The next IASA conference this Autumn will shed light on many of the facets of access – as Access for All is the main theme of the conference. The conference will take place in India, in a region that is very much in the eye of the storm of transition.

It is with great pleasure that I invite you to join me and the hosts of the conference in New Delhi, India. I am absolutely convinced that it will be a memorable conference!

Yours *soundly*,
Jacqueline von Arb
President IASA

DAT-ERROR STATISTICS: AGE AND BRAND CORRELATIONS AS REGISTERED IN VRT'S DIVA-PROJECT[1]

Brecht Declercq, VRT, brecht.declercq@vrt.be

1. Introduction

Digital Audio Tapes, commonly known as DAT, are of huge importance to audiovisual heritage because large parts of radio station archives (among others) in the nineties were stored on this kind of media. Recording quality was very high and many broadcasters bought at least some DAT-players and recorders. Often these machines recorded radio programs from a central control unit between studio and antenna. DAT was also commonly used to make witness-recordings (e.g. at DR in Denmark) for the purpose of the right of reply, legal deposits or concert recordings in situ. Nearly parallel to the CD, the introduction of DAT as a production standard meant for many radio broadcasters the entering of the first digital recording era as well as the last non-file-based one.

Recording quality might have been high (48kHz was feasible and quite exceptional at the time), early enough it appeared that demagnetisation and an irreparable loss of signal (*Block Error Rate*, BLER) made the DAT a support of no use for long-term archiving. Moreover, Sony stopped selling DAT devices in December 2005, officially making DAT an obsolete medium. As a result, somewhat paradoxically considering its young age, a special focus on DAT appeared in audio collection digitisation projects of audio collections from 2007.

Is it possible, or even necessary, to assign a priority to certain DATs to be ingested (transferred to files) and if so, what should be the criterion? Are age and/or brand significant factors? Is there a risk of losing valuable content by setting wrong ingest priorities within your DAT collection? Considering the large numbers of DATs still to be ingested worldwide, does the DAT sound quality degradation advance in a straight line with age? Do certain brands represent a higher risk? The relevance of this issue might even become clearer when compared to the analog domain. A lot of studying has been done on the brand and age related degradation of analog audio and videotapes. Among the results is the infamy of e.g. 'red Ampex' in many audiovisual archives and the fame of Richard Hess's 2008 article.[2] Without the slightest ambition to reach Hess's writings in quality or eminence, this article tries to provide a first, partial answer on the questions above.

2. DAT ingest in VRT's DivA project

The DAT collection of Flemish public broadcasting service VRT is, to West-European norms, quite exceptional in its size. VRT purchased its first DAT-machines and tapes in 1986. The experimental phase took about five years, but from 1991 it became a particularly popular support for recording radio programs and concerts. Between 1991 and 2006, six DAT-recording devices were part of a so-called *automatic recording unit* where radio programs were recorded directly between studio and broadcast antenna. Most of the radio programs had a standard length of one or two hours; DAT was considered a perfect support for this purpose. For this reason a collection of ca. 44,500 DATs grew, with a total duration of ca. 71,200 hours, including about 32,000 DATs with radio programs (58,700 hours) and 12,500 DAT with music (13,700 hours).

In a mass digitisation project named DivA (Digital VRT Archives, 2008-2011) the threat of the sound quality on short term was decided to be a first rate selection criterion for digitisation

1 This article is an extended version of the paper ***DAT-error statistics: typology, risk-analysis and correlations as registered in VRT's DivA project,*** as presented on IASA 2011 in Frankfurt A.M., Germany, September 8th 2011. Kindest acknowledgments go to VRT archivists Lieve Vanhamel, Thomas Eyskens, Els De Vuyst and Klaas Janssens and IASA editor Cassandra Gallegos.

2 HESS, Richard. 2008. Tape degradation factors and challenges in predicting tape life. *ARSC Journal XXXIV* (2): 240-274.

and considering audio, DAT was chosen as the focus support. But as resources for ingesting the complete DAT collection were not available, a second selection step had to be defined: which DATs needed priority and for which ones digitisation would be postponed? Although sometimes very interesting on other DAT related questions, literature provided little answer, so we had to find clearer indications in the VRT collection itself.

The solution we figured out for this problem was twofold: on one hand we had the quality of the signal of a reasoned sample of our DATs tested on different criteria by a specializing company, which also compared the results of our collection to their large experience from ingesting DATs. Although this survey provided us with interesting results and a far better understanding of the mechanisms of signal loss, there was no clear one to one relation of any of the audio signal parameters with audible errors in the file. In other words: some peak values appeared to be false positives while listening and some clearly audible errors were not reflected in any of the signal graphs. The assessment results proved to be indicative, but not clear enough to serve as a base for a solid sound degradation analysis.

