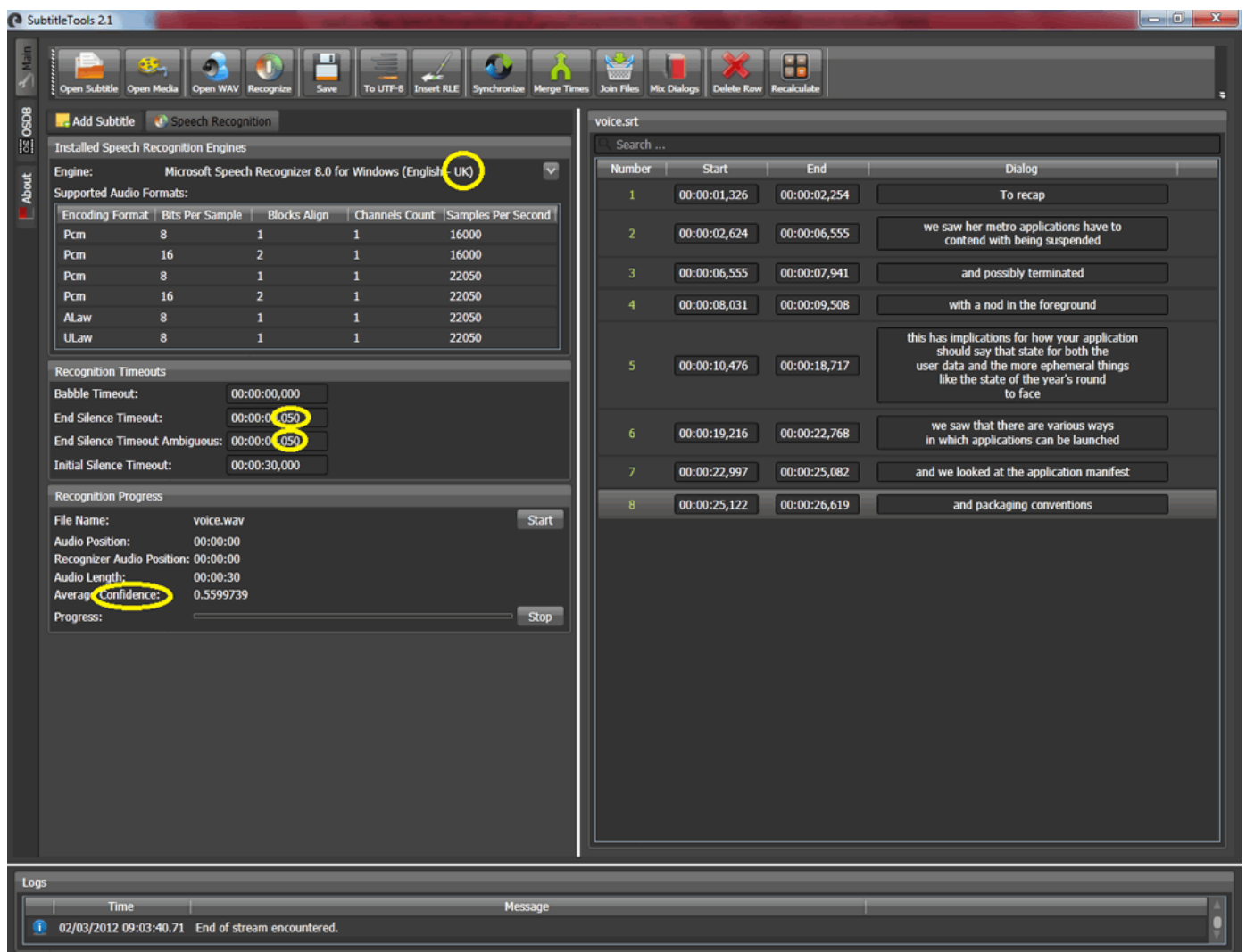


از ویندوز ویستا به بعد، ویندوز به صورت توکار دارای یک موتور تشخیص صدا شده است که در این مسیر قابل مشاهده می‌باشد:

