



National Twelve-Bell Striking Contest

Defining Good Ringing (or to be more precise, how to define imperfections in ringing)

Survey Results

Thank you very much to all 29 participants for your thoughtful and detailed responses.

Participants:

Jonathan Agg	Phil Barnes	Matthew Beadman	Mark Bell	Graham Bloom
Wendy Bloom	Tim Bradley	Alex Byrne	Martin Cansdale	Julia Cater
David Dearnley	Jennie Earis	Mark Gill	Tom Griffiths	Tom Hinks
John Hughes-D'Aeth	Chris Kippin	Colin Lee	Simon Linford	Ed Mack
Tom Mack + other SRCY	Paul Mounsey	Chris Poole	Ian Roulstone	Philip Saddleton
Katie Town	George Unsworth	Stef Warboys	Lucy Woodward	

So that you can follow an individual's answers, each respondent has been assigned a number, R1, R2, etc. The numbers have been randomly assigned to preserve anonymity.

To preserve anonymity, we have blanked out a few words where for example a respondent has written "when I judged in xxxx" or "when we won in xxxx". This only affects 3 or 4 sentences in the entire document.

Thank you very much from the Working Group (Neil Buswell, David Pipe, Chris Poole, and Stef Warboys).

1. What are the key metrics you'd look for in defining good ringing, and how important are they? Please mark each with a score of 1-5.

	1 (not very important)	2 (less important)	3 (doesn't massively matter)	4 (important)	5 (very important)
<i>Average speed for the touch</i>	*	***	****	*	*
<p>4-star replies: R8. Needs to be appropriate to the weight of the bells to sound good</p> <p>3-star replies: R1. Many rings of bells have a 'natural' rhythm which, if found, supports good ringing, but it's not a requirement of 'good ringing' per se. 'Good ringing' can be faster or slower on any ring of bells. R2. Either a slow or fast touch can perform well if executed well, but if a touch is too fast then it almost inevitably loses accuracy and can't perform well either by ear or within the strikeometer than more measured pieces of ringing. Slow pieces can perform well on the strikeometer if clean but lack momentum and can be less pleasant to listen to unless exceptionally well executed. R7. ... as long as the band can cope with the speed they are ringing at.</p> <p>2-star replies: R3. The average speed isn't anything other than interesting caveat to a judge. If I was ringing I would probably rank it higher. R6. In itself, not very important other than at extremes (e.g. ASCY band at Aston). But choosing the "wrong" speed can obviously have a powerful influence on other metrics. R9. If the ringing is good, speed should not be a factor in judgement.</p> <p>1-star replies: R5. This metric should not carry too much / any weight as a range of approaches can be successful and I would not want to appear to be judging on the basis of personal preferences around speed. It is, however, the case that decisions about speed can have an impact on a band's ability to ring well on particular bells – sometimes this is worth reflecting on in comments, but only if carefully crafted to make it clear that the speed did not affect ranking of teams.</p> <p>(Non-marked): R11. How is the average being calculated?</p>					

Consistency of speed across the touch			***	*****	
<p>4-star replies:</p> <p>R2. The strongest teams typically have a very strong and consistent beat set by the back bells. Without this, and with a changing pace, the ringing tends to be of lower quality, particularly if it speeds up significantly within the touch.</p> <p>R3. The best ringing to participate in or listen to has consistent timings across every lead</p> <p>R6. Important to create the sense of a coherent, controlled piece of ringing.</p> <p>R8. Changes in speed should be consistent across the change and ideally not too many per touch</p> <p>R9. Important rather than very important, as if there are small speed variations and the whole band respond, it should not detract too much.</p> <p>3-star replies:</p> <p>R1. A consistent speed helps the touch improve as it goes along, so the best ringing ends up being consistent, but it's not a requirement of 'good ringing'.</p> <p>R5. This is more important – though in practice, speed across the touch has not tended to vary massively in many of the test pieces I have judged. The key thing with any change in speed would be whether it was executed successfully by the entire band.</p> <p>R7. As above, and as long as they all speed up or slow down together.</p> <p>(Non-marked): R11. Consistency with respect to what?</p>					
Consistency of leading (with a handstroke gap), denoted by a standard deviation				**	*****
<p>5-star replies:</p> <p>R1. This is a requirement of 'good ringing'.</p> <p>R3. The corner stone of the change?! It may be just as useful to analyse the relationship of the bell ringing in 2nd and 1st place.</p> <p>R6. Critical. Erratic leading is at the root of most poor touches.</p>					

R7. Bad leading causes many problems.

