

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

Summary of Survey Results

Participants

29 people participated in the survey, including 7 Chief Judges from recent finals, 4 other experienced judges with many finals between them, and 18 ringers who ring in teams regularly in the final. The quality of replies was excellent, with very many thoughtful and detailed replies and an average of 600 words on each returned survey. Many people had lots to say, which was excellent!

Comments on the survey design

Overall, the survey was received well. Three judges commented “The survey was thought-provoking”; “Makes me think carefully about judging going forward”; “An interesting exercise, makes me stop and think”. But it must be recognised that this is a difficult subject area to get your head around, and it was a difficult task for us to design the survey questions. A few people commented that they struggled with the survey: “The questions were quite awkward, difficult to answer”. Perhaps it is because, as one person commented, “Judging is intuitive and difficult to analyse in detail”.

“Working together as a team”

Working together as a team is arguably the top theme coming out of this survey and is a phrase we may well want to make good use of. Most respondents marked this as very important. “Team’s sense of purpose, when a team has set out to achieve a certain style”; “How quickly the touch settles is key; “Ringing regularly together helps this”; “This also applies to big bells and small bells working together”; “And indeed any other combination of bells”.

One respondent thought of the “Mental image of a flock of birds such as a starling murmuration and how they flex and move relative to one another”.

Consistency

Importance of consistency of leading. Erratic leading is at the root of many problems. Consistency of tenor: “Plays a special role”. “Best pieces have even length rows. Dragging out or dropping in destabilises the rhythm.” Consistency is important – if speed changes over the touch, OK as long as everyone does it together (hard to do).

Speed

Most respondents realise speed is a difficult subject area. There’s nothing wrong with fast ringing or slow ringing: speed in itself isn’t important (apart from extreme speeds) – but will certainly influence the quality of the ringing. Most respondents said the speed of ringing should be appropriate to the weight of the bells, and appropriate to the band.

Rhythm

How rhythmical the ringing is, is important. Rhythm is all bells ringing to a similar plan. Back bells generating a good structure for others to follow.

Feature rows

Feature rows are no more important than any other row. However, some judges commented that it is an easy row to comment on in the results. Some ringers commented that they’d prefer judges to comment on a section of ringing leading up to a specific change, rather than just the feature row itself.

Overall impression / Musical performance

Overall impression is key, but very subjective - how do you measure it? Many respondents commented “we know good ringing when we hear it” but they found it much harder to tangibly define this. Overall impression is subjective and qualitative.

Quite a few respondents spoke about ringing being an artistic or musical performance. This is an area that requires further exploration. Are we trying to ring in a fault-free, very regular style, or is ringing meant to be a musical performance with musical flair and style considered as part of the judging? If the latter, how is it defined so that people know what to aim for? (one judge commented that there used to be scope for “impression” to have a significant influence on the overall result, but much less so now - the most accurate touch wins every time).

This is a tricky area, which needs further discussion.

Overall assessing good ringing

Most respondents commented that assessing good ringing was a combination of all the questions / factors in the survey, but most importantly: quality of rhythm, relentless rhythm, band working together with a sense of purpose. Good ringing: solid positive rhythm, relentless rhythm.

Many respondents said: “Ringing which is pleasing to listen to”; “I know it when I hear it”. The idea of this survey was to try and tangibly describe this – some respondents have had a good go at doing this, but some respondents had difficulty tangibly describing exactly what they like.

Strong rhythm, accurate striking, purposeful ringing flowing well is probably the best overall description.

Jargon

There is a plea from some respondents to get rid of the jargon, and some phrases which are not particularly well understood. The phrase “touch’s compass” is not understood by many perhaps it is time to drop this phrase and come up with a more meaningful / understood phrase.

Many people struggle to understand the meaning of standard deviation / SD scores.

There is also a suggestion to come up with a common set of word/phrases to be used by judges when giving results.

