A Software Radio Approach to Commercial FM Content Indexing

Bruce Denby, Olivier Romain, Sid-Ahmed Hariti

Laboratoire des Instruments et Systèmes d'Ile de France (LISIF)
Université Pierre et Marie Curie
B.C. 252, 4 place Jussieu, 75252 Paris Cedex 05, France
email: denby@ieee.org

Abstract. – **The article presents a feasibility study for a Software Radio based commercial FM content indexing system, focussing on the possibility of future portable and embedded implementations.**

Keywords: broadcast monitoring; audio indexing; software radio; FM radio

I. INTRODUCTION

There is considerable interest today in the use of *audio indexing* techniques to facilitate classification and retrieval of information contained in commercial broadcast media [1-7]. Such technology is in demand both by the broadcast monitoring community - for intellectual property rights issues, and by the entertainment industry - for providing consumers with novel media search and retrieval services.

Audio indexing may be *content-based* - for example speech/music discrimination based on the audio waveform - or *metadata-based*, exploiting explicit tags broadcast along with the audio content - such as in the Radio Data System (RDS) stream contained in most commercial FM radio broadcasts.

Commercial broadcast monitoring systems are typically rather massive installations, due to the large number of channels to be treated. For many consumer applications, however, it would clearly be advantageous, to imagine portable audio indexing solutions - for example in conjunction with an FM car radio, to see if a particular song has been played in the past few minutes, or to tag station content in real time as voice or music. Recently [8], *software radio* has been proposed as a possible enabling technology for this purpose. Originally conceived to provide adaptable air-interfaces for the wireless telecommunications industry, software radio is now discussed for commercial FM radio applications as well [9,10]. In an audio indexing context, software radio is interesting in that it digitizes entire chunks of radio spectrum and analyzes them in software. Thus, coupled with a memory module or a hard disk, a software radio solution with sufficient processing power could provide a means of transforming the FM radio band into an indexed, searchable 'Hertzian Internet.' The programmability of software radio architectures, furthermore, would make them all the more attractive as future DAB and iBOC services - also foreseen to make extensive use of metatdata tags - begin to roll out. Suitably equipped receivers could take advantage of these new services without hardware upgrades, simply by downloading a new software module.

The article presents the results of a feasibility study for a software radio based FM band 'navigator' intended to enable both content- and metadata-based audio indexing processes. The navigator concept is first introduced in section II. Section III then describes the hardware and algorithmic studies performed to assess the feasibility of a portable or embedded FM navigator solution. In the final section, a conclusion and a discussion of planned future developments are presented.

II. THE SOFTWARE RADIO ENABLED FM BAND NAVIGATOR CONCEPT.

The FM navigator concept is depicted in figure 1. After reception by the antenna and low noise amplifier (LNA), the downconverter and bandpass filter (BPF) move a segment of radio frequency spectrum - in this case, the entire 20 MHz-wide commercial FM radio band - down to baseband for digitizing. The downconvert is 'optional' in the sense that if a digitizer of sufficient speed is chosen, downconversion is unnecessary.

This mode of operation is to be contrasted with that used in a standard FM radio, where a phase-locked loop (PLL) is used to downconvert and demodulate one station at a time. If all channels were desired in parallel, separate downconverters, filters, and digitizers would thus be necessary. In the software radio approach, it is not until