R8. Same gap spacing between bells is striking

4-star replies:

R2. Same principle as consistency of speed above. The quality of the leading is key in defining the rhythm and without this it is harder to settle into even ringing across the touch.

R5. This is important in setting the tone for the rest of the change. Inconsistencies in leading are also relatively easily spotted by human judges!

(Non-marked): R11. Deviation from what? How much handstroke gap?

*** Consistency of tenor ringing, denoted by a standard deviation**

**

5-star replies:

R1. This is a requirement of 'good ringing'. I believe the tenor plays a special role in setting the pace and style, so faults from the tenor have a greater impact (not a greater level of fault) compared to other bells. The same applies, although to a lesser extent, to bells 9-11.

R2. Almost always differentiates the best teams in my experience.

R5. Also important due to the impact it has on the overall rhythm of the touch

R7. The tenor is the "easiest" bell to pick out so, in that sense, it is more important, and more obvious when wrong.

R8. All bells should be treated equally. Confusion from competitors may come from Treble or Tenor being easier to comment on.

4-star replies:

R6. Very important in Cinques, much less so (although still significant) in Maximus.

(Non-marked): R11. Deviation from what?

* Consistency of treble ringing, denoted by a standard deviation		*	**	***	****
<p>5-star replies: R1. This is a requirement of ‘good ringing’.</p> <p>R8. All bells should be treated equally. Confusion from competitors may come from Treble or Tenor being easier to comment on.</p> <p>4-star replies: R5. Important – and the little bells more broadly (see below)</p> <p>3-star replies: R2. A poorly rung treble can stand out (negatively), particularly in certain methods or touches, but otherwise I wouldn’t say it’s materially more important than other bells, expect that the quality of the leading in rounds at the start of the touch helps settle the pace and cadence for the remainder of the piece.</p> <p>R3. No more than any other bell!</p> <p>2-star replies: R6. Not at all important in Cinques. Slightly (but not much) more so in Maximus.</p> <p>(Non-marked): R11. Deviation from what?</p>					
* Consistency of not the treble or tenor (i.e. another bell), denoted by a standard deviation	*		**	***	****
<p>5-star replies: R1. This is a requirement of ‘good ringing’.</p> <p>R8. All bells should be treated equally. Confusion from competitors may come from Treble or Tenor being easier to comment on.</p> <p>4-star replies: R3. Bells 10 and 11 would need to be as consistent as the tenor in well performed touches</p> <p>R5. Little bells working together (runs etc) and around the back bells; back bells other than the tenor supporting the rhythm and framework</p>					

3-star replies:

R2. All bells need to be rung well for a strong touch – any one bells underperforming will both really bring down strikeometer results and stand out to the listener. I'd say that these are of equal importance to the treble.

1-star replies:

R6. Of no relevance beyond accrual of individual faults.

(Non-marked): R11. Deviation from what?

Consistency of length of change (stringing bells out at the back of the change, or failing to fully reach the back of the change)

5-star replies:

R1. This is a requirement of 'good ringing'.

R2. The best pieces have even length rows which remain consistent throughout the touch. Bells dragging out, or dropping in at lead, destabilise the rhythm and negatively stand out. Small inconsistencies in individual bells are less important than an overall strong rhythm.

R7. Like poor leading, cause of many problems.

R8. Likely to impact on leading/covering (in cinques years) if don't achieve this.

4-star replies:

R5. This is very noticeable when it goes wrong – dropped or stretched blows at the back spoil that individual row and potential disrupt the leading going into the next row.

R6. Can be really annoying, especially the front bells. Very rarely justified in my book, whatever some others may say!

(Non-marked): R11. Consistency with respect to what?

