JOURNAL OF THE KOREAN ASTRONOMICAL SOCIETY
34: 1~8, 2001

A FAST REDUCTION METHOD OF SURVEY DATA IN RADIO ASTRONOMY

YOUNGUNG LEE
Korea Astronomy Observatory/Taeduk Radio Astronomy Observatory, Whaam-dong 61-1,
Yusong-gu, Daejon, 305-348, Korea
E-mail: yulee@trao.re.kr

(Received Dec. 13, 2000; Accepted Jan. 17, 2001)

ABSTRACT

We present a fast reduction method of survey data obtained using a single-dish radio telescope. Along with a brief review of classical method, a new method of identification and elimination of negative and positive bad channels are introduced using cloud identification code and several IRAF (Image Reduction and Analysis Facility) tasks relating statistics. Removing of several ripple patterns using Fourier Transform is also discussed. It is found that BACKGROUND task within IRAF is very efficient for fitting and subtraction of baseline with varying functions. Cloud identification method along with the possibility of its application for analysis of cloud structure is described, and future data reduction method is discussed.

Key words : spectra–ISM: survey-methods: data reduction

I. INTRODUCTION

The universe is the fascinating place forever. Perhaps it is because the space is tremendously vast with its countless bizarre celestial objects and mysteries that bend common sense. They are often capable of humbling the experts as well as the beginners. Naturally human beings have wanted to *see* those objects more clearly as much as possible. Through this strong desire, observing tools and devices in 20th century is fortunately being developed in many record-breaking ways, including larger telescopes, space telescopes, interferometers, larger and more sensitive CCDs, array receivers, and many other ingenious devices. Moreover, these observing tools are now covering almost the whole range of electromagnetic radiation with unprecedented resolution and sensitivity with the help of computer and high-technology. The amount of data is also increasing in record-breaking ways, which astronomers have to deal with. Ten years ago, doctoral students, majoring radio astronomy with single dish telescope, usually reduced a few megabytes of data to complete a successful doctoral program. But these days they have to deal with a few scores of megabytes or more to get a successful degree. Or alternatively, they should get higher sensitivity data with higher resolution. In fact, there have been many survey projects constructing huge amount of database in almost all range of electromagnetic radiation, such as IRAS, MSX, 2MASS, HIPASS, SDSS, etc., and many new projects are going on in much larger scale. Therefore, it has been a big issue to successfully reduce and to construct a huge amount of data. Optical astronomers have been developing the necessary program packages to do fast reduction job, such as IRAF (Image Reduction and Analysis Facility), and several other attached system packages (e.g. STSDAS), while radio astronomers have to deal with spectra from their radio telescopes using mostly stand-alone reduction package. A universal reduction package, AIPS, for radio astronomers is more or less focused on interferometer and VLA data reduction. Observers of the 14 m telescope at Taeduk Radio Astronomy Observatory (TRAO) have been using the program, *SPA* (*SP*ectrum *A*nalysis), reducing the obtained spectra mostly one by one. In fact, most of single-dish radio astronomers have reduced the spectra one by one. However, as the number of spectra obtained is being exploded since the multibeam-array receiver with much higher sensitivity has been developed, the volume of spectral database has been substantially enlarged. Thus, there is request for a faster reduction method to take off heavy load of data reduction. In this paper we'll introduce a faster reduction method of spectra accumulated using a single-dish radio telescope.

Classical reduction methods of spectra from single-dish radio telescope are reviewed in Section II. We present a fast method of data reduction and some examples in Section III. In Section IV an efficient method of analyzing cube-data and cloud identification technique is described. In Section V the results and some possibilities of future reduction methods are discussed, and we summarize our results in Section VI.

