



### تشخیص خودکار گفتار فارسی

برقراری ارتباط گفتاری با کامپیوترها به جای استفاده از صفحه کلید و ماوس یکی از زمینه‌های تحقیقاتی مهم چند دهه‌ی اخیر بوده است و شرکت‌های بزرگی چون Philips، ALIT، IBM و Microsoft سالانه هزینه‌های هنگفتی را برای این منظور پرداخت کرده و می‌کنند. به عنوان یک کاربر کامپیوتر، احتمالاً با قابلیت گفتاری مجموعه آفیس به عنوان یکی از ویژگی‌های جذاب و تا حدی فانتزی برخورد کرده و یا با آن کار کرده‌اید. به کمک این قابلیت شما به جای استفاده از صفحه کلید برای تایپ مطالبتان، به راحتی با خواندن متن مورد نظر و انتقال گفتارتان به کمک یک میکروفون معمولی به کامپیوتر، آنرا در محیط Word تایپ شده می‌بینید. حتی برای ذخیره کردن، کپی کردن، گذاشتن عکس در متن و... به جای کلیک‌های پشت سر هم و گاهی با تعداد بالا، می‌توانید فرمان مربوطه را به کمک گفتار به نرم‌افزار داده تا کار شما را انجام دهد. جدای از اینکه توانایی درست کارکردن این قابلیت آفیس چقدر باشد، یک محدودیت بزرگ در سر راه استفاده از آن برای ما ایرانیان وجود دارد: این قابلیت فقط برای زبان انگلیسی است. (البته زبان‌های چینی یا ژاپنی را نیز می‌توان استفاده کرد!).

ایجاد چینی سیستمی که آنرا تشخیص یا بازشناسی گفتار (Speech recognition) می‌نامند، در زبان فارسی، چندین سال از تحقیقات محققان، اساتید و دانشجویان دانشگاه‌های مختلف کشور را به خود اختصاص داده است. اما جدی‌ترین جهشی که در حدود ده سال پیش در این زمینه ایجاد شد، ایجاد دادگان گفتاری فارسی و یک سیستم اولیه تشخیص گفتار فارسی در مرکز هوشمند علائم بوده است. در ادامه و در طی یکی-دو سال اخیر، مهم‌ترین دستاورد در این تکنولوژی برای زبان فارسی، سیستم‌های تشخیص گفتار شرکت عصرگویش پرداز است. این شرکت که توسط گروهی از اساتید و دانشجویان دانشگاه صنعتی شریف ایجاد شده است، تنها فعالیت تخصصی خود را در زمینه پردازش سیگنال گفتار و بویژه تشخیص خودکار آن قرار داده است و نرم‌افزارهایی را برای انجام کار تشخیص خودکار گفتار توسعه داده است. نرم‌افزار نویسا که برای دیکته خودکار می‌باشد و نرم‌افزار نیوشا که جهت تشخیص گفتار از پشت خط تلفن توسعه داده شده است، از دستاوردهای آنهاست. از نظر فنی، معیارهایی چون وابسته یا متعلق بوده به گوینده بودن سیستم، اندازه و تعداد واژگان، پیوسته یا گسسته بودن گفتار ورودی، استفاده از محدودیت‌های زبانی و کارایی در محیط‌های واقعی توانمندی سیستم‌های تشخیص گفتار را مشخص می‌کند. در کاربردهایی مانند تشخیص گفتار تلفنی، اطن سیستم لزوماً باید مستقل از گوینده باشند اما سیستم دیکته می‌تواند وابسته به گوینده خاص باشد و برای آن گوینده سیستم بهترین کارایی را داشته باشد. در کاربردهای واقعی معمولاً سیستم را مستقل از گوینده می‌سازند و موقع استفاده به صدای گوینده خاصی آنرا اصطلاحاً تطبیق می‌کنند. این کار در قابلیت گفتاری

مجموعه آفیس به کمک خواندن متون اولیه در ویزارد سیستم انجام می‌شود، چنین قابلیت‌هایی در سیستم نویسا نیز وجود دارد. هر چه تعداد واژگانی که سیستم می‌تواند تشخیص دهد بیشتر باشد، شباهت میان کلمات بیشتر شده و در نتیجه کارایی سیستم به علت افزایش اشتباهات کاهش پیدا می‌کند. از اینرو در کاربردهای واقعی معمولاً فقط واژگان متناسب با کاربرد مورد نظر را انتخاب می‌کنند تا محدود شود. در قابلیت گفتاری آفیس هم که ادعا می‌شود اکثر کلمات انگلیسی را دارد، کارایی به شدت پایین است (کارایی پایین آن به علت این است که ما غیرانگلیسی زبان هستیم!) ولی در محصولات محدودتر این شرکت کارایی به مراتب بهتر است. گفتار کاربر می‌تواند پیوسته و طبیعی و یا با مکث میان کلمات همراه باشد، بدیهی است که حالت اول مطلوب هر کاربری است. یکی از مشکلاتی که محصولات نویسا و نیوشا تا حدی زیادی آن را حل کرده‌اند، استخراج و به‌کارگیری قابلیت‌های زبانی، زبان فارسی در حد نسبتاً کاملی است. این اطلاعات زبانی می‌تواند در سایر نرم‌افزارهایی که نیاز به اطلاعات زبانی دارند، مانند مترجم‌ها و نرم‌افزارهای OCR نیز بکار گرفته شود. اثر صداهای اضافی و ناخواسته در کاربردهای واقعی نرم‌افزارهای تشخیص گفتار را در عمل دچار افت شدید کارایی می‌نماید، در محصولات فارسی ارائه شده با رویکردهای مختلفی این نقصان تا حد زیادی جبران شده است. برخی از محصولات مشتق شده از نرم‌افزارهای تشخیص گفتار فارسی نویسا و نیوشا که در حال حاضر توسعه داده شده‌اند، بصورت زیر است:

- دیکته خودکار متون فارسی با قابلیت تطبیق به صدای گوینده و محیط کار، تطبیق واژگان مورد استفاده متناسب با کاربرد و قابلیت استفاده در هر محیط ویرایشی در ویندوز
- منشی خودکار تلفنی: به جای استفاده از سیستم تُن، فرد تماس‌گیرنده می‌تواند فقط اسم شخص مورد نظر، بخش مورد نظر و یا با بیان گفتاری شماره‌ی داخلی، به مقصد وصل شود.
- سیستم‌های اطلاع‌رسانی گویا برای ندای گویای قرآن، ندای بهداشت و اطلاعات‌رسانی در مورد سازمان‌های کشور.
- اتوماسیون خانگی و صنعتی جهت کنترل گفتاری وسایل منزل چه در خارج از خانه و چه به کمک خط تلفن و کنترل ربات‌ها و سایر وسایل صنعتی به کمک گفتار.
- فرامین صوتی در کامپیوتر برای اجرای صوتی نرم‌افزار و برنامه‌های مورد نظر و استفاده از این فرامین در نرم‌افزارهای خاص مثل M.S. Office یا Media Player برای انجام کارهای لازم.
- تایپ هوشمند با پیش‌بینی کلمه/کلمات مورد نظر کاربر در هنگام تایپ.
- مترجم صوتی برای ترجمه گفتار فارسی به زبان انگلیسی برای عبارات مکالمه‌ای رایج با قابلیت استفاده روی PDAها.

نوشته شده در سال ۱۳۸۴