The other part of the solution consisted of a straightforward ingest of a long-running radio program, recorded on a diversity of DAT brands. A full-length sound quality check was done by a team of cataloguers trained on recognizing typical DAT-errors as a symptom of degradation. With a growing number of DATs ingested and fully checked, a quantitative analysis of these error reports became ever more significant. Doing so, we were able to draw some first conclusions about the relation between the sound quality, age and brand.

3. Research method

The researched sample consisted of exactly 1,000 fully recorded 120 min DATs, each with one episode of the Monday to Friday radio show *Het Vrije Westen* (VRT Radio 1), recorded from January 8th 1992 until May 2nd 1996. In the absence of any further possibility to know the exact production date of the DAT, this record date was as close we could get. The 1992-1996 period was also at the very heart of the era of the use of DAT at VRT.

The research sample contained 327 Sony tapes, 549 BASF tapes and 134 Ampex tapes. 218 DATs were recorded in 1992, 224 in 1993, 256 in 1994, 216 in 1995 and 86 in 1996, as represented in fig. 1.

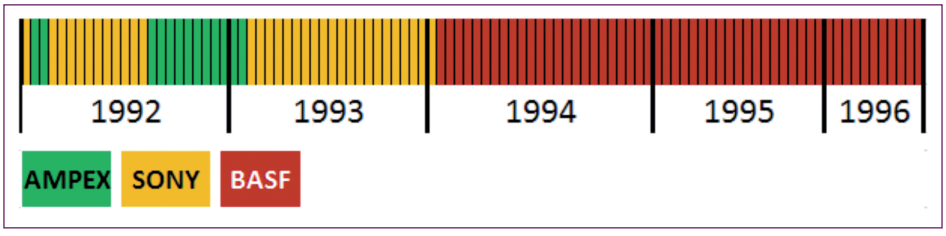

Fig. 1: brands and record years in the sample.

The transfer to file of these 1,000 DATs took place from August 24th 2009 until November 11th 2009, using six Sony PCM-7040 DAT-devices, digitally connected to an Intel Duo E8400 (3.00 kHz) standalone core, running Steinberg Nuendo 4.2.2 recording and editing software. Before ingest, all tapes were fast forwarded and rewound on a standalone Tascam device. The reading heads of all devices were regularly cleaned for 10 seconds with a cleaning cassette.

During the ingest, the DATs were controlled on physical tape problems such as broken tape or blockage in the machine. The ingest operators also performed a quick visual scan of the waveform, checking for major drops or peaks and documented these in the workflow files of the ingest process. Later on, documentalists were asked to pay special attention to all kinds of

audible errors while cataloguing the content during the full listening process (from November 18th 2009 until March 1st 2011). They documented the kind of artifacts they heard and included the length and time codes where these errors occurred. Every DAT in which the sound file contained at least one audible error was double checked to make sure that this error was already present on the DAT and not an unintentional result of the digitization process. All errors were taken into account and only after this full listening check were the affected DATs given a rerun with a Sony PCM-R500, connected to the same computer.

4. Results

4.1. Age

Out of 1,000 DATs 81 or 8.1% contained at least one audible error in the sound file. Fig. 2 reports the numbers of DATs affected, grouped according to their record year.

RECORD YEAR	DATs AFFECTED	TOTAL # OF DATs	PCT
1992	25	218	11.4
1993	24	224	10.7
1994	17	256	6.6
1995	14	216	6
1996	1	86	1.1
TOTAL	**81**	**1000**	**8.1**

Fig. 2: DATs containing at least one audible error, per year.

When represented graphically (fig. 3) we see clearly that figures drop over the five consecutive years in the sample. 1996, the last year, shows an exceptionally low number of only one affected DAT out of 86 in the sample, or 1.1%. However, when we consider all DATs recorded in 1996, of whom the major part is indeed out of the sample, the result is 5.4% of affected DATs, a figure more in line with the previous years.

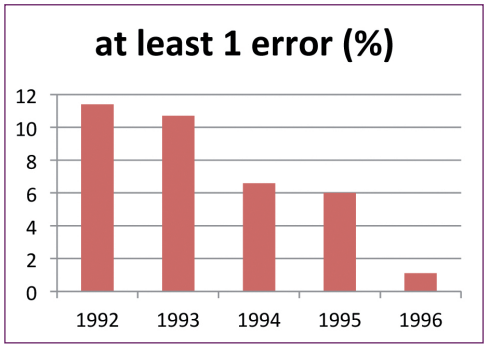

Fig. 3: DATs containing at least one audible error, per year (%).