II. CLASSICAL REDUCTION METHOD

Most of the observatories have their own reduction programs. At TRAO, we have used two versions of *SPA*, MODCOMP version and UNIX version to reduce the spectral and continuum data obtained from the 14 m radio telescope. *SPA* was originally written by Nick Scoville in 1978 for the reduction of millimeter wave data generated by the 14 meter telescope of the Five College Radio Astronomy Observatory (FCRAO) using FORTRAN 66 programming language. Due to the severe limitations of memory and hard disks available on computers in that era, the program was written with very small memory and disk requirements. In 1989, the program was rewritten in FORTRAN 77. In addition, machine dependent code was isolated into blocks so that a preprocessor can translate the code to any of five target architectures and operating systems. These include SunOS UNIX, VAX/VMS and PC-MSDOS. In 1990-91, the command parser, directory structures, and tape format were radically altered to remove the major limitations to the program which included the 1000 scan capacity of the data file, the 300 scan limit to the input arguments, and the fixed record length of a scan. In addition, the control loop of the program was modified to accommodate a secondary scan structure specific to data from an array receiver (Brewer & Heyer 1992). Some of reduction procedures in *SPA* are exported into IRAF as a user package, called as *fcrao* package, including ELIM, BASELINE, etc. Major reduction procedures in two versions of *SPA* could be divided into three different categories; bad channel elimination, baseline subtraction, and ripple pattern subtraction.

To eliminate a bad channel, one should search through a spectrum for bad channels and replace them with interpolation process. There are three methods the user can employ to identify bad channels using parameter ETOL, which stands for elimination tolerance: (1) If ETOL > 0, then any channel within the specified interval that deviates more than a factor of ETOL from the RMS of the spectrum will be eliminated. (2) If ETOL < 0, then any channel within the specified interval that has an antenna temperature (in absolute value) greater than or equal to abs(ETOL) will be marked for elimination. (3) If only one channel is specified, the channel will be marked for elimination regardless of the value of ETOL.

BASELINE subtracts a polynomial baseline from a spectrum. The coefficients in the polynomial are determined from a least square fit to the observational data in the specified intervals. After the fitted baseline has been subtracted, the rms noise level is calculated for the data in the specified intervals. Though there are higher order polynomials, linear, second and third order fittings are usually used.

To subtract ripple pattern(s) we can use SINE command within *SPA*. It subtracts a sine wave for which the period, phase, and amplitude are obtained from a least squares fit to the observed spectrum in the specified intervals. After the fitted baseline has been subtracted, the rms noise level is calculated for the data in specified intervals. This routine is designed to subtract one ripple pattern at a time. When there are more than one ripple pattern, this routine should be redone again.

III. A FASTER REDUCTION METHOD

(a) General Scheme

The above-mentioned reduction method is mostly based on manual mode, reducing one-by-one, though one may run a limited batch job for a stack of spectra. For a small number of spectra, manual mode may be good enough as one may conduct the reduction process very carefully. However, if the number of spectra becomes larger, for example, more than a few thousands, one may not be able to conduct a consistent reduction job from the first to the last, though one may be getting used to the reduction process itself. Human error factors (tiredness, and gradual generosity) grow bigger as the process goes on. Eventually, the reduction process may loose its consistency.

Before reducing spectra one by one, one may transform a stack of raw spectra onto a three-dimensional grid, constructing a cube data. Cube data comprises of velocity and two spatial coordinates, (v, l, b), velocity and galactic coordinate, or (v, α, δ), velocity and equatorial coordinate. Each section of the cube data can be displayed on a two-dimensional image displayer at once, saoimage or ximtool, when using IRAF. In *SPA* one may see the line profile of each spectrum, including the bad baseline(s) and bad pixel(s). The transformation of spectra to image can allow us to recognize the patterns of the baselines, bad pixels, and the quality of data. This process can be done within several astronomical reduction packages, but in this paper we will focus on IRAF, which is widely being used in astronomical community. Transformation of spectra to image has previously been suggested by FCRAO software group (Heyer & Carpenter 1991), when they developed *fcrao* reduction package. Within IRAF the obtained spectra can be transformed to IRAF data format or FITS format, and this can be handle as a 3-dimensional image. Image displaying and recognizing the patterns are the major merits of this method. Another merit is displaying the composite spectra of any image section up to a few hundred spectra at once. However, *fcrao* package does not include statistics and ripple pattern subtraction routines, thus the efficiency of data reduction is not enough to handle a large amount of spectra. In fact, there are several hundreds of reduction tasks within IRAF which can be manipulated for the more efficient reduction process, including 'imhistogram (or phistogram)' and 'imstatistics'. These two tasks are excellent for the statistics of the whole image or partial image section.

