

РАЗРАБОТКА ФОРМАТОВ ХРАНЕНИЯ ДАННЫХ РАДИОЗОНДИРОВАНИЯ ИОНОСФЕРЫ

Щирый А.О.

Марийский государственный университет, г.Йошкар-Ола

saovu@mail.ru

Для сохранения оцифрованного сигнала радиозондирования ионосферы выбран формат WAV. Оцифрованный сигнал хранится аналогично звуковым данным; хранение параметров сеанса реализовано посредством chunk'a INFO. Для хранения ионограмм выбран формат PNG. Основные данные ионограммы (спектрограмма) помещены в виде, "готовом для просмотра" стандартными программами просмотра картинок; для хранения параметров сеанса используется секция zTXt; возможность прямого доступа к данным спектрограммы при этом достигается путем сохранения в zTXt значений смещений спектрограммы на изображении.

По данным радиозондирования ионосферы коротковолновыми (КВ) сигналами можно получить информацию о процессах в ионосферной плазме, о ее структуре и состоянии [1,2]. Данные радиозондирования ионосферы также крайне важны для систем КВ связи и загоризонтной радиолокации [2].

Результат работы ионозонда наклонного зондирования обычно представляется в виде ионограммы, являющейся растровым изображением зависимости времени группового запаздывания и амплитуды радиосигнала от частоты зондирования (рабочей частоты). Ионограмма формируется как спектрограмма из спектров мощности сигнала разностной частоты с выхода приемника ионозонда. При этом величина амплитуды в спектрах отображается цветом или яркостью.

Итогом систематической работы систем радиозондирования является большой объем экспериментальных данных, основными видами которых являются ионограммы и файлы оцифрованного сигнала с выхода приемного устройства ионозонда (вторые сохраняются, потому, что содержат полную информацию о принятом сигнале, в то время как при построении ионограмм происходит потеря информации, необходимой при исследовании тонких радиофизических эффектов).

Форматы хранения данных радиозондирования создавались разработчиками соответствующих программно-аппаратных средств. Например, в [3] описаны форматы данных ионозондов (разных поколений), созданных в ИСЗФ СО РАН (Иркутск). Автор хорошо знаком с форматами ионограмм ("IG2") и оцифрованных отсчетов ("SEQ"), разработанных в МарГТУ Егошиным А.Б. [4] и Батухтиным В.И. под руководством Шумаева В.В. Существуют и другие форматы.

Общим недостатком всех этих форматов является неуниверсальность, т.е. ориентация на конкретное программное (а иногда и аппаратное) обеспечение, выражающаяся в невозможности использования этих данных в существующем "стандартном" программном обеспечении (ПО) обработки данных (ПО обработки сигналов, математическом ПО, и т.п.); а также затруднении интеграции научного ПО, разработанного в разных научных коллективах (и использующего несовместимые форматы данных). С целью преодоления последнего недостатка, предпринимаются попытки использования формата XML (см. например [5]). Что нельзя признать хорошим решением, так как, во-первых, все-таки не решаются проблемы со "стандартным" ПО, и, во-вторых, формат XML является текстовым и не приспособлен для хранения бинарных данных (хранение бинарных данных в формате XML, вообще, возможно – например путем преобразования в печатные символы посредством схемы base64; но это приводит к увеличению размеров файлов в разы по сравнению с бинарными).

Поэтому автором были выбраны универсальные бинарные форматы, позволяющие устранить указанные недостатки. Однако, при этом пришлось решать задачи приспособления этих форматов для реализации некоторых несвойственных им функций.

Файл оцифрованного сигнала с выхода РПУ (WAV)

Для сохранения временного ряда оцифрованного сигнала с выхода радиоприемного устройства (РПУ) ионозонда был выбран формат WAV. Такой выбор был сделан для облегчения использования существующих программ цифровой обработки сигналов (например, SpectraLAB) и программ для цифровой обработки звука (например, SoundForge). Стоит также отметить, что популярный математический пакет MATLAB (ставший в своей нише буквально промышленным стандартом) позволяет работать с WAV файлами.

Формат WAV (являющийся частным случаем формата RIFF) допускает создание в файле различных дополнительных (“информационных”) секций. Эти секции можно использовать в т.ч. для хранения информации о параметрах сеанса зондирования.

Основной концепцией RIFF-формата является т.н. chunk — порция данных с заголовком; заголовок состоит из сигнатуры, указывающей на содержимое chunk’а, и размера данных chunk’а [6].

“Основные данные”, т.е. оцифрованный сигнал с выхода РПУ хранится аналогично звуковым данным (т.е. последовательность двух chunk’ов: "fmt " и "data" — как принято в “обычном” WAV файле для хранения звука).