Although a far lower result than originally feared at VRT, 8.1% of DATs affected may also seem high to some. Therefore it is useful to go beyond these first figures and have a look at the total time of audio affected. The research sample contained 120.000 minutes of audio, of which 1,211 minutes or slightly more than 1% was affected (fig. 4).

RECORD YEAR	1 SEC		2->10 SEC		11 SEC -> 1 MINUTE		1 MINUTE -> ENTIRE DAT		ENTIRE DAT	
	#	%	#	%	#	%	#	%	#	%
1992	6	2.7	15	6.8	2	0.9	1	0.5	1	0.5
1993	6	2.6	7	3.1	4	1.7	2	0.9	5	2.2
1994	7	2.7	7	2.7	3	1.1	0	0	0	0
1995	4	1.9	4	1.9	3	1.3	1	0.5	2	0.9
1996	1	1.1	0	0	0	0	0	0	0	0
TOTAL	**24**	**2.4**	**33**	**3.3**	**12**	**1.2**	**4**	**0.4**	**8**	**0.8**

Fig. 4: Total duration of the affected audio within the DAT, per year.

The difference between the total number of DATs affected and the total duration of affected audio is logically explained by the fact that most errors have a short duration. In every record year in the sample, most errors have a duration of one to ten seconds (fig 5). Only the number of the *one-second errors* and the *two-to-ten-second errors* tend to decrease in the researched period. The longer kinds of errors show no very clear tendency.

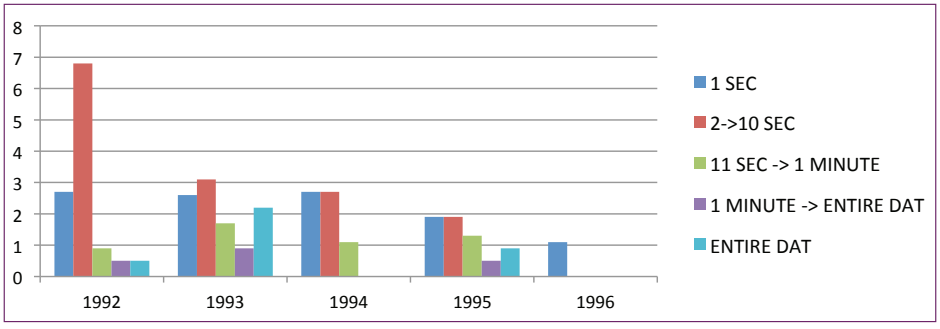

Fig. 5a-e: Total duration of the affected audio within the DAT (%), per year.

DATs with completely corrupted audio, a case widely feared for amongst audio archivists, proved to be quite exceptional in our sample. From only one (a SONY recorded in 1993) out of eight completely corrupted DATs did the full length sound had to be considered lost because even multiple reruns with different machines did not result in a better readout. Of the remaining seven the audio signal was captured with a Sony PCM-R500, not with a perfect result, but still far better.

4.2. Brand

A first examination of the brand results concludes a clear ranking lead by BASF (5.6% of DATs affected), followed by Sony (9.7%) and Ampex (12.6%) (fig. 6). Nevertheless a complicating factor is the fact that in our sample (and at VRT) not all brands were in use simultaneously. The fact that the BASFs, the youngest in the sample, are also the least affected, raises the question if and to which extent age and brand interfere as deterioration factors.

BRAND	# DATs IN SAMPLE	# DATs AFFECTED	% DATs AFFECTED
SONY	327	32	9,7
AMPEX	134	17	12,6
BASF	549	31	5,6

Fig. 6: DATs containing at least one error, grouped by brand.

The correlation between the brands of the DATs with the length of their errors (fig. 7) provides us with the argument that the older the DAT is, the longer its errors tend to be. When grouping the errors by length (fig. 8), the BASFs, youngest in the sample, have fewer errors and these even tend to be shorter than the errors of the Ampex and Sony cassettes. And even with some irregularities, two other facts confirm this indication: Sony cassettes typically have far more *two-to-ten-second errors* and the entirely corrupted audio comes most often from Ampex cassettes.