(b) Bad Channel Elimination

i) Negative bad channel elimination

When handling bad channels, one may consider them as a part of excessive noises. Bad channels are usually caused by bad channel performance, and can be easily identified as they are prominent comparing with other normal noise with negative or positive values. Bad channels (or pixels) with negative temperature values will be nominated as negative bad channels, and bad channels with positive temperature values will be called as positive bad channels. In this paper, elimination means interpolating with neighboring pixels, instead of nullifying the bad channels.

Within IRAF there is a useful task, generating the histogram and its plot of an image, which is 'phistogram' or 'imhistogram'. Either task can show the distribution of antenna temperature of an image very effecti-vely. For well-reduced image, the noise distribution would be a normal one in the negative temperature region of the histogram. If there are substantial number of negative bad pixels in an image (see Fig. 1), they will be represented as an 'excessive noise' in the histogram plot (Fig. 2a). Threshold value of these excessive noise can be determined as a displacement of normal distribution marked as an arrow in this figure. Identification of these bad pixels from the image can be conducted in several ways. We were able to manage this problem with 'Cloud Identification Code', which was developed ear-

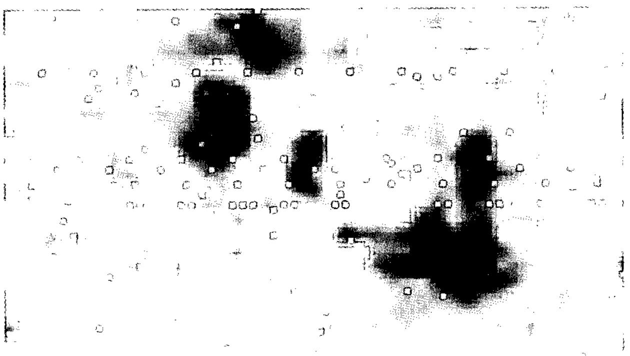

Fig. 1. Heavily contaminated image with many negative bad pixels (represented as white dots). The original image is one channel map (^{12}CO J=1–0) of Galactic Anticenter molecular cloud taken from Lee et al. (2000). The grey scale ranges from −1 to 5 K.

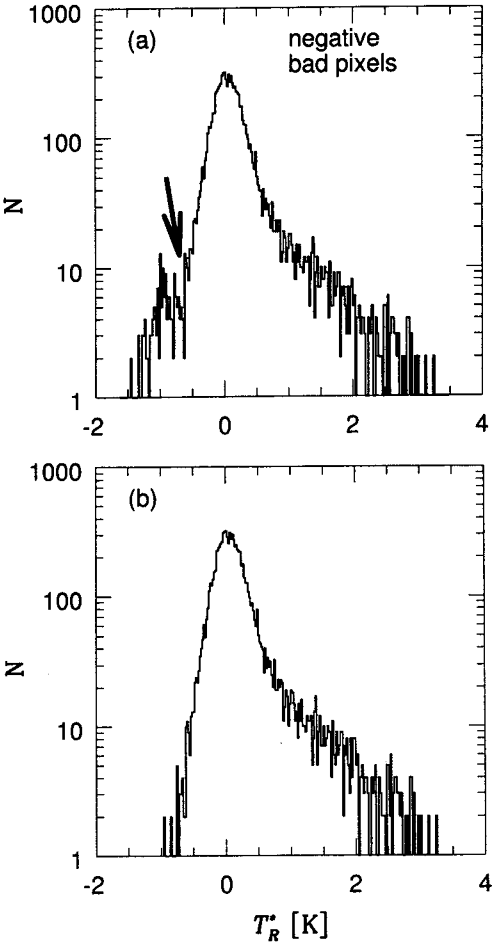

Fig. 2. (a) Distribution of CO temperatures in all the spectra of Fig. 1 is represented as a histogram. The temperature bin is 0.1 K. The arrow shows a threshold of excessive noise temperature (b) Fixed temperature distribution of negative emission after eliminating the bad pixels.