Control Panel\Ease of Access\Speech Recognition

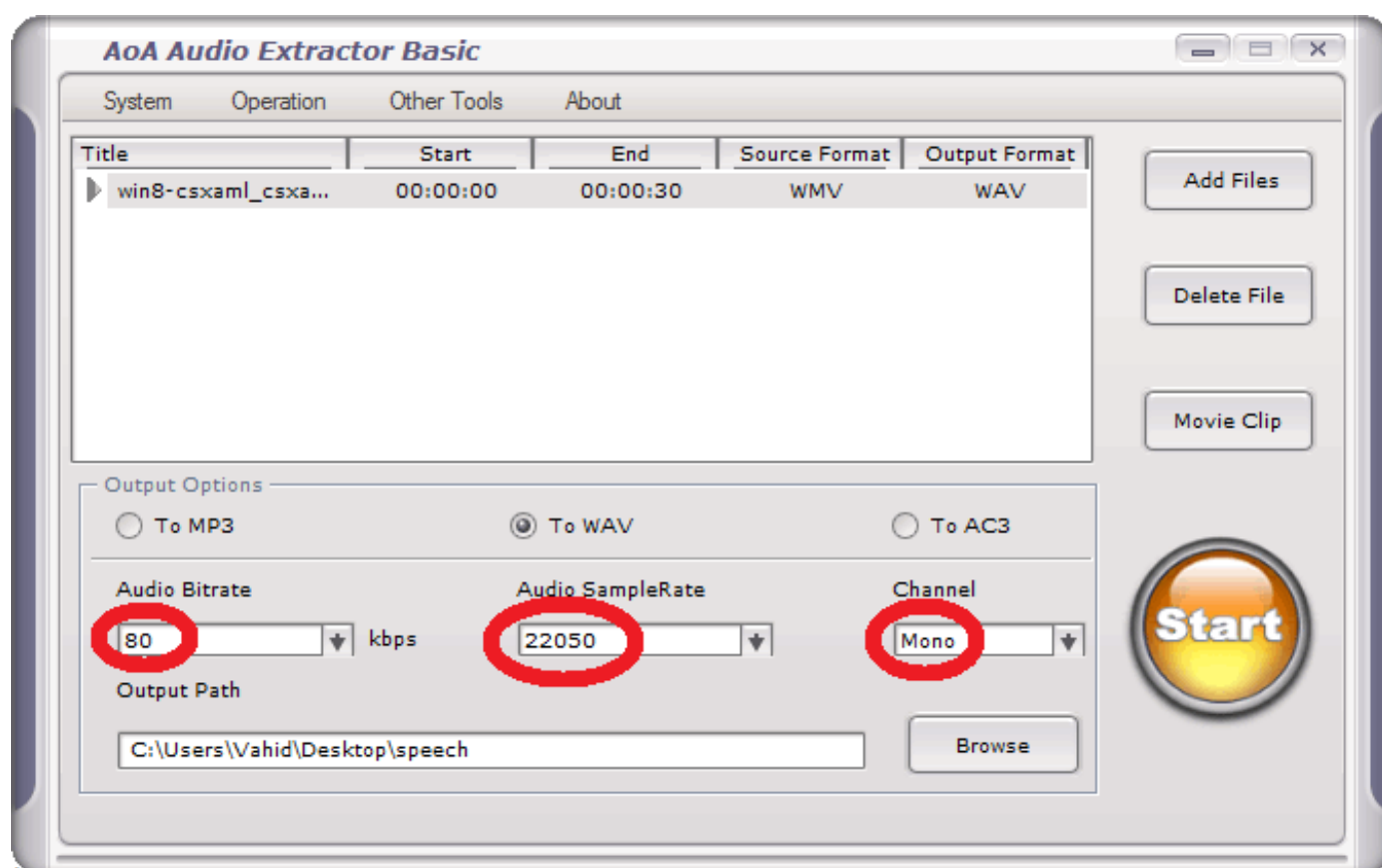
این سرویس از طریق اسمبلی استاندارد System.Speech در دات نت فریم ورک قابل استفاده است که اکنون با برنامه‌ی [Subtitle tools](#) یکپارچه شده است.