HawkEar

HawkEar is used by judges to help validate what they have heard, and to answer, clarify or confirm the issues judges are discussing. Currently, fault counts and SD scores are both valued as useful. More transparency is needed around HawkEar: publish metrics that are being used and explain any differences between judges and HawkEar.

There is probably a need for further work with HawkEar software writers to discuss additional output(s) to further assist judges and ringers.

* * * * *

The following pages give more detail of the replies for each question in the survey.

1 | What are the key metrics you'd look for in defining good ringing, and how important are they?

Summary of star ratings

Respondents ranked each feature (the names of which are shortened) on a 5 point scale:

Not very important, less important, doesn't massively matter, **important** or **very important**.

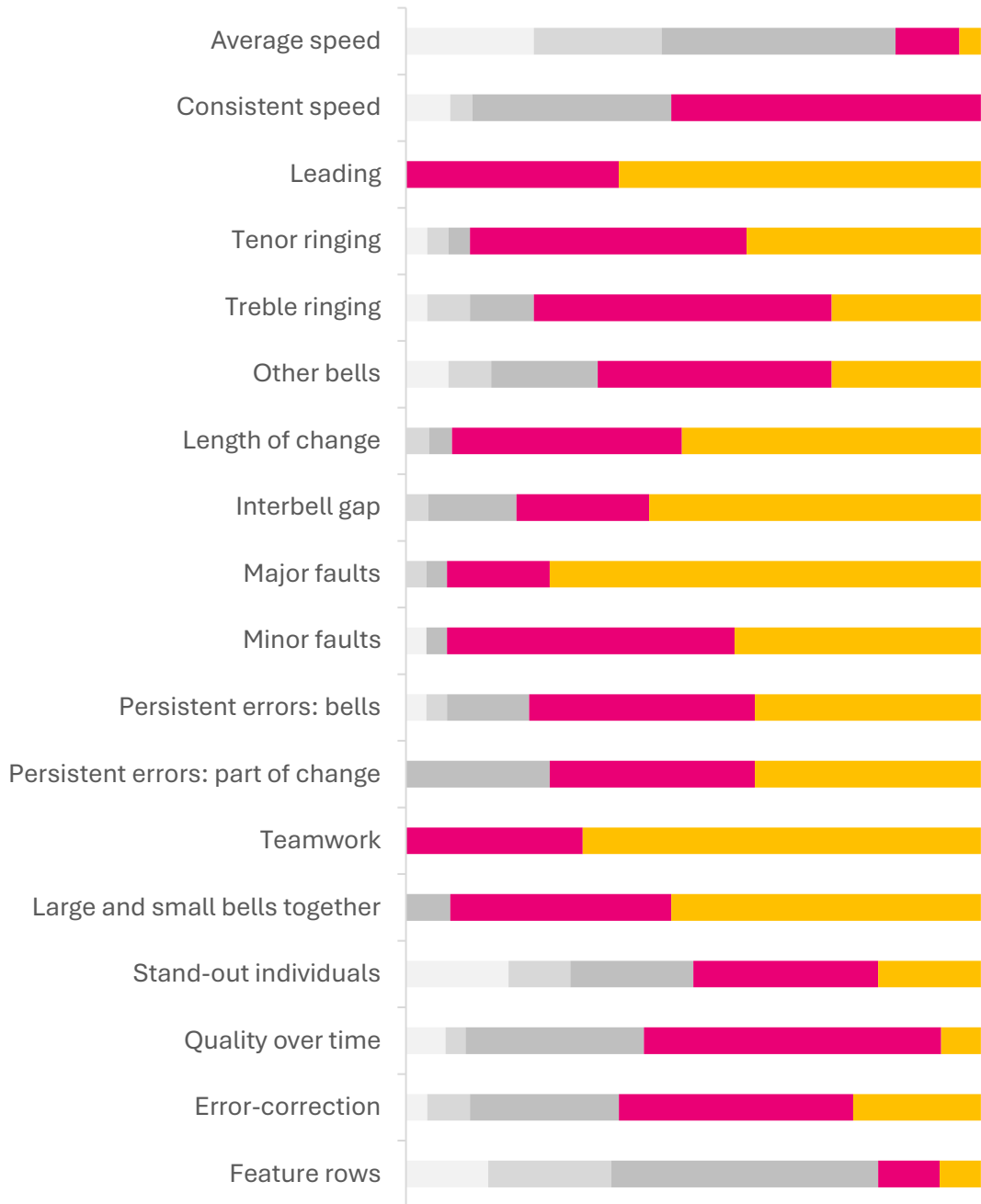


Figure 1: Importance of different features of ringing.

There were very few categories which were felt to be unimportant.

Overall ranking of importance

If the score (1 – 5) given to each feature is averaged across all respondents, some clusters emerge, which are categorised in the table below. Of course, these are only averages, and some respondents did attach a lot of value to features lower down in the list; for example, two people said that the **quality of a feature row** was very important.

	All respondents
Very important average score > 4.6	Teamwork Major faults Leading
Important average score > 4.3	Large and small bells Length of change Inter-bell gap
Quite important average score > 4	Minor faults Tenor ringing Persistent errors: part of change Persistent errors: bells
Some importance average score > 3.7	Treble ringing Other bells Error-correction
Less important average score > 3	Quality over time Consistent speed Stand-out individuals
Least important average score < 3	Feature rows Average speed

Comments on individual features

The comments on some of these individual features are now considered, in the order of those thought to be most important.

Teamwork How well the bells (or groups of bells) work together as a team

“This is key to producing good ringing!”

“What we are all striving for!”

However, there is less consensus in the comments about why this is so important, or how it should be measured.

- Several comments touched on the idea that a band ringing as a team will be able to produce a more consistent piece of ringing, where errors or imperfections do not propagate.
