethical Approval and Consent
    Written informed consent was obtained for all participants involved in the study. For those patients with language disorders, consent was obtained from the next of kin and the researcher explained the study as far as possible to the patient with gesture and modified language. The Joint Ethics Committee for Newcastle and North Tyneside Health Authority, the University of Newcastle, and the University of Northumbria at Newcastle granted ethical approval for the study.

References

1. Leslie P, Carding P, Wilson J: Investigation and management of chronic dysphagia. *BMJ 326:*433–436, 2003
2. Langmore SE, Terpenning MS, Schork A, Chen Y, Murray JT, Lopatin D, Loesche WJ: Predictors of aspiration pneumonia: how important is dysphagia? *Dysphagia 13:*69–81, 1998
3. Ekberg O, Hamdy S, Woisard V, Wuttge-Hannig A, Ortega P: Social and psychological burden of dysphagia: its impact on diagnosis and treatment. *Dysphagia 17:*139–146, 2002
4. Martino R, Pron G, Diamant N: Screening for oropharyngeal dysphagia in stroke: insufficient evidence for guidelines. *Dysphagia 15:*19–30, 2000
5. McCullough GH, Wertz RT, Rosenbeck JC, Mills RH, Ross KB, Ashford JR: Inter- and intrajudge reliability of a clinical examination of swallowing in adults. *Dysphagia 15:*58–67, 2000
6. Linden P, Kuhlemeier KV, Patterson C: The probability of correctly predicting subglottic penetration from clinical observations. *Dysphagia 8:*170–179, 1993
7. Mann G, Hankey GJ: Initial clinical and demographic predictors of swallowing impairment following acute stroke. *Dysphagia 16:*208–215, 2001
8. Mari F, Matei M, Ceravolo MG, Pisani A, Montesi A, Provinciali L: Predictive value of clinical indices in detecting aspiration in patients with neurological disorders. *J Neurol Neurosurg Psychiatry 63:*456–460, 1997
9. Kuhlemeier KV, Yates P, Palmer JB: Intra- and interrater variation in the evaluation of videofluorographic swallowing studies. *Dysphagia 13:*142–147, 1998
10. McCullough GH, Wertz RT, Rosenbeck JC, Mills RH, Webb WG, Ross KB: Inter- and intrajudge reliability for videofluoroscopic swallowing evaluation measures. *Dysphagia 16:*110–111, 2001
11. Wilcox F, Liss JM, Siegel GM: Interjudge agreement in videofluoroscopic studies of swallowing. *J Speech Hear Res 39:*144–152, 1996
12. Cichero JAY, Murdoch BE: Acoustic signature of the normal swallow: characterization by age, gender, and bolus volume. *Ann Otol Rhinol Laryngol 111:*623–632, 2002
13. Zenner PM, Losinski DS, Mills RH: Using cervical auscultation in the clinical dysphagia examination in long-term care. *Dysphagia 10:*27–31, 1995
14. Division of Access and Continuing Studies. Research Methods Glossary. University of Bath, http://www.bath.ac.uk/dacs/gold/glossary.html#N1934, 2000
15. Stroud A, Lawrie B, Wiles C: Inter- and intra-rater reliability of cervical auscultation to detect aspiration in patients with dysphagia. *Clin Rehabil 16:*640–645, 2002
16. Leslie P, Drinnan MJ, Ford GA, Wilson JA: Resting respiration in dysphagic patients following acute stroke. *Dysphagia 17:*208–213, 2002
17. Hamlet S, Penney DG, Formolo J: Stethoscope acoustics and cervical auscultation of swallowing. *Dysphagia 9:*63–68, 1994
18. Takahashi K, Groher ME, Michi K: Methodology for detecting swallowing sounds. *Dysphagia 9:*54–62, 1994
19. Pikus L, Levine M, Yang Y, Rubesin S, Katzka D, Laufer I, Gefter W: Videofluoroscopic studies of swallowing dysfunction and the relative risk of pneumonia. *Am J Roentgenol 180:*1613–1616, 2003
20. Landis J, Koch G: The measurement of observer agreement for categorical data. *Biometrics 33:*159–174, 1977
21. Scott A, Perry A, Bench J: A study of interrater reliability when using videofluoroscopy as an assessment of swallowing. *Dysphagia 13:*223–227, 1998
22. Bamford J, Sandercock P, Dennis M, Burn J, Warlow C: Classification and natural history of clinically identifiable subtypes of cerebral infarction. *Lancet 337:*1521–1526, 1991