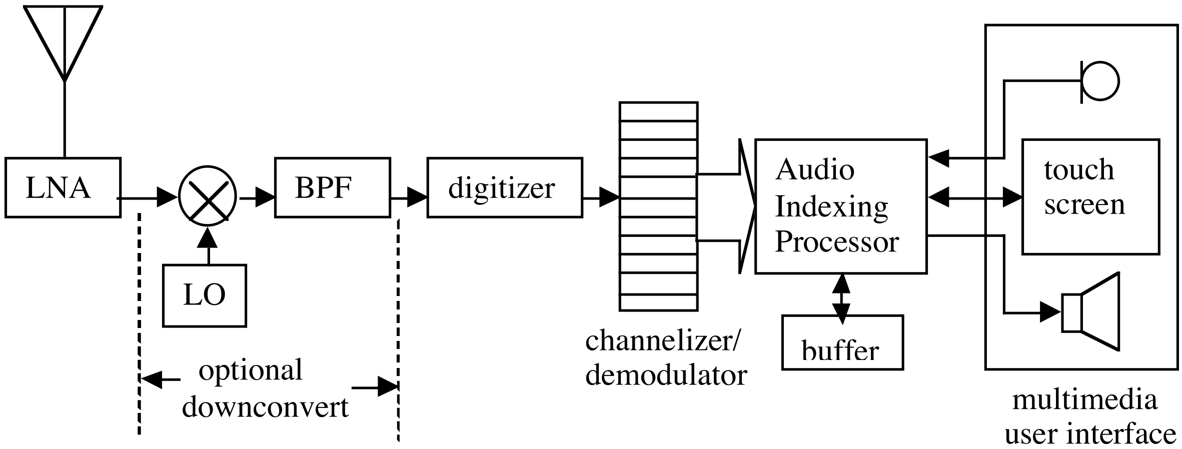

Fig. 1. Basic architecture of a software radio enabled FM navigator.

after digitization that the channels are separated, using digital filters, in the module labeled 'channelizer/ demodulator.' This element is also responsible for demodulation, i.e., extracting the baseband information from the RF signal of each radio station contained in the band.

After digitization, the audio indexing may begin. It is in this module that speech/music discrimination, audio track identification, word spotting, and or other indexing algorithms may be performed. It is important to realize that only a relatively small buffer would be needed to store several minutes of audio content from all stations, thus allowing 'rewind' capabilities and extending searches backwards in time. The RDS, DAB, or iBOC information is also decoded in this module.

The multimedia interface outputs an audio-visual representation of the retrieved indexing information - for example a screen icon for each station indicating something of its current content - along with the audio stream itself, and also allows input of voice or touch screen commands from the user.

III. Feasibility study

A. *Extraction of Audio and RDS Waveforms*

For the study, the FM input signal bandwidth was limited to 2 MHz in order to reduce the data transfer rates, as the post-acquisition processing was done using Matlab and was thus much slower than what can be achieved on a DSP architecture. The RF signal from the antenna - a simple piece of wire - was first downconverted to an intermediate frequency of 10.7 MHz, then sampled at 4.8 MHz using *bandpass sampling*. Bandpass sampling can be advantageous since it allows to sample at twice the total signal *bandwidth*, rather than twice the maximum frequency contained in the signal, as in Nyquist sampling. This may permit the use of simpler and less costly ADC, which is an important consideration for embedded applications. The resulting spectrum appears in figure 2, where peaks corresponding to 4 FM radio stations are visible. The multiple narrow peaks at low frequency are noise due to inadequate shielding. In addition, poor downconverter passband rolloff caused substantial attenuation of the first and last peaks, so that only the central 2 stations were fully exploitable, as described below.

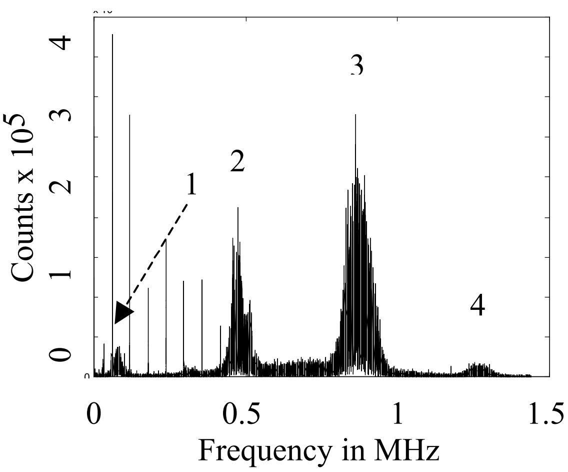

Fig. 2: Frequency spectrum of the undersampled RF signal, showing the presence of 4 commercial FM radio stations (labeled 1 to 4) with a 400 kHz separation.