После этих данных находится chunk с “нашим” заголовком. Это реализовано в виде стандартного chunk’а INFO, включающего стандартные chunk’и ICMT, INAM, ISBJ, ISFT, ISRF.

В целом структура файла выглядит так (условно):

```
chunk "WAVE" {
  chunk "fmt ";
  chunk "data";
  chunk "INFO" {
    chunk "ICMT";
    chunk "INAM";
    chunk "ISBJ";
    chunk "ISFT";
    chunk "ISRF";
  }
}
```

В chunk ICMT помещается поясняющая информация (комментарий) из одноименной колонки расписания (“Поясняющая информация”).

В chunk ISRF помещается заголовок с параметрами сеанса. Заголовок преобразован из бинарного вида в печатные символы. Это (преобразование) сделано потому, что программы редактирования звуковых файлов ожидают увидеть в chunk’е ISRF именно текстовую информацию, так что могут отказаться открывать файл с бинарным содержимым chunk’а ISRF или могут его испортить. В chunk ISRF помещается структура с параметрами сеанса, предварительно преобразованная в печатные символы.

В chunk ISBJ помещается полное строковое наименование трассы. Хотя это название можно сформировать (по координатам пунктов, хранящимся в заголовке в ISRF и названиям пунктов, хранящимся в файле описания приемных/передающих пунктов), это сделано на случай если будет утерян файл пунктов.

Chunk INAM зарезервирован для будущего использования.

Формат файла описания пунктов

Упомянутый выше файл пунктов (по умолч. "stations.txt") является служебным файлом, а не файлом данных результатов радиозондирования, и представляет собой текстовый файл,

строки которого являются записями о пунктах (кроме строк, начинающихся с символа # — это комментарий).

Каждая строка-запись имеет следующий формат:

<широта> <долгота> <"полное название пункта"> <сокr_nazv_punkta> [GPS]

где <широта>, <долгота> — указываются в градусах (т.е. если требуется указать еще и минуты, то переводить в доли градуса, десятичный разделитель — точка); положительные значения задают, соответственно, сев.ш., вост.д., отрицательные — юж.ш., зап.д.;

<"полное название пункта"> — полное название пункта в кавычках, внутри кавычек допустимы любые печатные символы;

сокr_nazv_punkta — сокращенное название пункта без кавычек, без пробелов, желательно латинские буквы (т.к. сокr_nazv_punkta предназначено для автоматического формирования имен файлов данных), рекомендуется 3-4 симв. (но не обязательно), значимы первые 8 символов (пример значений: yola, mosk, nnoy, kazan, kpr, inskip).

Последний столбец (“[GPS]”) — флаг "точности", возможные значения (без кавычек):

GPS — означает что координаты точные (“получены с GPS”);

~ — означает что координаты известны приблизительно (не по GPS, а по карте или литературе, например);

? — означает что координаты известны предположительно, т.е. нет уверенности в том, что это вообще именно такой пункт.

Для пунктов, помеченных <?>, но ставших впоследствии известными (пеленгация, данные радиолюбителей и др.) предполагается вести список соответствий в отдельном файле (чтоб не исправлять флаг точности в основном файле пунктов).

При поиске названия пункта по координатам, совпавшими считаются координаты, если разница между ними (координатами в запросе и координатами в записи базы) не более 1 град. Если же подходящих записей больше одной, то среди них ищется запись с наименьшей разницей координат.

При старте программы регистрации и первичной обработки, GPS дает координаты приемного пункта, по координатам определяется название пункта: оно ищется в базе, если нет совпадений, то предлагается ввод названия приемного пункта; если нет GPS (т.е. данная аппаратная реализация ионозонда не имеет GPS), то сначала предлагается и ввод координат.

В расписании (по которому работает ПО сбора данных в автоматическом режиме, т.е. без непосредственного участия человека) задаются только координаты (по которым находится и отображается название) передающего пункта (для неизвестных пунктов задаются неопределенные координаты, отображающиеся как "Неизвестен"). В файлах данных (временные ряды, ионограммы) сохраняются координаты и начального, и конечного пунктов.

Файл ионограмм (PNG)

Для хранения ионограмм выбран формат PNG.

Интересно отметить, что рассматривался также формат GIF, показавший лучшие результаты (меньший размер файла) при сохранении ионограммы с точностью 4 бита на пиксель (16 цветов), чем PNG при идентичных условиях. Однако, GIF “проиграл” при сохранении ионограммы с точностью 8 бит на пиксель (256 цветов), а именно такую точность решено было использовать.