یکی از خصوصیات مفید این موتور تشخیص صدا، امکان دریافت فایل‌های صوتی نیز می‌باشد. فایل صوتی دریافتی باید مطابق یکی از فرمت‌های پشتیبانی شده توسط آن، تهیه شود؛ که این مورد را ذیل قسمت Supported audio formats شکل فوق می‌توانید مشاهده کنید.

برای نمونه توسط برنامه [AoA Audio Extractor Basic](#)، می‌توان این تبدیلات را انجام داد و یکی از تنظیمات قابل قبول توسط

موتور Speech Recognition ویندوز 7 را در تصویر ذیل می‌توانید مشاهده کنید: (و در غیراینصورت هیچ خروجی را نخواهید گرفت؛ خیلی مهم!)



پس از انتخاب و گشودن فایل صوتی در برنامه Subtitle tools (کلیک بر روی دکمه Open WAV در اینجا) و سپس کلیک بر روی دکمه‌ی Recognize یا Start، کار موتور Speech Recognition ویندوز شروع شده و برنامه هم در اینجا از فرصت استفاده کرده و دریافتی نهایی را تبدیل به رکوردهای فایل زیرنویس می‌کند که نمونه‌ای از آن‌را در شکل فوق می‌توانید ملاحظه کنید.

نکاتی در مورد استفاده بهینه از موتور تشخیص صدای ویندوز:

الف) برای آزمایش برنامه، یک فایل voice را [از اینجا](#) دریافت کنید. این فایل voice از همان سری مترو PluralSight تهیه شده است. ابتدا موتور تشخیص صدای انتخابی را بر روی حالت US قرار داده و تست کنید. در ادامه یکبار هم بر روی حالت UK قرار دهید و کار تشخیص صدا را آغاز نمایید. نتایج کاملاً متفاوت خواهند بود و با توجه به لهجه انگلیسی گوینده، تشخیص‌های حالت UK، به واقعیت نزدیکتر هستند. این مورد را در گزینه‌ی Average confidence هم می‌توانید مشاهده نمایید. مثلاً در اینجا موتور تشخیص صدا در کل به 60 درصد خروجی تولیدی‌اش اطمینان دارد و مابقی ... آنچنان اعتباری ندارند. مثلاً متن صحیح سطر چهارم در تصویر فوق باید «when they are not in the foreground» باشد!

ب) تنظیمات Timeouts

اگر به فایل voice فوق دقت کنید، گوینده یک نفس از ابتدا تا انتها صحبت می‌کند. اینجا است که به کمک مقادیر Silence timeout، می‌توان تعداد رکوردها را بر اساس فواصل تنفس کوتاهتری، بیشتر کرد. مثلاً با اعداد پیش فرض سیستم، با فایل صوتی فوق به 5 خروجی خواهید رسید؛ اما با توجه به تنظیماتی که در تصویر مشاهده می‌کنید، به 8 خروجی متعادل‌تر می‌رسیم.

مزایا:

کار زمانبندی زیر نویس خودکار می‌شود.
تا حدود 60 درصد، خروجی متنی مطمئنی را می‌توان شاهد بود.