Fig. 3. Reduced image after eliminating the bad negative pixels Grey scale range from −1 to 5 K.

lier (Lee et al. 1997). Cloud identification process will be discussed in Section IV. Taking the clustering of negative pixels as imaginary negative clouds, bad pixels can be identified as one- or several-pixel clouds with a proper threshold temperature. The next procedure is interpolating the identified bad pixels with neighboring 2 or 4 pixels. The reduced image after in-terpolation of bad channels is presented in Fig. 3, which also shows a normal noise distribution in Fig. 2(b).

ii) Positive bad channel elimination

Identification of positive bad channels (or pixels) is more difficult than that of negative bad channels as they are often entangled with real emission (Fig. 4 and Fig 5a). However, some of these bad channels can be handled in the same manner as the negative bad channels; the positive bad channels in no emission regions can be easily identified using the cloud identification process with a proper threshold temperature. Taking the clustering of positive pixels as imaginary clouds, bad pixels can be identified as one-pixel clouds. Notice that the clouds identified with more than a few pixels are presumably real clouds. The next procedure is interpolating the identified bad pixels with neighboring 2 or 4 pixels. When they are mixed with real emission pixels within real clouds, it might be a tedious job to pick out and to eliminate them. These bad channel candidates should be checked by comparing neighboring images frames of three directions. The reduced image after interpolation of bad channels is presented in Fig. 3, which presumably has the same temperature distribution (Fig. 5b) as in Fig. 2b.

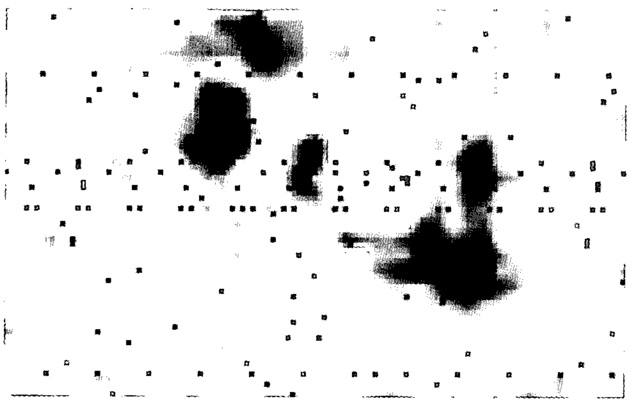

Fig. 4. Heavily contaminated image with many positive bad pixels (represented as darker dots). The grey scale ranges from −1 to 5 K.

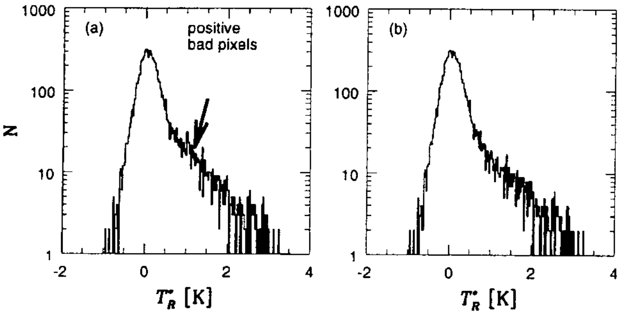

Fig. 5. (a) Distribution of CO temperatures in all the spectra retained in Fig. 4 is represented. The temperature bin is 0.1 K. (b) Fixed distribution after eliminating the positive bad pixels.

(c) Ripple Pattern Subtraction

Since the receiver gain variations often have a spectral period considerably longer than the width of the spectral lines being observed, the effects due to the receiver itself are usually removed using software by subtraction of a low or even higher order polynomial. However, there is more insidious standing waves of RF power between the parts of the antenna, the receiver, and the telescope environmental enclosure, such as radome. This standing wave is sometimes called as ripple pattern, and shows up in spectra as a frequency modulation of the total power at a period equal to the inverse of the round trip travel time between reflecting the surfaces (Goldsmith & Scoville 1980). The ripple patterns have substantially been reduced as the several techniques have been applied in various ways to the quasi-optics part of the antenna and receiver system. However, the small coherent reflection and unknown instrumental errors are always spreaded in several parts of the telescope

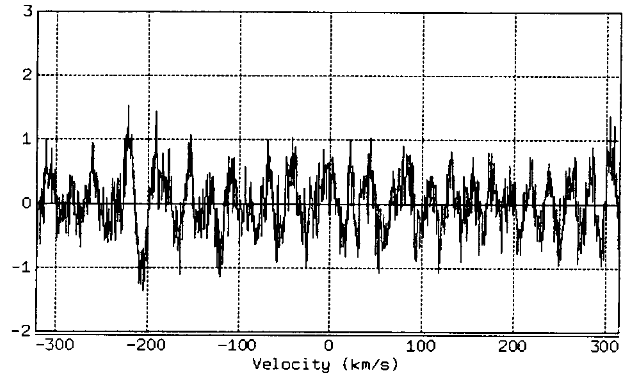

Fig. 6. A spectrum contaminated with several ripple patterns.

