DAT vs. MD

Is Minidisc recording quality good enough for prosodic analysis?

© Nick Campbell & Parham Mokhtari

ATR Human Information Sciences
Seika-cho, Hikari-dai, Kyoto 619-0223

{nick,parham}@atr.co.jp

Abstract

This paper reports an analysis of speech data recorded on Minidisc, and shows that for the purpose of prosodic analysis, it can be considered equivalent to that recorded on DAT tape. The advantage of MD over DAT is its lightness and portability, but differences were found, resulting from signal compression, which may affect signal-processing techniques and prevent precise numerical comparisons of MD and DAT data.

1. Introduction

In order to collect really spontaneous speech data [1,2] it is desirable to use light and unobtrusive recording devices. Although DAT recorders are now very small, they are not yet pocket-sized, and are still too heavy to wear comfortably on the body during everyday activities. Minidisc technology is much lighter, but makes use of compression to reduce the amount of data stored on disc.

The ATRAC perceptual-masking compression [3] used in Sony MD recorders may render the recorded speech unsuitable for conventional signal processing techniques. We therefore carried out tests to determine the extent to which traditional methods of e.g. voice pitch estimation, formant-tracking, spectral analysis, and cepstral encoding may be degraded as a result of using speech data which has undergone perceptual-masking for compression of the recorded signal.

2. Generation of the data

To measure the difference between recording quality on DAT and MD, we used a single condenser microphone (*Sony C-355*) to record a 5-vowel sequence (a-i-u-e-o) from a male and a female speaker, taking the signal to a portable DAT recorder (*Sony TCD-100*) and a small MD recorder (*Sony MZ-R900*) simultaneously. We also recorded a 1-10kHz chirp tone and a 200-800Hz sweep tone, each with a sinusoidal waveform, produced by an NF Electronic Instruments DF-194A variable phase digital function synthesiser.

韻律処理に向けて: DAT と MD の比較
◎ ニック キャンベル、パーハム マクタリ、
ATR 人間情報科学研究所・JST/CREST-ESP

The recording levels of the two devices were adjusted to an approximately equivalent setting using these tones. The signals were transferred directly to computer disc using optical fibre via a Canopus MD-Port, and down-sampled to 16kHz 16-bit using Wavesurfer software [4]. Both Wavesurfer and Entropic's ESPS software were used for pitch-estimation, spectral display, and formant analysis. We also compared cepstra of vowel sequences generated by the ESPS *fftcep* program.

3. Results of the analysis

For reasons of space, we limit our report here to only a subset of the data, but full results will be presented in the poster presentation and further details can be obtained from the authors on request. In all cases, the visible representations of the signals were perceptually equivalent (compare Figures 1 and 2), but the derived values were not identical. Table 1 shows that the differences in estimated F0 and formant values are small but significant. Since the start points of the MD and DAT waveforms were aligned manually, and the signals were processed using identical (default) settings of the software, we would expect the estimated values for fundamental frequency and formants to be identical, even though there are, inevitably, small differences in signal power arising from differences in the recording level settings of the two devices. Figure 3 and 4 confirm the spectral similarity, showing almost identical peaks, but slightly less energy in the troughs for the MD signal.

Table 1. *Comparison of prosodic (top: fundamental frequency) and spectral parameters (bottom: formants and their bandwidths) derived from signals recorded simultaneously on DAT and Minidisc (MD) recorders.*

	Male		Female	
	Mean f0	Sd	Mean f0	Sd
DAT	100.108	8.301	171.06	34.7
MD	100.128	8.409	169.31	38.6

	F1	F2	F3	F4
DAT	369 (171)	1421 (376)	2497 (395)	3533 (379)
MD	370 (171)	1420 (376)	2490 (401)	3556 (402)
	B1	B2	B3	B4
DAT	134 (104)	305 (208)	351 (219)	426 (237)
MD	132 (99)	305 (200)	353 (223)	428 (239)

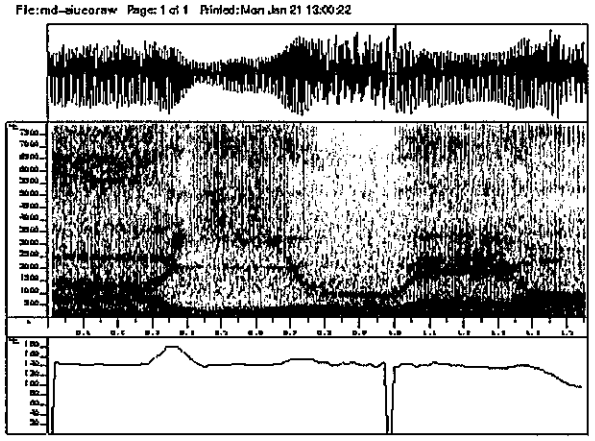

Figure 1. *Spectrogram and f0 of the 5-vowel sequence (from the male speaker) recorded using a Minidisc.*

