

Revisiones	Fecha	Comentarios
0	28/11/06	
1	02/02/07	Ampliación
2	28/01/11	actualización; aP89W24, ampliación: agrega familia aIVR

Presentamos en esta oportunidad una familia de chips para reproducción de voz. Estos chips, por oposición a los tradicionales grabadores/reproductores, son solamente reproductores, la voz o sonido se graba una vez, y luego se los utiliza para reproducción solamente. El proceso de grabación es similar al de una PROM o EPROM, en el cual la voz grabada se incorpora al dispositivo en tiempo no-real, es decir, se genera un archivo que luego se envía al chip. El hecho de ser programables sólo una vez (OTP), y que su proceso de grabación sea como una EPROM, permite que la arquitectura interna sea mucho más simple, lo cual, a su vez, hace que su precio sea mucho más accesible.

Una característica distintiva de esta familia de chips, es que no sólo pueden grabarse frases, sino que la grabación se efectúa por secciones. Luego, estas secciones se agrupan formando grupos, los cuales son seleccionados para reproducir la frase que se necesita. Por ejemplo, es posible grabar una sección con la palabra “acceso”, otra con “no”, y otra con “autorizado”. Luego se formarán dos grupos, uno con las secciones 1 y 3, y el otro con las secciones 1, 2 y 3. El primer grupo, entonces, hará oír la frase “acceso autorizado”, mientras que en el segundo caso será “acceso no autorizado”. De esta forma, no sólo es posible economizar el espacio necesario para las frases que se intenta reproducir, sino que no es necesario generar una compleja tabla con las direcciones en memoria del chip donde comienza cada una de las palabras o frases e ir alternándolas durante el proceso de “habla”. En este caso, simplemente informando al chip la intención de reproducir el grupo 2, el mismo localiza, direcciona, y reproduce, las secciones 1, 2 y 3, que forman parte del grupo.

Este tipo de dispositivos encuentra particular aplicación en sistemas donde se utiliza la voz para guiar o reportar al usuario, permitiendo la compaginación de frases a partir de palabras fundamentales, y sin que sea necesario grabar la voz de éste o cualquier otra para el funcionamiento.

## Grabación de la voz

El procedimiento de grabación de la voz es algo más complicado que el presionar un botón y hablar delante de un micrófono; pero si pensamos en todo el trabajo que se simplifica después para la generación de frases a partir de palabras principales, resulta sencillo. Por supuesto que si solamente necesitamos decir una frase, esta tal vez no sea la mejor opción, pero sigue siendo válida, particularmente por el costo más bajo de la implementación final.

Para grabar una frase, deberemos hacerlo en el sistema que más nos agrade. Desde ya los músicos tendrán el sistema de su agrado, pero el resto de los mortales generalmente dispone de una placa de sonido en su computadora de trabajo, y un micrófono. Esto es más que suficiente.

La primera decisión a tomar, que puede corregirse luego, es la velocidad de muestreo. Estos chips graban la voz de la siguiente forma:

	<i>One Time Voice</i>	<i>Instant Voice</i>
PCM 8-bits <sup>1</sup>	X	X
ADPCM 4-bits <sup>2</sup>	X	X
PCM 5-bits, ley $\mu$ <sup>3</sup>		X

Si bien la documentación menciona 6Ksps, digamos que la calidad de la voz no es del todo óptima a esta velocidad. Si partimos del hecho de que un teléfono fijo es PCM a 8Ksps, vemos que 8K nos va a dar una voz

1 Pulse Code Modulation, asignación lineal de magnitud analógica a número binario.

2 Adaptive Differential Pulse Code Modulation, descripta en ITU-T G.726 (anteriormente G.721)

3 PCM con función de transferencia alineal (compresión), descripta en ITU-T G.711

algo “telefónica”, pero sumamente utilizable<sup>4</sup>. El proceso de compresión ADPCM deteriora un poco la calidad, por lo que si se busca una voz de una calidad más “radial”, deberemos pensar en 16K.

En el caso que no estemos muy seguros, podemos grabar a una velocidad alta y luego hacer una conversión de velocidad de muestreo. Es conveniente realizar un buen filtrado para evitar la aparición de componentes no deseadas (aliasing).

Una vez obtenidas todas las palabras necesarias, las separamos una a una en archivos, los cuales grabaremos en el formato WAV en 8-bits mono<sup>5</sup>. Por ejemplo, en el caso anterior, grabamos “acceso no autorizado”, y generamos tres archivos VOC, uno con la palabra “acceso”, otro con “no” y el último con “autorizado”. La edición, conversión de velocidad de muestreo, y grabación en formato VOC, puede realizarse con programas de tipo shareware como Goldwave, o programas profesionales como CoolEdit, por ejemplo, a elección del usuario. Si bien esto parece algo difícil para quien no tiene experiencia en este tipo de programas, la identificación de las palabras resulta simple luego de familiarizarse con la visualización. Si se habla de forma pausada, se observan claramente los silencios entre las palabras, los cuales pueden borrarse o minimizarse para economizar tiempo de chip y hacer el habla más fluida.