در مورد ویندوز XP :

ویندوز XP به صورت پیش فرض دارای موتور Speech Recognition نیست. دو راه برای نصب آن در این سیستم وجود دارد:

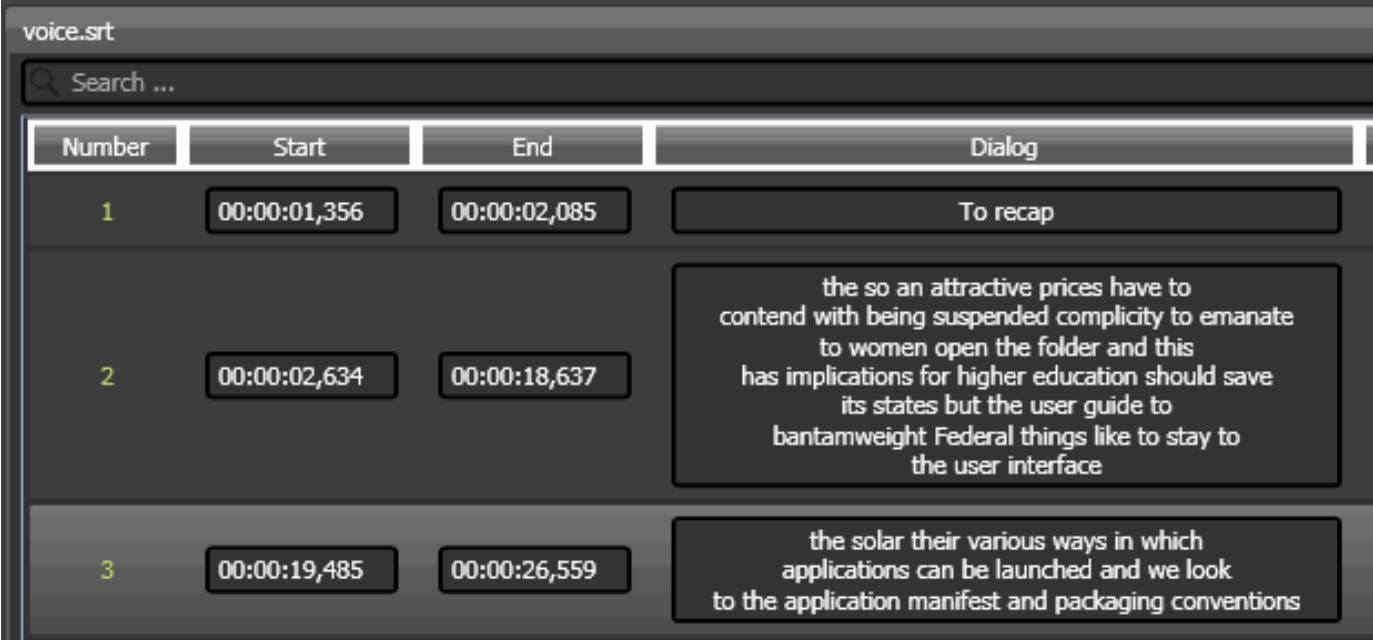
الف) استفاده از بسته نرم افزاری آفیس XP

به کنترل پنل مراجعه کرده، گزینه‌ی Add/remove programs را انتخاب نمائید. در اینجا Microsoft Office XP را انتخاب و بر روی دکمه Change کلیک کنید. نیاز است تا یکی از ویژگی‌های نصب نشده آن را نصب کنیم. به همین جهت در صفحه ظاهر شده، Add or Remove Features را انتخاب و در ادامه در قسمت Features to install ، گزینه‌ی Office Shared Features را انتخاب کنید. ذیل مدخل Alternative User Input، امکان انتخاب و نصب Speech مهیا است.

ب) استفاده از Microsoft Speech SDK Setup 5.1

بر روی ویندوز 7، نگارش 8 این برنامه نصب است؛ اما برای ویندوز XP تا نگارش 5.1 بیشتر ارائه نشده است. فایل‌های آن را [از اینجا](#) می‌توانید دریافت کنید. نصب آن هم [در اینجا](#) توضیح داده شده.