Consistency in maintaining the same inter-bell gap throughout the change			*	*	*****
<p>5-star replies: R1. This is a requirement of ‘good ringing’.</p> <p>R3. The difference between good ringing and excellent ringing?</p> <p>R5. Very important – and the ‘throughout the change’ is really key here I think. Developmental comments by judges often focus on leading / the back of the change, both of which are important – but the best pieces achieve consistency all the way through the row, and it’s important teams strive towards this.</p> <p>R8. See above</p> <p>4-star replies: R2. I think this is just a combination of length of change and standard deviation of the individual bells? In the end, the inter-bell gap, and whether this is too big or not big enough, is the definition of good / bad striking, so is very important. It is possible to have a good touch without this being perfect if the overall beat is strong though.</p> <p>3-star replies: R6. Isn’t this basically the same as consistency of speed?</p> <p>(Non-marked): R11. Consistency with respect to what?</p>					
Counts of overall major faults (noticeable large clips, clashes / noticeable large gaps)		*		***	*****
<p>5-star replies: R2. Ultimately, good striking is about what the listener can hear. Major faults stand out much more and detract more significantly from the quality of ringing in my view. To illustrate, I’d say that a row with 10 perfect bells and 2 major fault bells would sound worse than a row with 12 bells all of whom are just about right but a little bit out.</p> <p>R5. Fundamentally, avoiding errors (large or small) is what bands in striking competitions have been trying to achieve for centuries – bands achieving the best ringing on the other metrics listed in this survey will inevitably generate a low fault count. I would emphasise, however, that I do not believe fault counting by human judges is an approach which can achieve consistent judging across a contest with up to 10 teams ringing on 12. Hawkear generating a fault count measure certainly is helpful, in conjunction with other measures.</p>					

<p>4-star replies: R3. In a national competition I would argue the basic aim is to get through the touch without making a mistake. Major faults don't happen very often and would be picked up by the judging team as a matter of course</p> <p>R6. Clearly major faults leave a bad impression, even if quickly corrected.</p>					
Counts of overall minor faults (small, noticeable errors)			*	*****	*****
<p>5-star replies: R3. Bread and butter!</p> <p>R5. See above</p> <p>4-star replies: R2. These accumulate in the strikeometer and are harder to pick up in fast ringing, but remain important if noticeable.</p> <p>R6. Important, but for a different reason – as indicating a lack of attention to the fine detail that distinguishes a good performance from an average one.</p>					
Persistent errors from a particular bell or group of bells			**	*****	****
<p>5-star replies: R1. It's a requirement of 'good ringing' that this doesn't happen, so it's very important. This issue suggests a problem with either individual bells or ringers.</p> <p>R2. This can bring down a touch, particularly if a single bell has a consistent material error at one stroke. When judging in the past, I've seen this significantly impact the final placing of a band. I think what's difficult though is when it's clear that a bell cannot be heard properly from within the tower, so it's only band that are "working to the strikeometer" that adjust for this. A clear example was at XXXX (8th or 10th, I can't clearly remember) was oddstruck at backstroke and some teams had consistent major errors at this stroke all the way through.</p> <p>4-star replies: R3. As judges we are discouraged from singling out individual bells as part of the comments. But it is useful to understand when dissecting a test piece.</p> <p>R5. Can have a significant impact on the overall piece.</p>					

<p>R8. Likely to be commented on – does not mean it counts more</p> <p>3-star replies: R6. It depends! For example, persistent errors in roll-ups will stick out a mile, but similar errors “buried” in the middle of a course may not.</p>					
<i>Persistent errors in certain parts of the change</i>			**	*****	***
<p>5-star replies: R1. It’s a requirement of ‘good ringing’ that this doesn’t happen, so it’s very important. This issue suggests a problem with the method.</p> <p>4-star replies: R3. This would be helpful in aiding constructive criticism!</p> <p>R5. As above</p> <p>R8. See above</p> <p>3-star replies: R2. Has the biggest impact on the listener if this is at the front or the back of the change.</p> <p>R6. See above. Possibly doesn’t add much to an overall impression of inconsistency.</p>					
<i>How well the bells (or groups of bells) work together as a team</i>				**	*****
<p>5-star replies: R1. This reflects the ‘style’ of ringing, which is very important. Roll-ups, for instance, should be a particular point of focus for any band in a performance and I would expect more ‘working together’ to get these right.</p> <p>R2. Very important. A band of 12 good ringers will be perform better than a band with 11 excellent ringers and 1 underperformer when ringing in a competition of the standard of the 12 bell. The top performing band will not have a materially weak link.</p>					

R5. Affects the ability of a band to rapidly generate rhythmic ringing – eg support from back bells in creating a strong framework. Affects a band’s ability to respond to changes in speed if they happen, and to avoid mistakes propagating across changes.

R6. What we are all striving for!

R8. Likely to lead to other good metrics, ie consistency.

How well large and smaller bells work together

**

5-star replies:

R6. One of the key indicators of “good” ringing.

R8. Yes, but also small-middle and middle-back and other permutations. Is important everyone works well together.

4-star replies:

R5. Also important – especially in the methods usually selected for the 12 bell contest, eg Stedman / Cambridge, where small and large bells often work together.