Формат JPG не рассматривался, т.к. это формат с потерей данных.

Случай сохранения ионограммы с точностью 4 бита на пиксель (16 цветов) проверялся потому, что, как правило, ионограмма представляется в определенной 11-цветной палитре (и только при создании картинок для печати требуется серая палитра). Однако, было принято решение для повышения точности представления амплитуды сигнала использовать 8 бит на пиксель. Поэтому, для сохранения видимости тех же 11 цветов (ставшими стандартом де-факто в определенном кругу специалистов) приходится формировать 256-цветную палитру с

11-ю диапазонами одинаковых цветов, т.е. количество одинаковых цветов в каждом диапазоне палитры примерно равно 256/11, а точнее: все диапазоны палитры, кроме последнего, имеют размер по 23 цвета, последний имеет размер 26 цветов. При использовании такой палитры остается возможность легко перейти к серой 256-цветной палитре, и, что важно, не менять при этом значения самих пикселей — достаточно только смены палитры.

Основные данные ионограммы (спектрограмма спектров мощности сигнала с выхода приемника ионозонда — зависимость амплитуды сигнала от частоты зондирования и времени группового запаздывания радиоволн) помещены в виде, “готовом для просмотра” стандартными программами просмотра картинок, т.е. имеются оси, надписи и пр. Поэтому для получения программного доступа к основным данным ионограммы, в заголовке записываются параметры, означающие сдвиг информационного поля основных данных ионограммы относительно верхнего левого угла изображения; учет этих сдвигов позволяет работать с данными ионограммы как с двумерным массивом.

Для хранения параметров сеанса измерения используется опциональная секция файла PNG, называемая zTXt (подробнее о секции zTXt см. стандарт PNG: “11.3.4.4 zTXt Compressed textual data” в [7]). В эту секцию помещается структура с параметрами сеанса, предварительно преобразованная в печатные символы. Такое преобразование делается потому, что стандартная программа просмотра изображений Windows не поддерживает PNG формат полностью, т.е. считает ошибочными дополнительные бинарные секции, несмотря на то, что они валидны с точки зрения стандарта PNG [7] (создание дополнительных бинарных секций было опробовано и такие PNG файлы просматриваются большинством программ просмотра и обработки изображений, к каковым не относится упомянутая стандартная программа просмотра изображений Windows; так что от такого — самого естественного — пути пришлось отказаться и помещать параметры сеанса в секцию zTXt преобразованными в печатные символы, как это сказано выше).

Последняя модификация формата ионограмм допускает также опциональное сохранение фазового спектра. При этом, аналогично спектру мощности, фазовый спектр сохраняется как еще одна прямоугольная область на PNG “картинке” ионограммы, а в заголовке хранятся ее размер и смещение для программного доступа к ней как к двумерному массиву данных.

Литература

1. *Филипп Н.Д., Блаунштейн Н.Ш., Ерухимов Л.М., Иванов В.А., Урядов В.П. // Современные методы исследования динамических процессов в ионосфере. — Кишинев: Штиинца, 1991. — 286 с.*
2. *Щирый А.О. // Разработка и моделирование алгоритмов автоматического измерения характеристик ионосферных коротковолновых радиолоний: Автореф. дис. ... канд. техн. наук: Спец. 05.12.04; Санкт-Петербургский гос. ун-т телекоммуникаций им. проф. М.А. Бонч-Бруевича. — СПб., 2007. — 19 с.*
3. *Ким А.Г., Матюшонок С.М. // Структура и формат данных ЛЧМ-ионозонда ИСЗФ СО РАН: Препринт №8-06. — Иркутск: Институт солнечно-земной физики СО РАН, 2006. — 12 с.*
4. *Егошин А.Б. // Автоматизированная система адаптивной обработки сигналов со сверхбольшой базой для радиозондирования ионосферных радиолоний: Автореф. дис. канд. техн. наук: Спец. 05.12.04; Марийский государственный технический университет. — Йошкар-Ола, 2003. — 25 с.*
5. *Богомаз А.В., Козлов С.С., Пуляев В.А. // О возможности применения формата XML для хранения ионосферных данных. Дистанционное радиозондирование ионосферы: Сб. тезисов докладов. — Харьков: Институт ионосферы МОН и НАН Украины, 2010. — с.48.*

6. RIFF [Электронный ресурс] // Wikipedia. – Режим доступа: <http://ru.wikipedia.org/wiki/RIFF>
7. Portable Network Graphics (PNG) Specification (Second Edition) – ISO/IEC 15948:2003 [Электронный ресурс] // W3C. – Режим доступа: <http://www.w3.org/TR/PNG/>