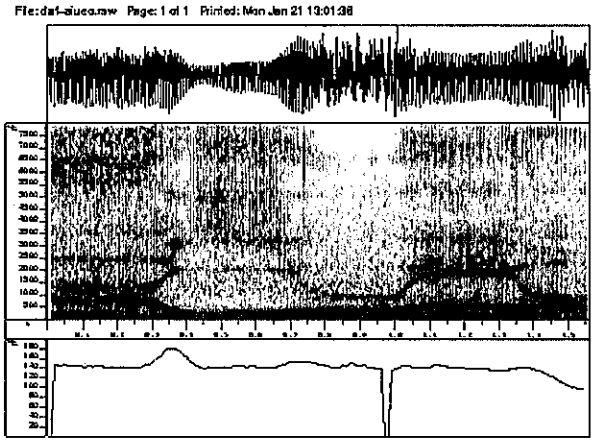

Figure 2. *Spectrogram and f0 of the 5-vowel sequence (from the male speaker) recorded on DAT.*

The cepstra showed the greatest differences, perhaps due to the small differences in signal power. Table 2 shows measured cepstral distances between equal length samples of 3 vowels. Inter-vowel distances measured within the same recording medium provide a baseline distance, against which intra-vowel, inter-media distances can be appreciated. In the perfect case, we would expect the bottom-row values all to be zero. We can see that the DAT-MD distance is almost half that of the /a/ - /i/ distance within each medium. (Signal range: DAT –9182 to 9606; MD –8146 to 9055).

Table 2. *Mean squared differences between sequences of vowels /a/, /i/, and /u/, (125 ms each) for DAT & MD. Figures are for spectral shape only, disregarding cep-0.*

	i – a	i – u	a – u	mean
MD – MD	0.0381	0.0297	0.0267	0.0315
DAT – DAT	0.0422	0.0268	0.0261	0.0317
	i – i	a – a	u– u	
MD – DAT	0.0207	0.0186	0.0060	0.0151

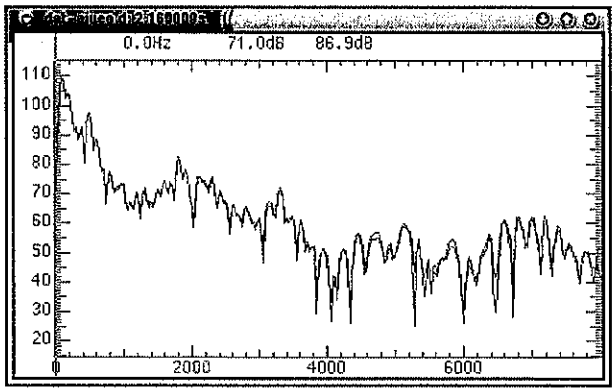

Figure 3. *Spectral slices of vowel /a/ (both DAT & MD). This small degree of difference is typical of most samples we measured in the 5-vowel sequence*

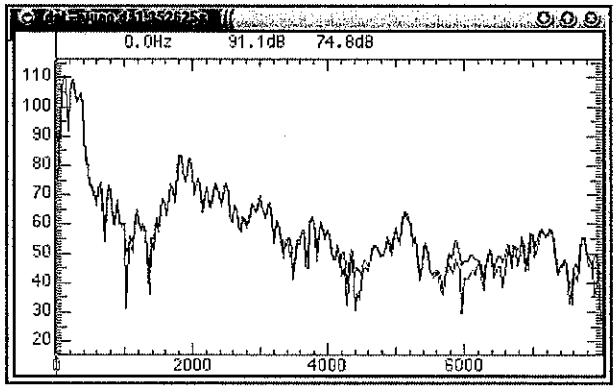

Figure 4. *Spectral slices of transition /a/ to /i/ (both DAT & MD) taken close to the f0 maximum. The large difference around 6kHz represents the worst mismatch we found in the data.*

4. Conclusions

This paper has presented results of a comparison of speech recorded on both DAT and MD. We found that the sampled data differ significantly, presumably as a result of the MD's ATRAC compression. However, the differences were small, particularly for estimates of formants and fundamental frequency, and we conclude that speech recorded on MD can be used for the analysis of spectral and prosodic characteristics. We note, though, that precise numerical comparisons should not be made with equivalent speech data recorded on DAT.

5. References

[1] Campbell, N., "The Recording of Emotional speech; JST/CREST database research", in Proc LREC 2002.

[2] Campbell, N., "Collecting really spontaneous speech", in Symposium on Prosody & Sp. Processing., Tokyo, 2002.

[3] ATRAC compression: www.minidisc.org/aes_atrac.html

[4] Wavesurfer: see http://www.speech.kth.se/wavesurfer