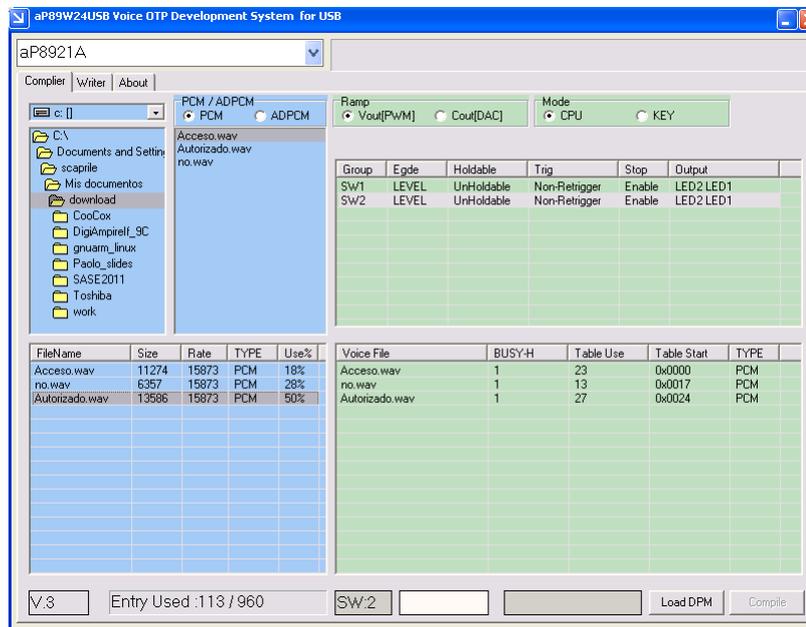
Desde ya que si esto resulta complicado, el lector puede optar por grabar las palabras de a una; la razón por la que sugerimos la grabación de la frase es porque la entonación resulta más natural.

### Armado de las frases

Una vez grabadas todas las palabras en sus correspondientes archivos, utilizaremos el software provisto por el fabricante para el armado de los grupos, conteniendo las frases. Según el chip que utilizemos, tendremos una capacidad de memoria que nos permitirá incluir más o menos tiempo, pero por lo general dentro de una misma familia, todos tienen el mismo número de secciones y grupos.

Seleccionamos las palabras y armamos los grupos. En el ejemplo que venimos desarrollando, la primera sección tendrá el archivo que contiene la palabra “acceso”, la segunda el que contiene la palabra “no”, y la tercera el que contiene la palabra “autorizado”. Luego, el grupo 1 contendrá las secciones 1 y 3; y el grupo 2 contendrá las secciones 1, 2 y 3.

La imagen siguiente es una captura de pantalla que muestra esta operación en el software para aP89W24, grabador de chips de la familia One Time Voice:

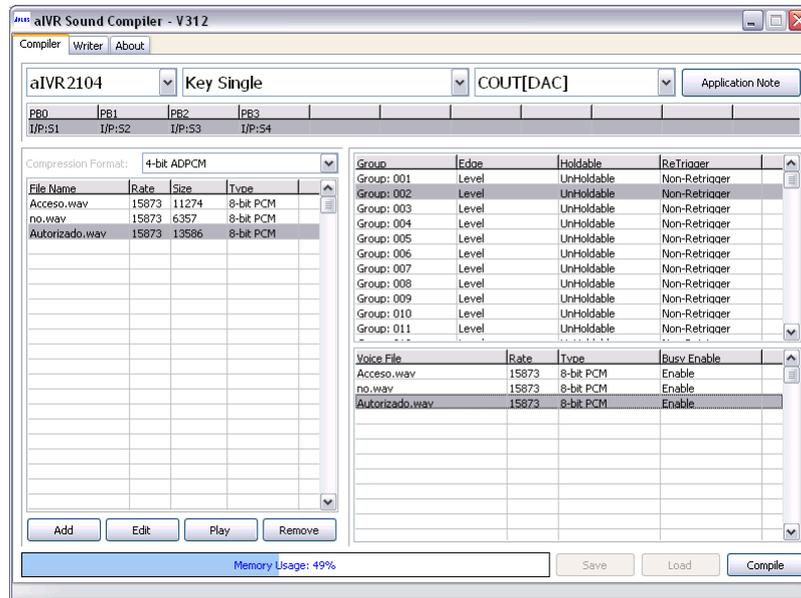


- 4 La telefonía analógica en Argentina utiliza G.711 ley-A, que consiste en tomar muestras de la voz a 8Ksps en 13-bits y codificarlas para ser transmitidas en 8-bits a 8Ksps (64Kbps). El formato PCM 8-bits presenta, por consiguiente, igual o menor “calidad” de voz que la telefonía analógica.
- 5 El software para la familia One Time Voice soporta además el formato VOC, el cual emplea ADPCM, y era el único soportado en versiones anteriores.