3-star replies:

R1. It’s often the case that many faults are created when large and small bells work together, but faults are just faults.

R2. Important, but I’d say equally important as any other bells working together. Trebles crashing down on big bells is always obvious to the listener, but no worse than middle bell clashes on the strikeometer.

Whether a particular bell or group of bells has been rung to a much higher standard than (most) other teams (e.g. specific odd-struckness or audibility challenges have been overcome)

*

*

**

5-star replies:

R2. See above – a single “difficult bell” and how it is rung can be a key differentiator between other similar teams.

R7. ... but judges are told not to comment on particular bells, good or bad.

R8. Achieving this will ensure that other metrics land. It is a mechanism of achieving results.

3-star replies:

R3. We have had a similar discussion about this recently and the conclusion was that it wouldn't be very well received by the teams. It was pointed out that if you can't single out an individual negatively, you probably shouldn't do this positively either. Keep it as a team event.

R6. In my book, not hugely relevant beyond its impact on the touch as a whole (i.e. don't elevate the whole band's performance merely because the "difficult" bell is rung well).

2-star replies:

R5. Can be useful in crafting comments (with sensitivity!) – less important than other metrics in coming to a judgement when ranking bands.

1-star replies:

R1. This is something for comments regarding a particular band's performance, but doesn't affect the overall mark (except, perhaps, in cases where a relative score is considered important – i.e. for score normalization purposes).

Quality over time – how quickly the touch settles and whether it continually improves

*

**

**

4-star replies:

R1. The overall impression of a piece of ringing and the 'definition' of a fault is certainly affected by how quickly it settles and whether it improves. Maintaining consistency in defining a fault is one of the hardest things to do as a (human) judge and, if the ringing improves over time, there's a tendency for fault counting to dry up. The opposite is true if the ringing starts well and deteriorates – the fault counting will tend to get more severe.

R2. Two different things here. How quickly the touch settles is key. The best bands settle immediately – it's usually obvious which band is Birmingham as the ringing settles immediately. Improving is important too (rather than getting worse towards the end!) but the best bands are simply good all the way through.

R5. Thinking a little differently – the wedding ringing / Sunday service ringing criteria is helpful here – settling quickly into a good rhythm and high-quality ringing would be what bands would want to achieve when ringing for Sunday service / for a wedding, so has to be an important criteria for defining good ringing.

R6. This metric seems to conflate two different things. Settling quickly is surely essential to a good overall performance (4-stars), but measuring improvement on its own is dangerous (2-stars) and may unjustly elevate a touch that, assessed as a whole, is no better than average.

R8. Quality is important and getting there quickly is important.

3-star replies:

R3. This is something that Judges can identify easily without the use of technology

How well bands cope with fault / error in a change (do they manage to quickly or immediately sustain the underlying rhythm?)

*

**

5-star replies:

R1. The best ringing is not affected by errors – i.e. they aren't compounded. The overall impression of ringing is dependent on the 'structure', or rhythmical framework, of the piece and the ability of any band to stick to this reduces faults.

R5. Important – bands which recover quickly and settle back into / are able to maintain the underlying rhythm will produce an overall much better effect.

4-star replies:

R8. Can help distinguish lower orders. I feel it is interesting this year the judges have decided to mark down mistakes, preferring persistent faults over mistakes.

R11. I would not rule out placing a team first if that team had fewer total errors and recovered from a mistake rapidly.

3-star replies:

R2. Same point as above – framework and underlying rhythm is key to a good touch. Marked as a 3 as the best bands don't really have material errors in changes from which to recover. I'd view this in the inverse – if a band cannot recover from a wobble then they will struggle to perform for the rest of the touch.

R3. Again this is something subjective that the Judges can easily deal with

2-star replies:

R6. Not beyond the fault itself. Much better not to have errors in the first place!