- Some associated teamwork with the ability of a band to sustain the rhythm or framework of the piece.
- Some people spoke more in terms of musical or stylistic attributes. Feature rows, such as roll-ups, were mentioned repeatedly.
- One comment said that bands that ring together regularly ‘just know’ where other bells are going to be, calling it “a genuine intangible”. Another asked how it might be objectively measured from the sound produced.

Leading Consistency of leading (with a handstroke gap), denoted by a standard deviation

This is important in defining the ringing and making sure the rhythm across the rest of the change is even. Inconsistent leading is also very audible in itself. All bells should leave the same handstroke gap, ideally creating a 25-bell rhythm.

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

“The best pieces have even length rows which remain consistent throughout the touch.”

Most respondents agree that if bells are dragging out a row, or dropping into lead, then it is hard to maintain a strong rhythm.

Inter-bell gap Consistency in maintaining the same inter-bell gap throughout the change

“This is exactly what good striking is.”

“OK to change gradually through the touch (as this changing is clearly the mechanism for the speed changing)”

“There are issues with using HawkEar to create metrics based on the inter-bell gap – if this is used to mark faults the quantization means that at certain speeds there is a step-change in the size of the error that will be counted as a fault”

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

“Almost always differentiates the best teams in my experience.”

“It would be a mistake to single out the consistency of the tenor (or any other individual bell) as being more important per se in judging the quality of the sound”

Opinion is split on whether the tenor is actually more important than any other bells, especially in even-bell methods.

- Some people feel strongly that the tenor plays an important role in setting the rhythm, pace and style, and inaccuracies make it harder for the rest of the band to establish a good rhythm.
- Others think the tenor is no more important than the rest of the band but is often commented on by judges because it is easiest to hear, and so it is more obvious if it is being rung badly. However, some people argued that because it was an easy bell to pick out, other ringers are also more likely to be distracted by inconsistent tenor ringing.

Treble ringing Consistency of treble ringing, denoted by a standard deviation

“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.”