BRAND	1 SEC		2->10 SEC		11 SEC -> 1 MINUTE		1 MINUTE -> ENTIRE		ENTIRE DAT	
	#	%	#	%	#	%	#	%	#	%
SONY	6	1.8	18	5.5	5	1.5	3	0.9	1	0.3
AMPEX	6	4.4	5	3.7	1	0.7	0	0	5	3.7
BASF	12	2.1	10	1.8	6	1.1	1	0.2	2	0.4
TOTAL	**24**	**2.4**	**33**	**3.3**	**12**	**1.2**	**4**	**0.4**	**8**	**0.8**

Fig. 7: Total duration of the affected audio within the DAT, grouped by brand.

BRAND	1 SEC		2->10 SEC		11 SEC -> 1 MINUTE		1 MINUTE -> ENTIRE DAT		ENTIRE DAT	
	#	%	#	%	#	%	#	%	#	%
SONY	6	1,8	18	5,5	5	1,5	3	0,9	1	0,3
AMPEX	6	4,4	5	3,7	1	0,7	0	0	5	3,7
BASF	12	2,1	10	1,8	6	1,1	1	0,2	2	0,4
TOTAL	**24**	**2,4**	**33**	**3,3**	**12**	**1,2**	**4**	**0,4**	**8**	**0,8**

Fig. 8: Total duration of the affected audio within the DAT (%), grouped by brand.

4.3. Point in timeline

With the number and the duration of all errors indicated with time codes, it also became possible to obtain some interesting statistics about when in the audio timeline these errors occurred. The experience of radio technicians and archivist already allowed us to presume that the first and the last minutes of the audio were the most vulnerable. The figures from VRT's DivA project confirm these presumptions (fig. 9). The number of errors per minute is by far the highest in the first minute. Gradually it decreases to a very low level between the sixth and the 114th minute. The last five minutes show again more errors, but not as much as the first five.

MINUTE	1	2 -> 5	6 -> 114	115 -> 119	120
YEAR	errors / minute	errors / minute	errors / minute	errors / minute	errors / minute
1992	6	1	0,08	0,5	1
1993	8	3,5	0,1	2	6
1994	2	0,25	0,1	0,25	0
1995	4	0,75	0,06	0,25	1
1996	1	0	0	0	0

Fig. 9: Errors per minute in the audio timeline, per year.

Here again, this effect seems to be reinforced by the age of the DATs (fig. 10). And again this relation is not very clear-cut: 1992 and 1993 on one hand and 1994 and 1995 on the other switched positions, compared to what could be expected if a strictly chronological degradation had been the case.

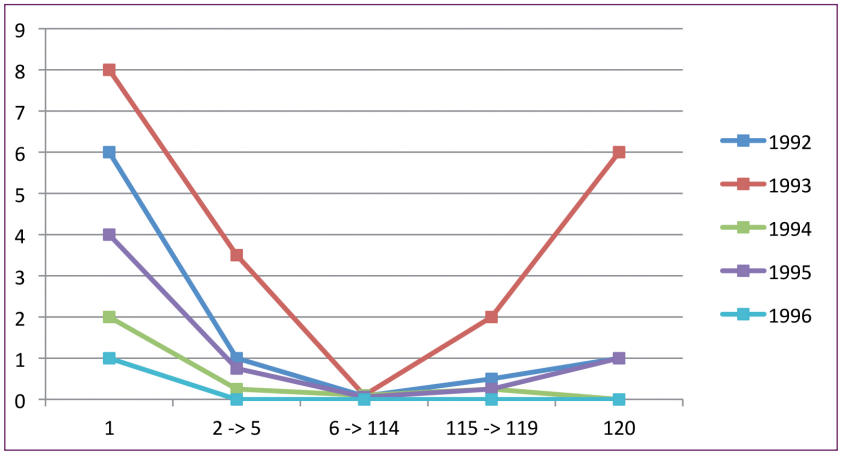

Fig. 10: Errors per minute in the audio timeline, outlined per year.

Considering the brands all three show the same effects of a considerably higher vulnerability for errors at the beginning and the end of the tape (fig. 11 and 12). BASF DATs are far less affected than Ampex and Sony. Ampex shows an additional vulnerability at the end, more than doubling the result of the Sony DATs.

MINUTE	1	2 -> 5	6 -> 114	115 -> 119	120
BRAND	errors / minute	errors / minute	errors / minute	errors / minute	errors / minute
SONY	9	1,75	0,14	1,5	2
AMPEX	7	2	0,08	1,25	5
BASF	5	1	0,22	0,25	1

Fig. 11: Errors per minute in the audio timeline, grouped by brand.