How well a particular row is rung (e.g. queens, or backgrounds)	*	***	****	**	*
<p>5-star replies: But only because all rows are important. It feels judges are increasingly commenting on ‘feature rows’ as it is a quick way of distinguishing touches and creating comments (due to pressures of time to develop the comments), this is creating the appearance to competitors that these rows are more important. When I have judged this has not actually been the case.</p> <p>4-star replies: R1. Special ‘moments’ in any piece of music should always be points of particular care and it’s right that judges should look for particular quality in certain rows in a piece (e.g. roll-ups – see above.). R9. The only reason this is “important” is because it is particularly noticeable!</p> <p>3-star replies: R6. Will inevitably have some impact. Human nature!</p> <p>2-star replies: R3. It would be something that you would comment on and listen out for as a feature of the ringing, but the reality is that it is of little significance. The approach to and away from notable changes is probably more significant. R5. Tricky one! These rows are very noticeable to judges – and often reflected on in comments – I would say overall though that the ‘feature rows’ should not be massively important in ranking bands and coming to an overall judgement. Teams who ring the feature rows very well are of course quite likely to have produced a really good piece of ringing, but that is not always necessarily the case!</p> <p>1-star replies: R2. No different from any row in my view – they are all marked the same!</p>					
<p><i>Additional metric – accuracy of method ringing: Method mistakes and recovery from mistakes</i> 4 out of 5. R5. Important in the sense that this can have a significant impact on the other metrics listed above – eg by knocking off the rhythm, unsettling the piece, or simply by introducing a large number of errors (whether measured by faults / inconsistency of inter-bell gap / SD)</p>					

* HawkEar, of course, treats all bells equally, but we are trying to understand whether the tenor, treble, or indeed any other bell is subconsciously perceived as more important than another to the ringer or human judge.

Comment from R9: I think you're right: the treble and tenor as instantly identifiable may well be subject to more intense scrutiny because of human nature. This is probably also true of a known "problem bell" in the circle.

2. How would you measure "overall impression" of a piece of ringing?

R1. I'm not sure 'overall impression' can be measured. It's about general aspects: consistency, style, structure, sense of purpose, general accuracy, improvement through the touch, no method mistakes (to pick a few).

R2. This is very subjective. I'd consider speed, consistency, accuracy (i.e. an errors) and overall framework. The biggest factor though is the framework and clear "beat" throughout, most frequently set by the back bells, and the quality of the way in which the bells work together in the touch.

R4. Team agreement on speed and pulse of the bell sounds, meaning that the rhythm is toe-tapping (or you could clap your hands to it like a metronome). So, the measure would be regularity of beat. Which for a human, is scientifically known to be pleasurable sound. Hence why people love pop music which they can dance to, or African drum beats which are such a strong part of culture.

R5. This is naturally a difficult thing to 'measure' – some metrics I would suggest are: consistency across each row and across the piece, the extent to which the ringing has a clear sense of purpose and plan (unity over rhythm between all the bells would be one aspect of this for example), an impression that the ringing is 'flowing', accuracy of placement of bells on a micro level (inter-bell gap) and a macro level (fitting into the overall rhythm). Good impression: strong rhythm, accurate striking from individual bells across each row, purposeful ringing flowing well. Not such a good impression: inconsistencies in rhythm / approach from different bells, bad blows/clips, method mistakes

R6. Inevitably, this is partly subjective and non-measurable. Beyond that, I think it is little more than a combination of the key points above; specifically, that the band is working together as a team, with a common purpose and a consistent shape to each change. I used to believe there was scope for "impression" to have a significant influence on the overall result, but over the years I have become less and less convinced that this is in fact the case. The most accurate touch wins every time!

R7. Was it enjoyable to listen to? Did the ringing flow and was everyone ringing to the same rhythm? Consistency and accuracy. No major errors (striking or method) and any minor ones quickly rectified.

R8. Hawkear! With judges giving human perception. If human measurement is required, should go back to marking numerical faults.

R9. Is it possible to "measure" an overall impression? Surely the phrase implies an artistic judgement which is not absolute but essentially subjective. In a clumsy attempt to answer the question, I would say something which sounds relaxed, unforced, consistent and is pleasing to the ear.

R10. I don't think it measurable, and that's the point. I guess many factors come into consideration, eg suitable speed for the bells, consistency of speed, lack of significant errors, or even those which just make me tut, ringing which exudes confidence. In the early days of the competition, Bristol usually came sixth out of eight, having turned in a performance which the churchyard judges liked but the judges, counting faults, marked down.

R11. Subjectively

3. How would you define a touch's "compass"? And is it important?

R1. Certainly important as this contributes strongly to some of the 'overall impression' aspects above: consistency, style, structure etc.. The 'compass' is the general rhythm of the piece and is set by the back bells – 9-12 (on 12) and 7-8 to a lesser extent. I like to hear these bells rung as though they're being rung by one ringer as a set rung perfectly together – this is the 'compass'.

R2. Generally set by the back bells – this is sort of pace but really framework and how this holds together throughout the touch.

R4. I'm not sure I understand what's meant by the word compass, or at least, I'm not sure it's a word I would typically use.