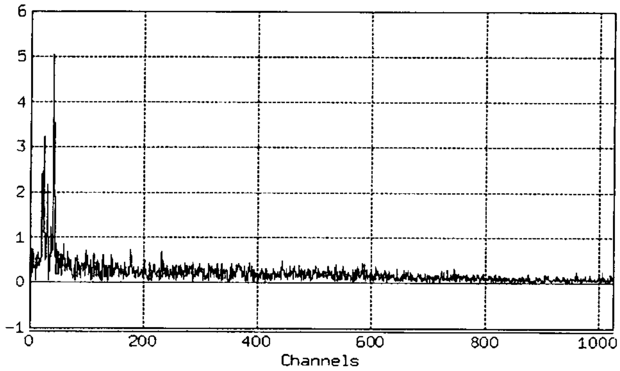

Fig. 7. Histogram of Fourier components in Fourier domain of above spectrum in Fig. 6.

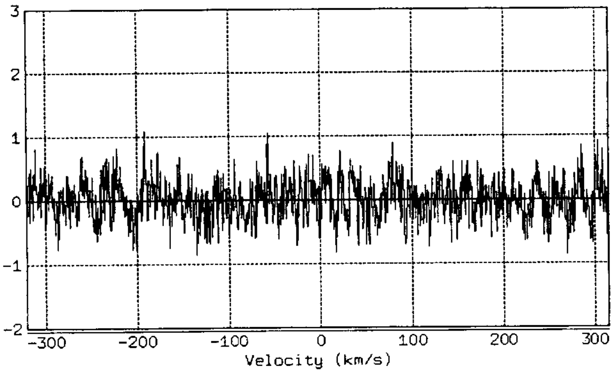

Fig. 8. Reduced spectrum after 3 most prominent ripple patterns are subtracted.

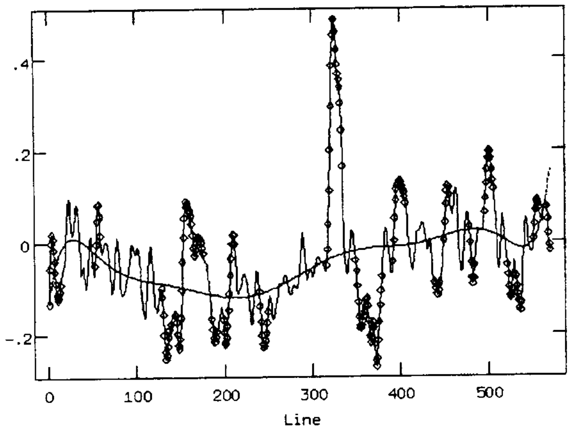

Fig. 10. An example of a spectrum with complex baseline. A 10th order Chevyshev function is fitted with pixels excluding the rejected pixels (marked with diamonds) below 3σ of noise level.

system, and thus, undetectable, but often significant when accumulated, ripple patterns are residing in almost all the spectral data. Thus, clear removal of the ripple patterns is very important process from a spectrum as well as a large amount of survey data. In *SPA*, this process can be done for each spectrum as discussed in Section II. We suggest to conduct this process in sequence for a stack of data, making a routine subtracting the ripple patterns in Fourier domain. In Fourier domain, the Fourier transform of the plotted channels of spectrum, then zeros out selected high amplitude channels, which are representing the ripple patterns in frequency domain, then re-transforms the selected channels and replaces them *in sequence*. This is a fast method of removing well-defined baseline ripple components. Several number of Fourier components can be zeroed out in this process. We may introduce the number of low-frequency components to ignore, since these may contain significant data. The data shown in Fig. 6 is transformed into Fourier components (see Fig. 7), and the brightest channels in the transform with channel number higher than a certain number are set to zero. The data is then re-transformed through inverse-Fourier transform (Fig. 8).

