پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

زبان‌شناسی

● ادراک گفتار / دکتر محمدرضا باطنی

ادراک گفتار*

محمدرضا باطنی

در آواشناسی، هنگام بحث از صداهای زبان، گفته می‌شود که صداهای زبان به وسیلهٔ اندام‌های گویایی (مانند تارهای صوتی، لب‌ها و غیره) تولید می‌شوند. ولی باید به یاد داشته باشیم که واقعیّت امر چنین نیست. توضیح اینکه اندام‌های گویایی ما صدا تولید نمی‌کنند، بلکه ذرات هوا را طبق الگوهای خاصی به هم می‌زنند، یا به بیان دقیق‌تر، در آنها ایجاد ارتعاش می‌کنند. این ارتعاش‌ها، محرک صوت هستند، نه خود صوت. ادراک صوت، ویژگی دستگاه شنوایی ما است، بدین معنی که دستگاه شنوایی ما آنچنان ساخته شده و سازمان یافته است که این ارتعاش‌ها را به صورت صوت ادراک می‌کند، وگرنه در جهان خارج صوت یا صدا وجود ندارد. آنچه وجود دارد تموّج انرژی در ذرات هواست. با این همه، ما می‌توانیم همچنان از تولید صدا به وسیلهٔ اندام‌های گویایی صحبت کنیم به شرط اینکه فراموش نکنیم که این کار متّکی بر نوعی تسامح است، تسامحی که کار ما را در بحث از صداهای زبان آسان می‌سازد. یک هشدار دیگر را نیز نمی‌توان نادیده گرفت: اگر ادراک صوت، و از جمله صداهای زبان، ناشی از ویژگی یا ویژگی‌های دستگاه شنوایی ما است، پس هنگام بحث از صداهای زبان، و در سطحی بالاتر، هنگام بحث از زنجیرهٔ گفتار، نمی‌توان از توجه به ساخت و کار این دستگاه به‌کلی غافل بود.

---

* این مقاله قبلاً چاپ شده و اینک با اندک تغییراتی در اینجا تجدید چاپ می‌شود. (نویسنده)
لازم به توضیح است که در مقاله (استنباط در درک زبان، ص ۸۴-۹۴ ، بخارا۴۱) به علت روی دادن یک اشکال فنی غیر قابل پیش‌بینی، در چاپ نهایی تمامی گیومه‌های سمت راست بصورت علامت سئوال درآمد. بابت آن اشتباه از دکتر باطنی و خوانندگان بخارا پوزش می‌طلبیم. (مجله بخارا)

بسیاری از درس‌خوانده‌ها، تحت تأثیر آشنایی خود با خط و نوشته، چنین می‌پندارند که صداهای زبان در هنگام گفتار، همانند حروف در نوشتار، یکی پس از دیگری ادا می‌شوند، بدین معنی که اندام‌های گویایی نخست صدایی را تلفظ می‌کنند و پس از فراغت یافتن از آن به تلفظ صداهای دوم و سوم می‌پردازند و این کار را تا پایان زنجیرهٔ گفتار ادامه می‌دهند، به طوری که هر صدا منفک و مجزا از صدای قبل و بعد آن قابل تمیز و تشخیص است. این تصور به کلی نادرست است: صداهای زبان در زنجیرهٔ گفتار نه چنین تلفظ می‌شوند و نه دستگاه شنوایی ما آنها را به این صورت ادراک می‌کند.

