Admap

November 2006

A Bewilderment of Meters

By Paul Kennedy, Research Director, RAJAR

Unedited Version

Audiometers (small portable electronic creatures that monitor exposure to broadcast audio) have always told the truth. I remember vividly the first time I carried an audiometer. At the same time I kept a log of my listening to check it against, and I still remember the sense of anticipation during the week or so that I had to wait for my data to be sent back to me, and the barely containable excitement when it finally arrived.

Could that little device really register my listening? I hadn't tried to confuse it by station-hopping - that would have felt like cheating - so I had stuck rigidly to my normal radio routine, and there it was, looking resplendent in 10-point Courier: entry no. 1, Monday, 07.41 – 07.56, BBC Radio 5. My log said 7.39 - 7.55, but what the heck – that was a formality. Meters worked! At that moment I knew that radio audience measurement was about to change for ever. The grail was within reach. The wishful past had collided with the fulfillable future and life would never be the same again.

It lasted all of two minutes, which was how long it took to spot the Sunrise Radio entry – Wednesday, 18.17 - 18.21. But I hadn't listened to Sunrise Radio. I racked my brain – had I been in a shop, or a taxi, or anywhere that I might be exposed to this station. No, I was fairly certain I'd been on the underground at that time. But if the meter got Radio 5 right, then why would it get Sunrise Radio wrong? Must have been my mistake. Sorry. Back to the celebrations.

Then I noticed an entry for Talksport, a station that I hadn't listened to that week, and a couple of references to BBC World Service, which wasn't in my repertoire, and clouds began to gather. And the rain really started when I realised that Classic FM was missing completely. I could have been wrong about one of those disparities, but not so many. Parade cancelled.

Subsequent and (you will be pleased to hear) much more objective testing confirmed my examples as being symptomatic of general experience, and that was the trouble with those early meters. They confused us by telling us the truth, but embellishing it with a sprinkling of mismatches – either short episodes of listening to stations we hadn't listened to (false positives), or missing out listening that had taken place (false negatives) – derivatives of the limitations of the technology. In other words, they told us the truth, but not the whole truth, and certainly not nothing but the truth.

As the industry body responsible for managing the UK's Radio Audience Measurement service, it is RAJAR's job is to ensure the UK radio industry (the fourth largest radio market in the world, worth around £1,000 million p.a.) has the finest measurement service it can afford.

Recently RAJAR awarded a new contract, thus reaching the culmination of years of planning; a process that included an industry-wide consultation on what kind of service was needed as UK radio moves into an unprecedented period of expansion and escalation across stations, brands and delivery platforms.

If the findings of that consultation could be distilled into a single sentence, it would read something like (draw a deep breath):

All subscribing stations to be measured in a single survey, regardless of their size, location, ownership or method of delivery, using a robust definition of "listening", with faster data turnaround and platform-specific reporting, at an affordable cost, to include electronic data capture as and when it is proven to be feasible.

The use of the word "feasible" is crucial. It means something that is possible, plausible, fit for purpose, and capable of being done within reasonable constraints of time, economics and technology.

Many people have considered audiometers to have been feasible for the past few years, as the manufacturers accentuated the positives in their messages, and industry bodies such as RAJAR were frequently labelled as enemies of promise for looking these gift horses in the mouth. But thank goodness we did. Otherwise we might never have realised that the small print was missing – caveats like "* recognises almost all music formats except classical", "* tends to inflate reach figures for speech-based services", or "* works pretty well if the listener sits still and avoids wearing nylon".

Happily, audiometers survived the bad old days and advanced, as did RAJAR's interest in the devices, and over the course of 3 years and a £2 million investment in testing, the UK's search for the feasible audiometer saw seven different devices being investigated (alongside several others that never made it off the drawing board). By 2005, when the final round of tests was taking place, the number of candidates had been whittled down to three, each of whom responded to RAJAR's specification for a new contract, those being the Eurisko Media Monitor (EMM), the Ipsos MediaCell and the Arbitron Portable People Meter (PPM).

Eurisko Media Monitor

The Eurisko Media Monitor uses audiomatching, whereby a monitoring station records everything that is output on every surveyed station and stores it. Meanwhile the device is worn by a person, and takes samples of ambient sound which are then compared to the recordings from all of the stations at that precise time, in much the same way that a fingerprint can be used to identify a person.

The EMM performed with startlingly accuracy in RAJAR's validation tests. Unfortunately, the fieldwork trials were hampered by logistical problems meaning that it was impossible to draw firm conclusions. Eurisko needed more time to prove their device could work; unfortunately, there was no more time available. RAJAR expects that the Eurisko meter will prove itself over the next year or so and can be a genuine contender for the next RAJAR contract.

Ipsos MediaCell

The IPSOS MediaCell device is quite possibly the perfect technical solution, as it uses people's existing mobile phones and converts them into audiometers using a software upgrade, very much like a SIM update. Ultimately it promises huge (as in HUGE) sample sizes very cheaply, with the option to have near real-time ratings. Unfortunately for Ipsos, their breakthrough has also come too late for RAJAR this time around, but as with Eurisko, we fully expect them to be serious contenders for the next contract. Like our third candidate, Arbitron's Portable People Meter (or PPM), the Ipsos MediaCell uses encoding rather than audiomatching.