R5. Rhythm – sense of purpose / direction – overall framework. I've tended to not use this term in comments or in coming to judgements – feels like a bit of a cliché and just another way of judges saying the word 'rhythm' in comments without repeating themselves...

R6. Again, if I have understood correctly, this is really no more than a distillation of some of the key factors above; a consistent pace, accurate leading, not stringing out the back of the change etc. As such it isn't a separate metric in itself, but is a useful way of capturing and describing the overall impact of other key metrics.

R7. Changes a consistent, even length.

R8. Talked a lot about in comments – it is classic judging wording. I would say it refers to the general shape and consistency of the piece. A good compass tends to reflect a confident band who are ringing well together.

R9. My definition of compass is framework. To me it is of great importance, as consistency is the key factor: handstroke gap before leading, each inter-bell gap, backstroke leading and so forth. If all these have a large degree of accuracy the compass of the change is retained and good rhythm is achieved.

R10. Right speed for the bells – ringing is an artistic performance, after all – and consistency of speed.

R11. A perception of a collective view on what a team is trying to achieve, demonstrated by a notion of consistency throughout the piece.

4. If you were comparing a quick piece of ringing to a slow piece of ringing, how would you discriminate between the two if both were very accurate?

R1. I think this is very difficult indeed and is one of the hardest problems to solve as a 12-bell judge. Accurate, quick ringing is exciting, vibrant and technically impressive. Accurate slower ringing is majestic and, perhaps, more stylish. In the absence of strikeometer information, I think this boils down to a matter of personal preference.

R2. Typically, the quick piece will sound better than it actually is on the strikeometer. Generally, ringing which is too slow is less pleasant on the ear unless it's very, very good. Overall though, the judges job is to measure the quality of the striking, using both manually methods and the strikeometer, and this should come first over a judgemental view of speed. Generally, anything that is too fast or too slow will fall down, though where those thresholds lie is judgemental.

R4. I would start no differently from any other comparison: listening for team agreement on pulse, beat, consistency of lead. And then beyond that, there's something about a team's sense of purpose kicks in. That's hard to describe. But you can usually tell when a band has deliberately set out to achieve a certain style (ie, they have a sense of purpose) because they start like that from the very beginning and stick consistently all the way through. It doesn't matter whether the handstroke gap is larger or smaller, but is clearly been pre-planned and managed that way. And that tends to produce more purposeful ringing which I like. So, I suppose I'm saying that the speed doesn't matter. What does is the teamwork.

R5. When two pieces are both very accurate (or indeed just very close to each other, wherever in the overall ranking), this is the stage at which Hawkear comes into its own as it is otherwise very hard for judges to maintain consistency in their judgements, with some of the problems described by Liz Orme in her RW series on the 12-bell (eg sequence effects where judgements are impacted by what came before/after) coming into play. I would look carefully at the data from Hawkear to differentiate between the two pieces, as well as using all the available metrics listed above to discriminate between them. It's important judges aren't influenced by personal preferences over speed – but difficult to achieve in practice!

R6. It depends! For example, on a light and "pretty" ring, fast ringing may sound exhilarating and slow ringing stilted. On the other hand, on a heavy and stately ring, a slow pace may confer a grandeur that is lost at a brisker speed. In the end, I don't think it's possible to be prescriptive or to give a "one size fits all" answer.

R7. Difficult, but very much as 2 above. The problem is that, to a human listener, faults in slow ringing can seem to be more obvious. Quick ringing can, sometimes, cover up minor faults. But if the overall effect is good ...?

R8. Ringing tends to sound better if the speed is appropriate to the weight of the bells, as it is easier to get good striking if not going too slow or too quick. However, accurate ringing is successful regardless of speed.

R9. I would try not to. Speed should not be judged in this context. I think the answer would be a personal one: my discrimination would be based on what I preferred, what sounded pleasing.

R10. I suppose it might depend on how far apart the speeds were. I think there's probably a tendency to mark slower ringing more harshly than quicker ringing, as there's more space between the blows to assess (or measure, if you prefer). The slower the ringing, the more difficult it becomes to maintain the compass, particularly when you get to a speed which too slow for the comfortable range of speed for the bells.

R11. By counting faults.

5. When looking at HawkEar output, what do you find most useful? And what would you like to see that's not there?

R1. I look at the overall absolute and relative (speed adjusted) error. It would be good if HawkEar had knowledge of the touch as I think this would help it disentangle some of the weaker teams, but it wouldn't affect the top teams' placings. Does it do fault counts now too? If not, this would be useful.