اندام‌های گفتار را به رقاص ماهری تشبیه کرده‌اند که در همان حال که مشغول انجام دادن حرکتی است خود را برای حرکت بعدی آماده می‌کند و از این‌رو حرکات او موزون و هم‌آهنگ جلوه می‌کند. وقتی به تلفظ صداهای زبان توجه می‌کنیم می‌بینیم که این تشبیه چندان بی‌مورد نیست. فرض کنید که ما می‌خواهیم واژهٔ تک‌هجایی «بید» را تلفظ کنیم. اولین صدای این کلمه /b/، دومین صدای آن /i/ و صدای پایانی آن نیز /d/ است. برای تلفظ /b/ باید دو لب نخست بسته و سپس باز شوند و برای تلفظ /i/ نیز زبان باید در وضع خاصی قرار گیرد. آنچه مسلم است ما این دو عمل را مجزا و منفک از یکدیگر انجام نمی‌دهیم، بلکه آنها را تواماً انجام می‌دهیم، بدین معنی که در همان هنگام که لب‌ها را برای تلفظ /b/ می‌بندیم زبان را نیز در وضعی قرار می‌دهیم که تلفظ /i/ ایجاب می‌کند. نتیجهٔ این کار این است که تقریباً همزمان با گشوده شدن لب‌ها و رها شدن هوایی که در پشت آنها فشرده شده تلفظ /i/ نیز آغاز شده است. این بدان معنا است که اساساً مرزی وجود ندارد که بتوان گفت در اینجا تلفظ /b/ پایان می‌یابد و تلفظ /i/ آغاز می‌شود. حاصل این درهمتنیدگی این است که مقداری از اطلاعات آکوستیکی که برای شناسایی /b/ لازم است در درون /i/ و مقداری از اطلاعات آکوستیکی که برای شناسایی /i/ لازم است در درون /b/ قرار می‌گیرد، یا به زبان غیرفنی، رد پای /b/ در /i/ و رد پای /i/ در /b/ نمایان می‌گردد. همین وضع نیز در مورد صدای پایانی کلمه یعنی /d/ پیش می‌آید. در اینجا نیز پیش از آنکه تلفظ /i/ پایان یابد، محفظهٔ دهان برای تلفظ /d/ آماده می‌شود، به‌طوری‌که بخشی از تلفظ /d/ با /i/ همزمان می‌گردد. در اینجا نیز مرز مشخصی وجود ندارد که بتوان گفت تلفظ /i/ در اینجا پایان یافته و تلفظ /d/ آغاز شده است. در واقع کلمهٔ «بید» در تلفظ، از یک زنجیرهٔ آوایی به هم تنیده و تجزیه‌ناپذیر تشکیل شده است. تجزیه و تحلیل آزمایشگاهی گفتار نیز مشاهدات بالا را تأیید می‌کند. طیف‌نگاشت‌های صوتی نشان می‌دهند که ویژگی‌های آکوستیکی صداهای زبان در زنجیرهٔ گفتار در قلمرو یکدیگر وارد می‌شوند و روی هم به شدت تأثیر می‌گذارند.

بد نیست به مثالی از نوع دیگر توجه کنیم. این بار تلفظ دو کلمهٔ تک‌هجایی «کی» /ki/ و «کو» /ku/ را مقایسه می‌کنیم. در سطح واجشناسی می‌توان گفت که این دو کلمه هر دو با صامت /k/ آغاز می‌شوند و با دو مصوت متفاوت پایان می‌یابند. واج‌نویسی آنها نیز به همین صورت خواهد بود. ولی هنگام تلفظ، کیفیّت آکوستیکی /k/ در آغاز «کی» با کیفیّت آکوستیکی /k/ در

• دکتر محمد رضا باطنی (عکس از ستاره سلیمانی)

آغاز «کو» تقریباً همان اندازه متفاوت است که کیفیّت آکوستیکی مصوت‌هایی که به دنبال آنها می‌آیند. برای تلفظ /k/ در هر دو مورد عقب زبان بالا می‌آید و به نرم‌کام می‌چسبد و راه خروج هوا را سد می‌کند و در نتیجه، هوا در پشت این مانع انباشته می‌شود. ولی در مورد «کی» همزمان با متراکم شدن هوا در پشت این مانع، حفرهٔ دهان خود را برای تلفظ /i/ آماده می‌کند، در حالی که در مورد «کو» حفرهٔ دهان برای تلفظ /u/ آماده می‌شود. در نتیجه، هوایی که در پشت زبان متراکم شده در دو محفظهٔ تشدید که شکل متفاوتی دارند تخلیه می‌شود و همین باعث می‌شود که صدای انفجار خفیفی که از رها شدن هوا ایجاد می‌شود هر بار کیفیّت آکوستیکی متفاوتی پیدا کند و دو صدای متفاوت با دو /k/ متفاوت تولید شود. پاسخ این سؤال که پس چرا ما فقط یک صدای /k/ می‌شنویم دیگر به تولید گفتار مربوط نمی‌شود بلکه به ادراک گفتار مربوط می‌شود. در این مورد، پاسخ سؤال این است: چون تفاوت آکوستیکی میان دو صدای /k/ در نظام آوایی (یا فونولوژی) زبان ما نقشی به عهده ندارد دستگاه شنوایی ما نسبت به تفاوت آنها بی‌اعتنا است و در نتیجه، آن دو صدای متفاوت یک صدا ادراک می‌شوند.