Arbitron PPM

There are several types of encoding, but all of them follow much the same principles. Two elements are required, the first being the insertion of a unique digital code into the broadcast chain via a box at the station itself. This carries information about station ID, platform, and date and time of transmission. The code is inaudible.

The second element is the meter that is worn or carried by the respondent. In the case of Ipsos, this is, of course, a phone. The PPM, on the other hand, is a pager-type device, deliberately designed to be plain. It contains a microphone and a motion sensor. Every few seconds it searches for codes and stores any that it finds. At the end of each day the device is placed in a docking station to download the data and recharge the battery. Information from the motion sensor is retrieved at the same time and this is used to determine whether an individual's compliance is satisfactory to include or reject them.

Codes stored in the PPM are then compared to those known to have been broadcast during the day in question and any matches are stored as listening.

Outcome

The outcome of RAJAR's audiometer test programme was the commissioning of a 2-year research & development panel using the Arbitron PPM. RAJAR chose the PPM because it was the best meter that was available at the time, and the only one that had run the full gauntlet of testing. However, this panel will not connect with the overall currency, which will remain a diary-based system.

Why? Well, having evaluated audiometers in the field we were left with a number of issues that needed to be resolved before we could commit to their full-scale implementation. These concern compliance (with particular reference to breakfast-time), the creation and application of Editing Rules, and data differences.

Compliance

The RAJAR tests have shown that audiometers are capable of delivering a reasonable interpretation of a person's radio exposure; however, the level of accuracy depends upon the conditions under which the respondent is exposed and their diligence in carrying the device throughout the waking day. In other words, while audiometers have made vast strides towards telling nothing but the truth, they still do not necessarily tell the whole truth.

There is an urgent need for knowledge concerning general compliance, and overall participation and churn. Because audiometer measurement happens in real time and exposure cannot be captured retrospectively, only a minority of respondents will deliver a complete set of usable data day after day, while a majority will provide data on most days, but not all, and even on those days when they do provide data, that data itself is an incomplete record because they do not carry the meter all of the time.

RAJAR's tests show that when the PPM is used as part of a panel, between 70-80% of the sample will deliver at least one days' data per week, and the average reporter will provide data for 66% of his or her time. In short, RAJAR can expect the entire installed sample to be compliant 50% of the time.

Breakfast Time

While there are several aspects of compliance that we need to know more about, one of the most critical is what happens in households operating to a deadline on weekday mornings. Despite drilling home the importance of good panel behaviour, is it realistic to expect people to put the audiometer before their own domestic pressures? If it is confirmed that listening is being "lost" at the breakfast peak period due to poor compliance, RAJAR must look to improve panel member behaviour, as well as investigating techniques to recover any lost listening through editing routines such as "bridging".

Editing rules

Editing rules themselves play a massive part in determining what is included as "listening", starting with how a minute is credited. Should it be 60 seconds of exposure, or 30 seconds, or even 4 seconds – as with the early meters? What should be done with short gaps in the listening records that are due to a failure on the part of the meter to pick up codes? RAJAR found that even in good listening conditions, a meter may miss up to 10% of exposure

because the person is moving, or the type of audio being broadcast is not spectrally rich enough to sustain sufficient codes. And what should be done with "listening" that is captured while the meter is not in motion? How these questions are answered will exercise a dramatic influence on the reported audience statistics.

Data differences

The third concern centres around differences in the data collected, and not just differences between the diary and the meter, but differences between different meters, and even differences between the same meters.

Diary v Meter differences are not necessarily as severe as people might imagine. In fact, RAJAR has found that the number of quarter hours containing listening in a diary is virtually identical to the number of quarter hours containing validated exposure by a PPM. Good news for both the PPM and the diary. However, every audiometer field test has yielded some results that are so unlikely as to be incredible. If the differences are valid, is that because one system is less accurate than the other, or because they are measuring something entirely different? But perhaps they are plain shortcomings, and if they are, and such shortcomings turn out to be an inherent characteristic of meter measurement, then there may be no choice but to accept them if we are to pursue an electronic solution.

Compliance is the Key to Currency

There is one further obstacle in the road to a totally electronic future – the ironic fact that the only meter that seemed to work just had to be the most expensive - and if the UK is to replace its diaries with an equally effective sample of audiometers, then prices will have to tumble. But regardless of expense, the uncertainty surrounding respondent compliance, and the lingering unanswered questions, mean it would have been irresponsible of RAJAR to convert to a meter-based currency. If we only have two thirds of the data for three quarters of the participating sample, then we only have half of the truth. That may not matter, but we don't know that yet.

Similarly, we don't yet know whether compliance can be improved or influenced by better panel management methods, higher incentives or more selective recruitment, and whether ascription and/or modelling techniques can be used to restore the data in a manner that is both credible and acceptable. However, the R&D panel affords RAJAR an opportunity to grapple with these issues, and while there is no guarantee that in the future we will use this particular meter in this particular form, we can be certain that without this phase there can be no small portable electronically fulfillable future.