Although most felt that the treble is not inherently more important to good ringing, in some methods other ringers are likely to react to a poorly rung treble. The treble’s role in setting the handstroke gap during opening rounds is important, and it is very audible at the lead-ends.

Other bells Consistency of not the treble or tenor (i.e. another bell), denoted by a standard deviation

“All bells need to be rung well for a strong touch”

“It’s not that it doesn’t matter, but perhaps bells other than 1 or 12 are less noticeable”

“Need enough rhythm-setters in the band to build and preserve the framework”

2 | How would you measure “overall impression” of a piece of ringing?

The most common answer was that it was impossible to do this objectively. However, lots of respondents attempted to define what they would look for, with some detailed responses. On the whole, answers here were a combination of features already mentioned in question 1.

The most commonly mentioned features are shown in these word clouds.



Figure 2: Features leading to a good overall impression



Figure 3: Features leading to a poor overall impression

Others used more subjective measures: “Is it nice to listen to? Would I like to take part in ringing like that?” Several people said that it was easier to compare teams than score them on this measure.

One person thought that overall impression should prioritise the best bits of ringing: “I think that ‘how good are the good / average bits’ is more important than ‘how bad are the worst few bits’.”

3 | How would you define a touch’s “compass”? And is it important?

“My mental image of this is of a flock of birds (such as a starling murmuration) and how they flex and move relative to one another.”

“A perception of a collective view on what a team is trying to achieve”

A vast majority of replies mentioned a consistent speed and rhythm, and often related this to the framework. However, seven of the respondents didn’t know what was meant by compass, and one thought it was a judging cliché. One person defined it as how the quality of ringing evolves over time.

Compass is generally thought to be very important (by those who were able to define it) but only a few commented on how they would measure it. This was usually through a combination of length of change, leading and inter-bell gap. Back bells were felt to be especially important, and so a measure of how these bells were rung could be used.

“A piece of ringing with a strong compass will have a consistent speed with front, middle and back bells in agreement”

“I like to hear [back bells] rung as though they’re being rung by one ringer, as a set rung perfectly together – this is the ‘compass’.”

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?

“Accurate, quick ringing is exciting, vibrant and technically impressive. Accurate slower ringing is majestic and, perhaps, more stylish.”

“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.”

Three answers came up repeatedly in this section.

- About a third of respondents said that they wouldn’t discriminate, or that they would try not to, or that they would rely on other evidence to make a decision.
- About a quarter of respondents would rely on their overall impression of the piece of ringing, although several acknowledged that this introduced more subjectivity into the judging process, and one person spoke of the need to avoid various forms of bias. Some people named specific things they would look for, such as the bands sense of purpose mentioned in the quote above.
- Five people said that their answer would depend on the bells – for example, slower ringing sounds better on deep-toned bells with lots of hum.

Two people felt that quicker ringing often sounds more pleasant, although not at the extremes.

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

A few respondents expressed some reservations about their understanding of HawkEar outputs.

Most useful output: the question about what is useful was answered from 3 different perspectives, and many respondents included more than one perspective in their answer.

Most useful for improving own ringing

Output	Number of mentions	Notes
Touch viewer (coloured visual representation of the touch)	6	One person specifically mentioned listening back alongside, and another mentioned the error bars that appear on the right when looking at an individual bell. Use of colour is appreciated.
Bell error histograms	5	The average error for each bell at hand and back seems to be most popular.
Individual standard deviations	3	Two comments implied they always look at this first but don't think it actually helps them improve.
Bell error tables	2	
Count of errors over 50ms	2	
Individual average errors	2	
Star rating	1	

Most useful as a team captain or for the band as a whole

Output	Number of mentions	Notes
Bell error tables	2	Good for picking up structural issues to improve performance.
Striking summary	2	Used to measure progress as a band.
Individual standard deviations	1	Useful for selecting the team.
Touch viewer	1	Used to analyse the touch lead by lead and look for groups of bells not working well together.
Count of errors over 50ms	1	Used to measure progress as a band.
Star rating	1	