شواهد زیادی در دست است که نشان می‌دهند دستگاه‌های گفتار و شنیدار ما با یکدیگر سخت هم‌آهنگ شده‌اند. از مطالعات آزمایشگاهی که روی ادراک گفتار صورت گرفته چنین برمی‌آید که دستگاه شنوایی ما نسبت به رد پایی که صداهای زبان روی یکدیگر می‌گذارند بسیار حساس است، تا جایی که به کمک همین تأثیرات آکوستیکی جانبی است که می‌تواند صداهای زبان را در زنجیرهٔ گفتار از هم باز شناسد. بنابراین، درهم‌تنیدگی صداهای گفتار نه تنها

مانعی در راه شناسایی و ادراک صداها نیست، بلکه خود کمکی در این راه است. مثلاً از شواهد آزمایشگاهی چنین برمی‌آید که صامت‌های انسدادی و واکبر/g,d,b/ تنها از روی تأثیری که روی مصوت بعد از خود می‌گذارند شناخته و ادراک می‌شوند. اگر سه هجای gâ, dâ, bâ را در نظر بگیریم مشاهده می‌شود که وجود /g, d, b/ در آغاز هجا، روی فورمان دوم مصوت /â/ اثر می‌گذارد، به این نحو که آغاز فورمان دوم پس از /b/ رو به بالا خیز برمی‌دارد، پس از /d/ تغییر محسوسی نمی‌کند یا خیزشی اندک نشان می‌دهد، ولی پس از /g/ به شدت فرو می‌افتد. با دستکاری‌های آزمایشگاهی می‌توان این سه آغاز متفاوت را برید و جداگانه به آنها گوش داد. در این صورت دیگر صدای /g, d, b/ شنیده نمی‌شود، بلکه صداهای غیرزبانی شنیده می‌شود که زیر و بمی آنها نسبت به هم تغییر می‌کند. ولی پس از اینکه این قطعاتِ بریده شده در جای خود قرار گیرند دوباره هجاهای gâ, dâ, bâ شنیده می‌شوند. شواهدی از این دست، دانشمندان گفتارشناس را به سوی این نظریه کشانده است که کوچکترین واحد ادراک گفتار هجا است و نه صداهای منفرد.

دلایل و شواهد دیگری نیز در جهت تأیید این نظریه ارائه شده‌اند. مثلاً محاسبه شده که در گفتار عادی، ما در ثانیه ۲۵ تا ۳۰ صدا را دریافت و ادراک می‌کنیم. حال اگر این تعداد صدا بدون هیچگونه همپوشانی زنجیروار به دنبال هم قرار گیرند و به گوش ما رسانده شوند، دیگر گوش ما گفتار نخواهد شنید، بلکه یک صدای واحد «ویز» مانند خواهد شنید، زیرا دستگاه شنوایی ما توانایی این را ندارد که مرز این تعداد صدا در ثانیه را تشخیص دهد و در نتیجه ما آنها را یکپارچه و به صورت صدای واحدی می‌شنویم، همانگونه که دستگاه بینایی ما هم توانایی این را ندارد که بین این تعداد تصویر متفاوت در ثانیه تمایز قایل شود و در نتیجه آنها را به صورت یک تصویر پیوسته ادراک می‌کند. اگر ما می‌توانیم ۲۵ تا ۳۰ صدای گفتار را در ثانیه ادراک کنیم و بین آنها تمایز بگذاریم به این دلیل است که صداهای گفتار همپوشانی دارند و به صورت بسته‌هایی که هر یک یک هجا هستند به گوش عرضه می‌شوند. بنابراین، معیار اندازه‌گیری قدرت پردازش دستگاه شنوایی ما تعداد صداهایی نیست که در مدتی معین، مثلاً یک ثانیه، دریافت می‌شوند بلکه تعداد هجاهایی است که در این مدت می‌توانند ادراک شوند. علاوه بر شواهد آزمایشگاهی، شواهد رفتاری نیز می‌توان در تأیید این نظریه ارائه کرد. دیده می‌شود که کودکان شش ساله در کلاس اول دبستان به سختی می‌توانند بفهمند که یک کلمه از چند صدا ترکیب شده است، در حالیکه با اندکی راهنمایی می‌توانند کلمه را به هجاهای سازندهٔ آن بخش کنند. همچنین افراد بی‌سواد نمی‌توانند بگویند یک کلمه از چند صدا تشکیل شده است، ولی با اندکی آموزش می‌توانند از عهدهٔ تقطیع هجایی کلمات برآیند، بدون اینکه بتوانند عمل خود را توجیه کنند. این شواهد همه بر این دلالت دارند که دستگاه شنوایی ما طوری ساخته شده که می‌تواند هجاهای زبان را به آسانی ادراک کند، در حالی که ادراک صداهای مجزای زبان برای آن دشوار و گاهی غیرممکن است.