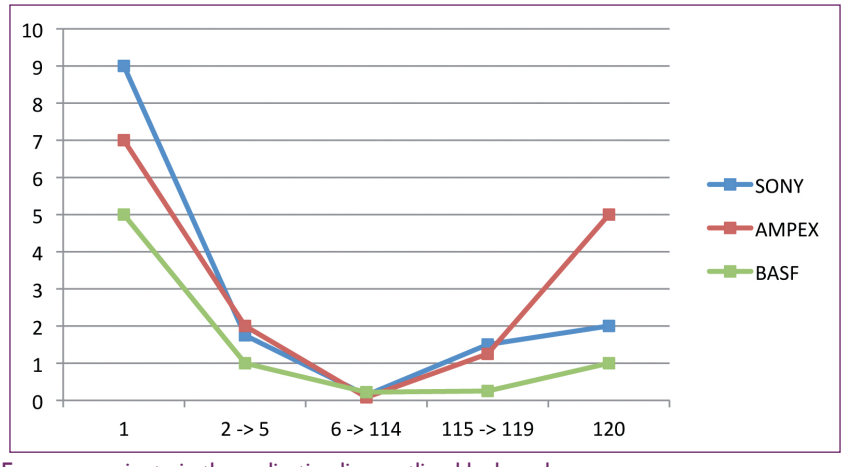

Fig. 12: Errors per minute in the audio timeline, outlined by brand.

5. Conclusions and nuances

In this article we have examined the results of the error logging by full human listening of 1,000 two-hour sound files, coming from an equal number of DATs. A recording period of almost 4.5 years was covered, from January 1992 until May 1996, with three major brands (Sony, Ampex and BASF) represented in the sample.

Answering the questions that have been lined up in the introduction, we have established that a total of 8.1% of all DATs were affected by at least one audible error. This number decreases every record year, leading to a first indication that there is a degradation of the sound signal through time. However, when we add up the full duration of all affected audio, we counted only 1,211 minutes (or slightly more than 1% of all audio), meaning that most errors are very short and only very few DATs are entirely corrupted. Considering length, only the numbers of *one-second errors* and *two-to-ten-second errors* tend to decrease through time. The longer kinds of errors show no clear tendency. DATs with a completely corrupted signal were very exceptional in our sample.

If we take a closer look on the brands Ampex seems to be the most vulnerable, followed by Sony and BASF. But as in our sample the different brands were not spread equally over the years (Ampex and Sony were in use from 1992 until early 1994, followed by BASF for the rest of the researched time span), it is still possible that age and brand interfere as factors of deterioration. Furthermore the errors coming from BASF cassettes were shorter and the entirely corrupted audio came most often from Ampex cassettes.

The results of our logging and research also reaffirm that the beginning and the end of the DATs are more vulnerable for signal loss. Here again there is a slight indication (but no explicit evidence) of gradual sound signal deterioration through time and a higher vulnerability of Sony and Ampex.

Should audio archives prioritize the DAT-to-file ingest with age or brand as a criterion? The answer can only be nuanced. There seems to be a slight relation between audio signal deterioration and the age of the DAT. Furthermore Ampex seems more vulnerable and faster in its degradation than Sony, and Sony in its turn seems more vulnerable and faster in its degradation than BASF. But the question remains to which extent the age relation and the brand relations influence each other.

To clear up these relations more diachronic (same brand, different periods) and synchronic (different brands, same period) studies are needed, also to exclude other possible factors such as differences in machine maintenance etc. But also the physical process of deterioration of the carrier and the signal stays an interesting subject for research. Regarding BLER we know that signal loss is gradual and that the possibility to calculate missing signals through the checksums stops at a certain threshold. In other words: an apparently perfect sound may conceal a very bad signal condition. 'Traumas' such as problematic recording conditions or poor conservation circumstances (e.g. exposure to extreme temperatures, humidity or magnetic sources) seem to play a role. In this respect the relatively low number of errors in our sample may be due to the professional recording and storage environments at VRT. If the history of the collection is known, it is therefore recommendable to give priority to 'traumatized' DATs.

If we consider the rather low quantity of lost audio in our sample, it may seem logic to shift the focus of DAT preservation strategies from the carriers to the (scarcity of) devices, which are often considered the bigger problem. According to Tim Bathgate however Digital Data Streaming (DDS) may provide the archival community with an escape route.[3] The signal loss problem though remains, up until today, irreparable and therefore continues to deserve our awareness.

3 BATHGATE, Tim. 2009. Took the one less traveled by. DAT migration at Radio New Zealand. IASA Journal 34: 50-64.