من در کل ویندوز XP را برای اینکار توصیه نمی‌کنم چون هم موتور تشخیص صدای آن قدیمی است و هم حالت Asynchronous آن درست کار نمی‌کند. برای مثال این یک خروجی تهیه شده از همان فایل voice فوق، توسط موتور تشخیص صدای مخصوص ویندوز XP است که بی‌شبهت به طنز نیست!



| Number | Start | End | Dialog |
|--------|--------------|--------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1 | 00:00:01,356 | 00:00:02,085 | To recap |
| 2 | 00:00:02,634 | 00:00:18,637 | the so an attractive prices have to contend with being suspended complicity to emanate to women open the folder and this has implications for higher education should save its states but the user guide to bantamweight Federal things like to stay to the user interface |
| 3 | 00:00:19,485 | 00:00:26,559 | the solar their various ways in which applications can be launched and we look to the application manifest and packaging conventions |

نظرات خوانندگان

نویسنده: Farhad Yazdan-Panah

تاریخ: ۱۳۹۰/۱۱/۱۵ ۰۱:۲۶:۱۶

عالی بود.

نمیشه از موتور تشخیص گفتار وب سایت youtube استفاده کرد؟ (وب سرویسی یا .. موجود نیست؟)
به دلیل حجم بالای آموزشش احتمالا نتیجه بهتری داره.
در ضمن وابسته به یه پلتفرم خاص هم نمیشه.

نویسنده: A. Karimi

تاریخ: ۱۳۹۰/۱۱/۱۵ ۰۲:۵۵:۵۹

خیلی ایده جالبیه! اگر مترجم گوگل هم به این کار اضافه بشه، یک ابزار کمکی خوب میشه اما خط به خطش نیاز به بازبینی داره (اشتباهات تشخیص گفتار ضرب در اشتباهات مترجم). ولی میشه به چند سال دیگه امیدوار بود.

نویسنده: وحید نصیری

تاریخ: ۱۳۹۰/۱۱/۱۵ ۲۰:۴۹:۰۶

بله، ممکن است. ولی به دو دلیل من ترجیح میدم به این سمت نروم:
- این مقدار پهنای باند آپلود در اختیار ندارم. (اینترنت حجمی ...)
- خیلی از این مباحث کپی رایت دارند و نمی‌شود به همین سادگی این‌ها را عمومی کرد، تست کرد و غیره.

نویسنده: Farhad Yazdan-Panah

تاریخ: ۱۳۹۰/۱۱/۱۵ ۲۲:۴۲:۴۹

حرف شما صحیحه.

فقط چند نکته:

من با اجازه خودتون یه سری آزمایش انجام دادم.

اول یه فایل رو به برنامه دادم تا پردازش کنه (نتایج A)

سپس همون فایل رو با یه برنامه پردازش صوت (Adobe Audition CS5.5) پردازش کردم (فقط از فیلتر Normalizer استفاده کردم تا سطح صدا بهتر بشه) جالب اینه که نتیجه (در حد 3-4 کلمه در - هر 30-40 ثانیه) بهتر شد.

یه بار دیگه با استفاده از امکان Settings موجود در Windows Media Player Play Speed فایل رو کش آوردم (از نظر زمانی)

بازهم نتایج عوض شد.

در صورت 0.8 (البته مطمئنا این عدد برای موارد مختلف متفاوت خواهد بود) نتیجه خارق العاده بود. تقریبا در 30 ثانیه 3-4 اشتباه داشت. (نسبت به 10-12 اشتباه).

قطعا شخص من کسی نیست که بتونه با شما در یک پروژه همکاری کنه، ولی به عنوان یه پیشنهاد فکر می‌کنم که اگه بتونید اطلاعات ورودی (فایل صوتی) رو با چند حالت پردازش کنید (سرعت بالاتر، پایینتر، نرمالایز شده و ...) فکر میکنم با فیوژن نتایج با هم میشه عدد 60 درصد رو به مقداری بالاتر رسوند.

ناگفته نمونه که فایل من لهجه US بود.