## CTC-042, Chips OTP para reproducción de voz de Aplus

En este programa, elegimos el formato a utilizar (PCM o ADPCM) al momento de abrir el archivo, seleccionando el botón que se encuentra sobre la lista de archivos disponibles.

La siguiente captura corresponde al software aIVR Sound Compiler, que soporta al aIVR20W, grabador de chips de la familia Instant Voice:



En este programa agregamos archivos presionando 'Add'. Los cargamos inicialmente en PCM, y podemos convertirlos a ADPCM o ley- $\mu$  presionando 'Edit', debajo del listado de archivos en uso.

Por supuesto que si esta operatoria de armado de frases resulta complicada, siempre es posible grabar una frase en una sección y asignar un grupo a una sección. Sin embargo, el modo sugerido permite optimizar el uso de la memoria y lograr un tiempo total de habla mayor; por ejemplo, si grabáramos "acceso autorizado" y "acceso no autorizado", estaríamos utilizando casi el doble de la memoria que utilizamos en el modo sugerido.

La cantidad de grupos y la duración máxima dependen de cada chip, por ejemplo el aP8921A permite 12 grupos y 21 segundos a 6Ksps en ADPCM; mientras que el aP8942A permite 35 grupos y 42 segundos. En la familia Instant Voice, por ejemplo el aIVR2104 permite 13 grupos y 21 segundos; mientras que el aIVR4208 permite 57 grupos y 42 segundos.

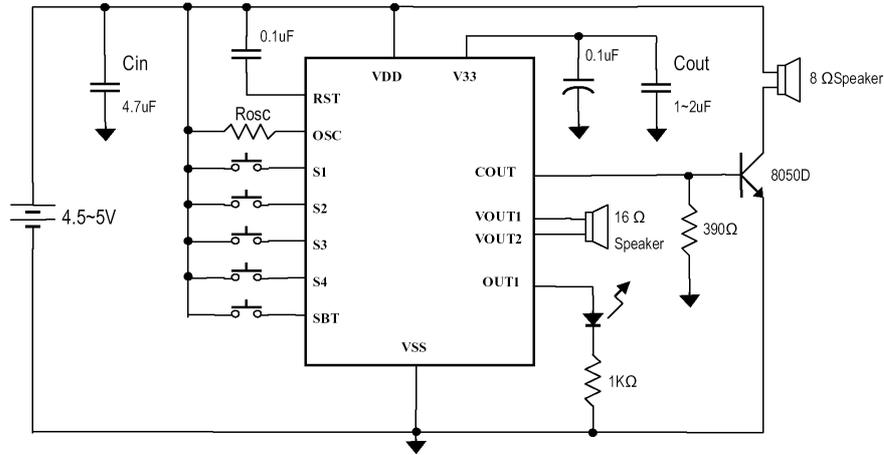
## Utilización

Antes de proceder a la grabación del chip, deberemos decidir si lo utilizaremos con una opción de bajo costo en la cual el chip genera PWM que excita directamente un resonador piezoeléctrico, o la opción de una calidad un poco superior en que un DAC interno genera una corriente que puede utilizarse para controlar un transistor o convertir a tensión y controlar un amplificador. De igual modo, deberemos decidir si la activación se realiza del modo "botones", en el cual la activación de un pin de selección de grupo ocasiona la reproducción del mismo, o en el modo "micro", en el cual estos pines forman un código binario representando el número de grupo, y otro pin realiza la indicación de comenzar la acción.

Grabado el chip en el modo correspondiente, el mismo responderá a este modo, permitiendo con una simple acción como presionar un botón, la selección de una de varias frases, o mediante un micro, informar al usuario la condición requerida. En el caso de ejemplo que venimos desarrollando, ésta será el hecho de si se permite o no su ingreso, al presentar éste su identificación electrónica a nuestro lector.

El circuito esquemático a continuación corresponde a una utilización típica del aP8921A. Para el ejemplo que venimos desarrollando, a 16K, el valor del resistor  $R_{OSC}$  es de 133K :

## CTC-042, Chips OTP para reproducción de voz de Aplus



Este chip funciona tanto a 3V como a 5V, pero existen otros en esta familia que son sólo del rango de 3V.

En el caso de la familia Instant Voice, del rango de 3V, las hojas de datos disponibles en la web del fabricante al momento de escribir esta nota son una versión reducida. La información necesaria para diseño puede obtenerse presionando el botón 'Application Note' en el programa de grabación. Dado que este programa verifica el firmware del grabador, es necesario tenerlo conectado para acceder a esta información.

El ejemplo a continuación corresponde a un aIVR2104. En este caso, el fabricante nos especifica que el resistor *Rosc* con el valor especificado soporta hasta 14Ksps. Para 16Ksps deberemos utilizar un valor menor. En las pruebas de laboratorio utilizamos 580K, obteniendo una calidad de sonido sumamente aceptable en ADPCM de 4-bits.