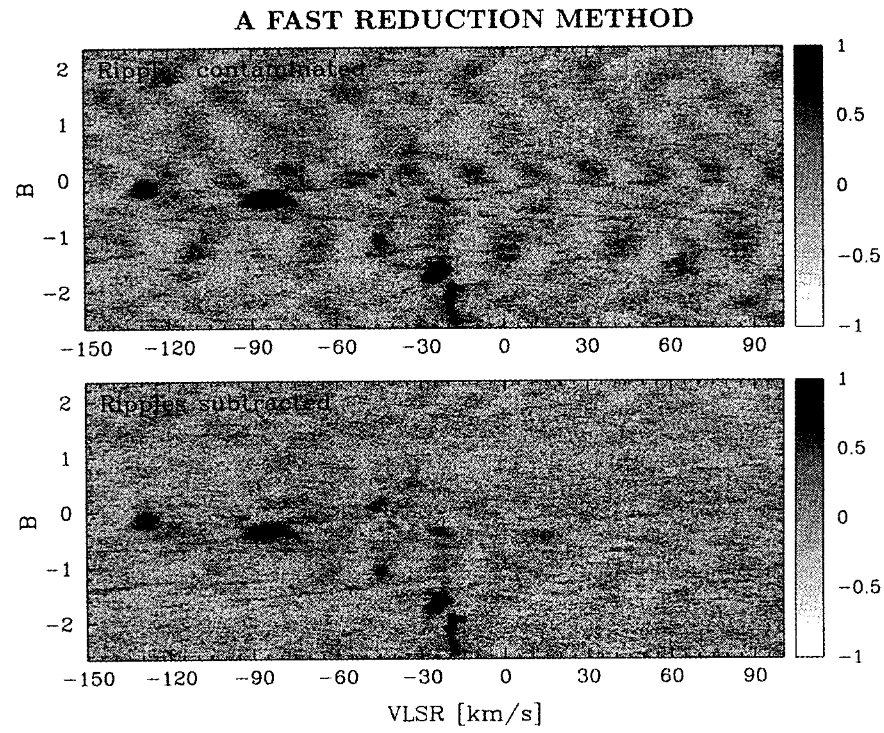

Fig. 9. Upper panel is an image heavily contaminated with ripple patterns, taken from Lane et al. (2000)'s B-strip survey using 1.7 meter telescope at Antarctic Submillimeter Telescope and Remote Observatory. Galactic longitude is $336^{\circ}.0$ and the observed line is CO(4-3). Lower panel is the reduced image after subtraction of 3 most prominent ripple patterns from each spectrum. Grey scale ranges from −1 to 1 K to make the ripple patterns noticeable.

If the single-precision data is converted to double precision and subsequent operations are all double precision, there is no loss of accuracy in the process of inverse-Fourier transform. Fig. 9 is an example of an image heavily contaminated with ripple patterns, taken from Lane et al. (2000)'s B-strip survey using 1.7 meter telescope at Antarctic Submillimeter Telescope and Remote Observatory. Galactic longitude is 336°.0 and the observed line is CO(4-3). Lower panel is the reduced image after subtraction of 3 most prominent ripple patterns from each spectrum. Applying this routine to the whole image, 2-dimensional or 3-dimensional in sequence, we were able to subtract the ripple patterns in an image successfully. In Section V, we'll discuss inclusion of this routine within IRAF.

(d) Baseline Subtraction

The most important reduction process is how to subtract the baseline properly. It may be a very good choice to use an IRAF task BACKGROUND for the baseline removing process, which is being used by optical astronomers handling spectroscopic data (twod.longslit), instead of using BASELINE in *SPA*. The major difference between BASELINE in *SPA* and BACKGROUND in IRAF is as follows: a batch job using BASELINE in *SPA* can be run using a fixed baseline range and a fixed polynomial. However, BACKGROUND task in IRAF allows us to conduct fitting each line or column interactively or in a batch job mode, with varying functions to the columns or lines specified by the sample parameter. This function is then subtracted from the entire line or column to create an output line or column. Before applying the functions to a spectrum, we can set up low and high rejection limits in units of the residual sigma of each spectrum. If low_reject and/or high_reject are greater than zero the sigma of the residuals between the fitted points and the fitted function is computed and those points whose residuals are less than low_reject × sigma and greater than high_reject × sigma are excluded from the fit. Points within a distance of grow pixels of a rejected pixel are also excluded from the fit. The function is then refit without the rejected points. This rejection procedure may be iterated a number of times given by the parameter 'n_iterate'. The fitting parameters (sample, naverage, function, order, low_reject, high_reject, niterate, grow) may be adjusted in interactive mode. Lines or columns from the image are selected to be fit with the 'icfit' package, which is other task package in IRAF. A single column or line may be chosen or a blank-separated range may be averaged. The actual image lines and columns are fit individually. The interactive cursor mode commands for this package are described in a separate help entry under 'icfit'. When an end-of-file or no line(s) or column(s) are given the last selected fitting parameters are used on each line or column of the image. This step can repeated for each image in the input list, or in each channel map. One example of a fitted baseline with 10th order Chevyshev function is presented in Fig. 10 with many rejected pixels for the fitting. The result after subtracting the fitted baseline is shown in Fig. 11.