R2. Individual bells SDs, number of material errors above a threshold by bell, how this varies between leads. Graphical analysis of certain parts of the touch, particularly for rocky patches. Both pace adjustment and non pace adjusted SDs are helpful. The overall metric measuring the SD for the touch is a little useful as a comparison between bands but should be used blindly to rank as it misses various other subtleties in the ringing (pace, errors, framework etc.)

R4. As a judge, I look at the grid of rows to see where clusters of issues are consistently happening. That often helps to validate what I think I've heard, and can sometimes be referenced in feedback/comments. This is the most useful output in my opinion. And secondly, I use one of the scores to help with the ranking of each team. But I treat the score as another judge on the judging panel. I don't believe in tech-driven score out-weighting the other judges' scores; it's an equal contributor. In the last 12-bell I judged, there were three different scores outputted for the rank of each team, and they didn't agree. The fact that they disagreed with each other deprioritises its reliability and hence importance frankly. We have to remember, that the final scores announced come as subjective %s based on ranking, not an accuracy score. I obviously get the data of peal speed. And finally, I sometimes drill down into one bell's ringing, if I want to pick out something complementary to say as part of feedback. What would I like to see: a score which I can rely on, rather than one which conflicts with others. And less agonising over analysis and data. It's a team-based musical performance at the end of the day. Much like a symphony orchestra, string quartet or jazz band. These can all go at different speeds and vary in their approach to accelerando/rallentando etc. But what makes it good is energy put into consistency, togetherness and agreement about how they aim to perform.

R5. Overall SD score. Data on individual bells if we need to drill down a bit. The blueline-style output for comparing to judges' notes on the test piece – eg is HawkEar picking up the same good features / less good features we did – useful for visualising, and also for jumping in to sections of the touch for relistening if needed. Peal speed data can be useful at times.

R6. I'm not really qualified to comment on this one, as I don't sufficiently understand what the various outputs mean. I am told that SD is the best guide to "accurate" striking, so I always look at that. I also tend to notice the number of major errors and whether individual bells are marked out as consistently fast or slow. But, for me, the most useful tool is to listen back to a touch while following the grid, comparing what I see with what I hear and noticing trends as I go.

R7. I have relied very much on the Strikeometrist providing, where possible, evidence or information which helps the judges answer, clarify or confirm the issues they are discussing, eg "Team C seemed to start well, had a bad patch in the third lead but picked up in the fourth and continued to improve" or "The back bells in Team D rang together better than those of Team F".

R9. I fear I don't understand HawkEar well enough to make a constructive reply to this question. With my limited knowledge what I find most useful is SD and the charts showing individual bell errors at each stroke.

R10. I haven't used HawkEar for so long that I don't have an opinion. It's probably changed a lot since I've used it.

R11. It's a long time since I looked at this closely. I used to focus on the overall ranking rather than the detail, and tension that against the judges' scores and ranking. If and where there were differences, then I found the histograms of handstroke/backstroke scores helpful.

6. Would you like to see fault counts or standard deviations?

R1. Both would be useful.

R2. Both! For fault counts, it's useful to see these at different thresholds for what constitutes a fault e.g. showing both 30ms and 50ms deviations is insightful. SDs are important but at the more granular individual bell / touch segment level as the "overall score" consolidates too much individual underlying information.

R4. The scoring systems were fault count, SD, and C-Count when I last judged. So 3 different systems. I found Fault and C count matched my scores more than SD.

R5. Both please! It is useful for judges to have access to as much data as possible – otherwise we get into quite a tricky debate about which measures should be presented to judges to support their judgements. It can be the case that different measures present the judges with differing results, but part of the job of human judges is to work through that.

R6. I'm sure both have their place, although it depends what you mean by "fault" (the very question begged by this survey!).

R7. Both, with appropriate caveats.

R9. I'm not sure I understand the question. I rely upon fault counts, as I tend to have an old-fashioned approach to judging. The SD figures of HawkEar are useful, as already stated.

R10. I've always counted faults, and then tuned the overall results into percentages, by various magic formulae (sometimes black magic). I wonder if the average punter listening to the results would understand results conveyed as standard deviations.

R11. Fault counts. Unless we know how the SDs are calculated, it's difficult to say more.

7. How do you assess good ringing?

R1. I assess it objectively using fault counts and HawkEar measurements, and subjectively against the criteria listed in Q2 about 'overall impression'. I would always mention points of excellence and points to improve.