Channelization was accomplished using a Weighted Overlap Add (WOLA) algorithm [11] coupled with a single 200 kHz low pass filter (LPF). It was desired to test the WOLA algorithm in our study as it is designed to afford substantial speedups in DSP implementations of systems having multiple equally spaced channels. The degree of speedup increases with the number of channels, and is due to the possibility of time-aliasing of the input signal - amounting to a reduction in the input window size - and to the use of an FFT to do the downconversion to baseband. WOLA may thus permit, in a final embedded application, to use a less expensive DSP with lower clock speed and hence lower power consumption. The algorithm also allows an arbitrary decimation factor. In our case this was set to 10, outputting thus for each radio station a baseband signal sampled at 480 kHz.

FM demodulation was accomplished by estimating the instantaneous frequency of the baseband signal at each point in time via a simple trigonometric identity applied to a triplet of samples centered on the time point. Except for peak 1, which is swamped by the noise, the resulting baseband frequency spectra clearly showed the L+R monaural, 19 kHz sub-carrier, L–R stereo, and 57 kHz RDS portions of the signal, as shown in figure 3 for largest peak (number 3).

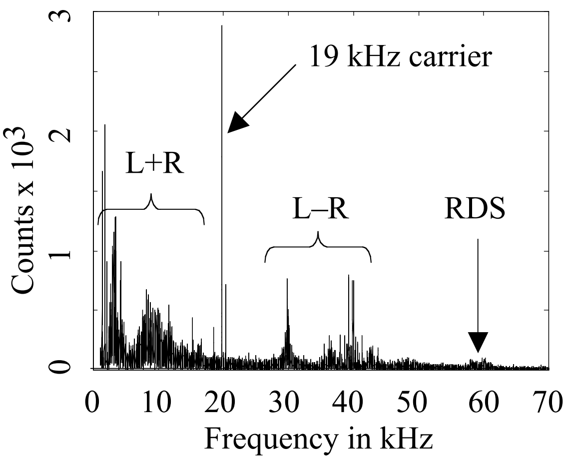

Fig. 3: Baseband signal of peak 3 after WOLA channelization/decimation and FM demodulation. The zero is suppressed to give a better view at higher frequencies.

The desired bands can now be selected using baseband filters and the appropriate treatments applied. The L+R (mono) waveforms, for the purposes of this study, were simply further decimated to 11025 samples/s and written to .wav files in order to check sound quality. An example waveform, from peak 2, is shown in figure 4 for a 0.6 second long segment. For the RDS band, a simple software decoder implementing a subset of the RDS data fields was developed. The L–R stereo signals were not processed.

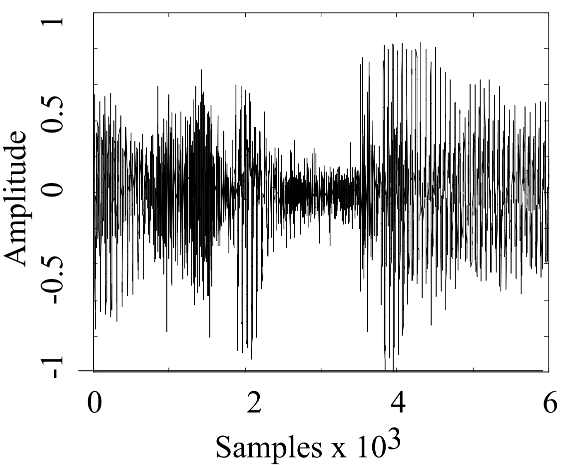

Fig. 4: Audio signal recovered from the RF peak labelled '2' in figure 1. The clip lasts 0.6 seconds and is sampled at 11025 samples/s. See also table I.