IV. CUBE DATA ANALYSIS

When a cube data is successfully reduced and constructed, the next step would be analyzing it and deducting several physical parameters. For survey data, the first thing would be identifying process of individual clouds efficiently. A cloud can be defined as an object

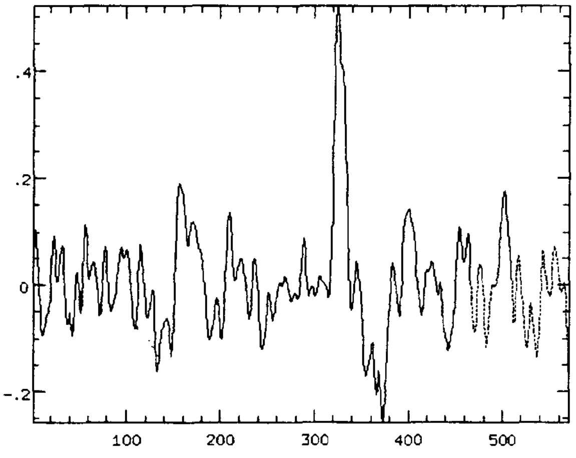

Fig. 11. Reduced spectrum of Fig. 10 after subtracting a 10th order Chevyshev function fitted.

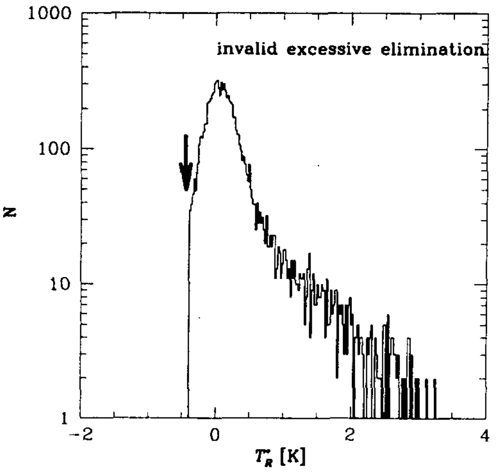

Fig. 12. Distribution of temperatures after eliminating negative bad pixels with an invalid threshold.

composed of all pixels in longitude, latitude, and velocity that are simply connected and that lie above some threshold temperature (Lee et al. 1997). Ideally, one would like to define clouds with a 0 K threshold temperature. However, low threshold temperatures are impractical in view of the noise level in the spectra and more importantly because of the blending of adjacent clouds which often occurs in crowded regions. On the other hand, with too high a threshold temperature, regions are severely truncated, and it is impossible to obtain a reliable estimate of the sizes and velocity dispersions, thus the related parameters of the clouds identified.

Several kinds of data set can be used for identifying clouds, such as ^{12}CO, ^{13}CO and HI cube data with proper header information. The most frequently used data format is FITS form. FITS data of an image usually includes the coordinate information with the first pixel values, and steps in three axes. Lee et al. (1997) used galactic coordinate system *v*, *l*, *b*, in order. If the cube data is not in *v*, *l*, *b* order, it is recommended to change it into this order as the code is designed in this way. If the different coordinate system is used, the code should be modified properly.