R2. All of the above! Key is that many different metrics flow in which, whilst themselves objective, need to be subjectively assessed in order to form a view. I think that the metrics available from the strikeometer are a very useful aid but are not conclusive (for example, different metrics rank in different orders). They are most useful for drilling into specific attributes of the ringing and ensuring consistency in "judging harshness" over the period of the long competition (otherwise the ear definitely becomes less critical as one gets tired!)

R4. See my answer to qn 4 and 5.

R5. Overall quality of the rhythm, effectiveness of the band working together to a common goal, faults/errors/method mistakes, quality of leading, interbell-gap, (these next two are quite subjective but...) 'am I enjoying listening to this ringing?', 'what would I say to this band if this ringing was a peal I was conducting?'

R6. See all my answers above!

R7. Quoting (from memory) from some past notes for judges: It should be virtually faultless, exciting to listen to and have a sense of purpose. You know it when you hear it!

R9. The answer is essentially the same as in question 2. Perhaps I would add accuracy and a "felt pulse" – a relentless rhythm which first of all captures ones attention in manner where less good ringing fails, and goes on to almost hypnotize you. Does that make sense?

R10. Ringing which is pleasing to listen to, taking into account appropriate speed for the bells, absence of significant mistakes and consistency of striking.

R11. I know it when I hear it.

Comments received in email replies:

R1. Many thanks for asking me to participate in this survey. It was thought-provoking!

R5. It was really interesting to complete and made me think carefully about judging going forward - I'll be really interested in the results of the survey, thank you for your work on putting this all together.

R6. It's been an interesting exercise and has certainly made me stop and think. Whether I've found the right words to describe what I actually believe is a different matter! I'll be really interested to see the combined results in due course. I know many people have strong opinions on the subjects raised and I suspect you will have fun trying to draw out common themes and extrapolate meaningful and useful conclusions from them. Good luck!

R7. I found some of the questions quite awkward, partly because just about all the metrics are very important and I felt I was simply saying the same thing over again, and partly because so much of what you do when judging is intuitive and difficult to analyse in detail.

R11. Please find attached my response, which, as you will see, is incomplete. The reason it's incomplete is that I can't honestly answer the first set of questions without knowing what is meant by "average", "consistency", and "deviation". As Stuart remarks in his excellent notes on the Kalman filter approach, "anyone.....has to choose an underlying model to calculate the 'ideal' in order to be able to measure deviations from it." So what is the choice? I guess this is, in part, what the survey is trying to address, but in order to do this, surely one has to pose questions without using the very terms one is trying to define!? I am not trying to be awkward here, because I think the survey is a worthy exercise, but care has to be taken when interpreting the responses, in light of the issue I raise above. I shall be interested to know whether the results of the survey will lead to a modification of the state vector, and/or the model description matrix, and/or the measurement matrix – eqns (1), (2) and (3) in Stuart's notes. And/or if any of the assumptions used in defining the initial conditions should be changed – e.g. the prescription of the handstroke gap at lead. If the current model uses those definitions as given by Stuart, which are surely a reasonable zeroth-order idealisation of "good ringing", then isn't there a risk of introducing a bias towards a style of ringing preferred by some bands if those definitions are changed? (I may be out of touch here, if the model used has changed radically since Stuart wrote his notes, but I suspect it hasn't.) It is my view that subtle changes to inter-bell spacing, and handstroke leading, can result in ringing that is a pleasure to listen to, which is why human subjectivity is important (and, as I've said many times, do you really want to hone your skills as measured by an algorithm that would put a pianola ahead of Andre Previn?). Of course, I could be barking up entirely the wrong tree here and there's no intention to modify the basic model, and indeed my own view is that the model is best left unchanged. It might be more helpful to produce a set of guidelines for judges to work to, and to emphasize the strengths and weaknesses of the strikeometer. For me, the strikeometer is a very useful tool for calibrating judges's scores and decisions. It should not replace human input (cf. the approach taken at the Final at Aston in 2016!), rather, as the initial pioneers intended, it helps to reduce the impact of human fallibility that is inevitable when concentrating intensely for a prolonged period of time. It is a long time since I looked at the data from contests, but when I did this (with Stuart's help in 2012), we found that the level of accuracy of ringing in the contests was such that the distributions were, at worst, "skewed-Gaussians". Having observed this, it would be interesting to compare, for example, the data from the Final at Southwark with that of the Final at Guildford. I certainly look forward to hearing about how you progress with this initiative.