Results of the audio and RDS baseband signal analyses are summarized in table 1. The quality of the audio signals of peaks 2 and 3 is quite good, with words clearly recognizable. The audio of peak 3 is noisy, while peak 1, perhaps not surprisingly, is unusable. Concerning RDS, energy was found at the 1187 Hz RDS bit frequency for peaks 2, 3, and 4, but the decoder was able to achieve synchronization only on peak 3. The lack of synchronization on the weaker stations may be in part due to the shortness of our clips - 0.6 seconds - whereas RDS capture time may be up to 4 seconds on commercial tuners. On peak 3, nonetheless, the Program Identifier field could be decoded without error correction, and the Program Name field using only lowest order error correction. The decoded strings, F220 and NRJ respectively, correspond to a popular Paris area rock station. It thus seems likely that with better shielding and filters and a higher gain in the RF stage, it should be possible to decode RDS on the majority of the stronger FM stations with our system.

Table 1: Summary of FM Channel Content Analysis

Peak	**Audio Content**	**RDS**
1	noise	not detected
2	speech (French) "…et sept ans…"	1187 Hz RDS bit rate
3	rock music lyric "…all weise…"	Program code: F220 Station name: NRJ
4	music+noise	1187 Hz RDS bit rate

B. Audio Indexing

Word spotting, song identification, and the like, are ambitious goals for a real time audio indexing system which is intended to be portable. The purpose of this audio indexing study was to examine simpler algorithms which will nonetheless be interesting for a first test of such a system. The algorithm chosen was a speech/music discriminator based on simple level crossing variables [5,8]. To test its performance, thirty each of 3.5 second speech and music clips were recorded from FM radio broadcasts. These clips did *not* undergo the signal processing analysis described in the preceding section, since the time to produce them using the Matlab code would have been exceedingly long. They were simply recorded from the radio using a sound card. As the sound quality obtained by our earlier method was good, however, the performance observed here should still be approximately indicative of what will be possible in a future system.

The results are presented in figure 5. The figure shows the number of level crossings lasting longer than 20 samples versus the mean duration in samples of such crossings, for a threshold level of 20% of the maximum normalized absolute audio signal value. Music tends to consist mostly of very short level crossings, whereas voice, due to the well known 4 Hz syllabic frequency, tends to contain level crossings of longer duration. The figure shows that a cut on level crossing duration at 41 samples retains 90% of the music clips with 0% false positives. Such performance is already rather interesting, making the speech/music detector a good candidate for inclusion in a prototype embedded real time audio indexing system. It should furthermore be possible to improve the performance, if necessary, using some of the simple variants proposed in the literature [see 12 and references therein].

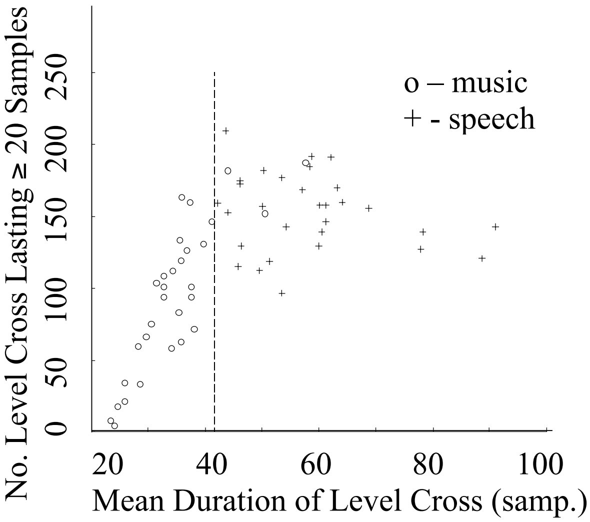

FIGURE 5. Mean number of level crossings of the (absolute value) audio envelope lasting at least 20 samples, versus the mean duration of these crossings (in samples), for 30, 3.5 second commercial FM music (circles) and 30 voice (crosses) clips. The threshold for level crossings was 20% of the peak absolute envelope value.