To make a FORTRAN program using IMFORT into a IRAF task, a CL foreign task interface is required to connect the program to CL callable task. We arbitrarily simplify the CL file for easy handling. One merit of making a task within IRAF is that every data can be handled in IRAF image form, which can be transformed to FITS form easily, and can be transported to other reduction packages, if necessary. Moreover, the identified clouds can be separately regenerated for further analysis within IRAF. When making a catalog of identified clouds, one may adjust the threshold pixel numbers, as one may not want to include the minimum number of pixels for defining a cloud. In any case, the minimum number of pixels in one axis is two. Thus, the minimum number of pixels of an identified cloud would be four. However, quite a large threshold pixel numbers is recommended for initial use of the code, and then choose a proper pixel number for more detailed study. Along with identification it is designed to identify clouds with their mean values of longitudes, latitudes, and velocities. After identifying the individual clouds, the dispersions of these parameters are also calculated. The clouds are given individual ID numbers, and all the pixels within the identified clouds are also given the same ID numbers as the cloud's. Thus, there would be no confusion with other clouds when estimating the dispersions of three axes. In addition, the total integrated intensity, the number of identified clouds, and the number of pixels involved are also estimated.

The cloud identification code can be applied to analyze the structure of giant molecular cloud. By applying several threshold temperatures to the target cloud, we may be able to see the hierarchical tree structure in detail, and the estimated physical parameters of subclouds can be statistically analyzed.

V. DISCUSSION

Our identification scheme of bad channels seems to be working well and the eliminating process is very fast comparing with classical method. Eliminating process can be done for any slice of a cube data in our method, not necessarily along the velocity direction. Usually, it has to be done along the velocity direction. However, care must be taken while applying to eliminate negative bad pixels. Negative bad pixel elimination with lower threshold than the excessive threshold temperature may cause false image with less negative emission pixels with more noisier positive emission pixels. The abnormal histogram of temperature distribution is shown in Fig. 12. The determination of a proper threshold for the excessive negative bad pixels is usually self-evident mostly in logarithmic scale histogram. Elimination of positive bad pixels is more complicated process as discussed above. However, this process can be solved if combination of image cursor ('imcur' within IRAF) and well-determined script code to identify and to interpolate the positive bad pixels, which is being developed. When completed, this process can be done just by clicking the mouse on the image displayer.

Fourier transformation routine has been developed in collaboration with A. A. Stark within *COMB*, a reduction package of Bell Laboratories 7 m telescope and Antarctic Submillimeter Telescope and Remote Telescope (AST/RO) during author's visiting Harvard-Smithsonian Center for Astrophysics in 1999. This routine seems to be working quite well for a 2-dimensional image. The next plan would be to revise this routine for up to 3-dimensional data cube, and to install it into IRAF.

BACKGROUND task is more efficient if there is more extent of baseline. Namely, with more number of channels involved, more probable fitting is expected. Although BACKGROUND task can be applied to a 3-dimensional image all at once, it is recommended to apply it to 2-dimensional image to make sure the process. When there is substantial emission around the edge of the image, a successful baseline fitting is not possible. In that case, low-order functions are recommended to fit the baseline, which seldom change the slope. When high-order function is applied, a small threshold values for the high and low rejection is recommended as higher order function can contaminate the emission itself.

VI. SUMMARY

We present a fast reduction method of survey data in radio astronomy. We focus on the spectral database obtained with a single-dish antenna. Classical data reduction method used at Taeduk Radio Astronomy Observatory is briefly reviewed. Identification and elimination of negative and positive bad channels can be done using cloud identification code previously developed and several IRAF tasks relating statistics. Removing of several ripple patterns is to be conducted in Fourier domain by getting rid of several prominent Fourier components. It is found that a BACKGROUND task within IRAF is very effective for fitting and subtraction of baseline with varying functions can be conducted using a BACKGROUND task within IRAF. Cloud identification method and the possibility of its manipulation for analysis of cloud structure is reviewed, and some future data reduction methods are discussed.

This work was supported by Creative Research Fund NZ-01-01-A-31 of Ministry of Science and Technology, and by grant No. 2000-1-11300-002-3 from the Basic Research Program of the Korea Science and Engineering Foundation.

REFERENCES

Brewer, M., & Heyer, M. H. 1992, *SPA* User's Manual, Five College Radio Astronomy Observatory, Amherst, MA, USA
Goldsmith, P.F., & Scoville, N.Z. 1980, A&A 82, 337
Heyer, M.H. & Carpenter, J.M. 1991, FCRAO Internal Memo
Lane, A., Lee, Y. & Stark, T. 2000, in preparation
Lee, Y., Jung, J.H. & Kim, H.G. 1997, JKAS, 12(1), 185
Lee, Y., Kim, H.G., Jung, J.H., Kim, B.G. & Ryu, O.K. 2001, in preparation