Most useful for judging

Output	Number of mentions	Notes
Overall SD for the band	5	Can be used to compare teams, although it can miss subtleties.
Touch viewer	4	Used to validate what judges have heard, spot trends and identify rocky patches. Useful when writing comments. Also used to find sections of the touch to relisten to.
Bell error histograms	3	
Individual bell errors and SDs	3	Can drill down into how an individual bell has been rung. Can be useful when writing comments.
Count of errors over 50ms	3	
Speed-adjusted SD	2	Also referred to as relative error
Peal speed	1	

Suggested areas for development

Several new features were suggested:

- Ability to analyse by types of work, e.g. dodges, places, hunting – the most popular request, mentioned by 3 people.
- Clearer indication of which parts of the touch (e.g. which leads) generated most errors
- A way to jump to each bell's largest errors in the touch viewer
- A metric for in-change evenness of feature rows
- Measures based on median and percentiles rather than mean and standard deviation
- Measurement of mean absolute error
- Visual representation of compass (“once defined!”)
- Although not mentioned in this section, in responses to other questions more than one person has asked for the ability to change the threshold for an error, or to introduce a graduated scale of errors to avoid a cliff-edge.

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

“Both are very valuable. For improving ringing SD is probably most helpful. For judging ringing, I think fault counts, and how those faults are scored, is most important.”

10 people wanted to see both, 12 preferred fault counts and no one just wanted standard deviation. A few people wanted neither, or did not express an obvious preference.

A number of comments said that fault counts were more easily understood, and therefore better for conveying results in a contest.

Three people wanted to be able to change the error threshold for a fault or display more than one threshold. One person wanted to introduce a point system where larger errors counted more.

7 | How do you assess good ringing?

Most respondents used this question as an opportunity to describe what the best ringing sounds like, and answers were very similar to question 2.



Figure 4: What people listen for when assessing good ringing

Most respondents said that they would take both Hawkear metrics and their overall impression of the ringing into account.

Several people mentioned perceived drawbacks with relying exclusively on Hawkear, or with ignoring it completely. There was a feeling that the relationship between Hawkear output and judges' decisions is not always transparent at the 12 bell contest, and some people would like to see more explanation of how judges have used Hawkear, and especially when they have changed the ranking it gives. Several people also commented that consistent fault counting or unbiased comparisons were very hard to do as a human judge alone.

8 | Additional comments

A large number of respondents said that completing the survey had been thought-provoking, and a few apologised for not being able to find the right words to explain their intuitive sense of good ringing.

“Perhaps the two most important limitations in the current system could be expressed as two issues from recent contests: (1) ‘I had to deliberately ring my bell too close for the strikeometer’ (2) ‘How did they come third with those scores?’”

“It would be interesting to manufacture two pieces of ringing electronically which achieved the same Hawkear score but in different ways. Or one piece of ringing that had everything perfect but a sprinkling of errors over 30ms which judges pick up, and then another piece which Hawkear might fault more heavily because there is lots of inconsistency but no single error is over say 25ms and hence could not be detected. Which would sound better or worse? Which should we think is the better piece of ringing?”

“I'd like to plead for as little jargon to be used as possible in communications about [Hawkear]. I'm sure it's done in complete innocence, but for those of us totally outside this area even using the word metric to head a column rather than something in more general use like criteria is intimidating”

What Next?

The working group will consider the results of the survey and provide recommendations to the 12-bell committee. In particular, we will consider producing some brief guides to assist the training of future judges and to assist teams with their contest practising. We will also use the survey results to decide if further HawkEar analysis software can be developed to produce additional output metrics based on this latest knowledge, and how this can be presented to judges.

The Working Group: Neil Buswell, David Pipe, Chris Poole, and Stef Warboys

*Plus a special thank you to **Catriona Agg**, who analysed the results and produced much of this summary document.*