C. Real Time Concerns and the "GO" Band Demonstrator

A number of signal processing issues must of course still be addressed before the FM band navigator can be realized. It is not known at this time which type of real time DSP architecture will be necessary to allow the extraction of the audio envelopes, RDS decoding, indexing algorithms, and multimedia interface to run concurrently for several FM stations in real time.

In order to help understand some of the challenges to be faced, a simpler demonstrator has been developed for the French "GO" long-wave AM band, situated at 157-260 kHz. Based on the SHARC ADSP21061 EZ-Kit Lite board [13], the board samples in bandpass mode at 80 ksamples/s and thus produces data rates which are much more manageable. Details of the system, which produces real time audio output for stations selectable by on-board pushbuttons, are being presented in a separate paper [14]. Experience thus far, however, has indicated that the EZ-Kit DSP is not rapid enough for an FM navigator.

IV. Conclusions and Future Perspectives

The study shows that a bandpass sampled FM radio signal can be used to produce both good quality audio and RDS outputs for multiple stations. The WOLA technique was shown to be applicable for channelization, which will be advantageous for a DSP implementation of our system. Level-crossing variables give good performance on voice/music discrimination and thus appear to be good candidates for a simple yet convincing real time audio indexing system. Finally, valuable experience in real time software radio has been gained *via* an AM "GO" band navigator, which is being presented separately.

The goal of future work will be to incorporate the above elements into a working demonstrator. A new RF front end comprising downconversion, amplification, and improved analog filtering and shielding, is currently under construction. This should appreciably improve the quality of the raw signals fed to the digitizer, and will also extend the bandwidth of the demonstrator to include a larger number of stations. The target DSP architecture for the FM navigator system is still under study.

V. References

[1] http://www.confirmedia.com/

[2] http://www.cyberalert.com/

[3] http://www.shazam.com

[4] Beth Logan, Pedro Moreno, JM Van Thong, "Approaches to Reduce the Effects of OOV Queries on Indexed Spoken Audio," Cambridge Research Laboratory, HP Laboratories Cambridge, HPL-2003-46, March 5, 2003.

[5] R. Jarina, N. Murphy, N. O'Connor, S. Marlow, "Speech-Music Discrimination from MPEG-1 Bitstream," *WSES/SSIP01*, Malta, September 1-6, 2001.

[6] W. Chou and L. Gu, "Robust Singing Detection in Speech/Music Discriminator Design," *ICASSP01*, Salt Lake City, May 7-11, 2001.

[7] J. Haitsma, T. Kalker and J. Oostveen, "Robust Audio Hashing for Content Identification," *CBMI01*, University of Brescia, Italy September 19-21, 2001

[8] B. Denby and Sid Ahmed Hariti, "Towards a Software Radio Enabled Broadcast Media Navigator," *EC-VIP-MC2003*, Zagreb, Croatia, July 2003.

[9] Flood D., Doyle L., Mackenzie P., Nolan K., O'Mahony D., "Exploiting The Dynamic Flexibility Of Software Radio In FM Broadcast Receivers," *ARRL and TAPR 22nd Digital Communications Conference*, Hartford, Connecticut, USA, September 2003.

[10] http://www.vanu.com/

[11] R.E. Crochière, "A weighted overlap-add method of short time Fourier analysis/synthesis," IEEE Trans. Acoust. Speech Signal Process. ASSP-28 (1980) 99-102.

[12] Tong Zhang, and C.-C. Jay Kuo, "Audio Content Analysis for Online Audiovisual Data Segmentation and Classification," IEEE Transaction on Speech and Audio Processing, vol. 9, no. 4, May 2001 441.

[13] http://www.analog-devices.com/

[14] O. Romain, B. Denby. S.A. Hariti, "A Real Time Software Radio for the AM Long Wave Band," these proceedings.