پژوهش‌های اخیر دربارهٔ ادراک گفتار ما را از این نیز فراتر می‌برند. گفته می‌شود که دستگاه شنوایی انسان طوری تکامل یافته که نه‌تنها صداهای گفتار را از صداهای دیگر تشخیص می‌دهد، بلکه آنها را در قسمت‌های مختلف مغز پردازش می‌کند. پیش از آنکه به ذکر آزمایش جالبی که در جهت تأیید این نظریه صورت گرفته بپردازیم باید دو نکته را یادآور شویم. یکی اینکه امروز بین اهل فن قبول عام یافته که مراکز تولید و ادراک گفتار در اکثریت قریب به اتفاق مردم در نیمکرهٔ چپ مغز قرار گرفته است. دیگر اینکه رشته‌های عصب شنوایی از هر گوش به هر دو نیمکرهٔ مغز می‌روند، ولی تعداد رشته‌های عصبی که از گوش راست به نیمکرهٔ چپ می‌روند بسیار بیشتر از رشته‌های عصبی است که از گوش راست به نیمکرهٔ راست می‌روند؛ همچنین تعداد رشته‌های عصبی که از گوش چپ به نیمکرهٔ راست می‌روند بسیار بیشتر از رشته‌های عصبی است که از گوش چپ به نیمکرهٔ چپ می‌روند. در آزمایش مزبور، زوج‌هایی از هجاهای بی‌معنا و متفاوت، مانند ta و ga، را انتخاب کردند و هر یک از دو هجای یک جفت را به‌طور همزمان به کمک دستگاه مخصوص در یکی از دو گوش وارد کردند. نتیجهٔ آزمایش نشان داد که هجاهایی که درست تشخیص داده شدند بیشتر آنهایی بودند که به گوش راست وارد شده بودند. تا اینجا نتیجهٔ آزمایش همان چیزی بود که از پیش انتظار می‌رفت، زیرا تعداد رشته‌های عصبی که از گوش راست به نیمکرهٔ چپ، یعنی مرکز تکلّم، می‌روند بسیار بیشتر است. ولی وقتی به جای هجاهای گفتار، ملودی‌های ساده یا نت‌های موسیقایی انتخاب شد و به همان نحو همزمان در دو گوش نواخته شد، نتیجه برعکس بود. این بار نتهایی که درست تشخیص داده شدند بیشتر آنهایی بودند که در گوش چپ نواخته شده بودند. این نتیجه، نشان داد که صداهای موسیقایی، برخلاف صداهای گفتار، کلاً یا عمدتاً در نیمکرهٔ راست پردازش می‌شوند. از کل این آزمایش می‌توان نتیجه گرفت که دستگاه شنوایی انسان بین صداهای گفتار و دیگر صداها تمایز قایل می‌شود و آنها را در مراکز متفاوتی در مغز پردازش می‌کند.

شواهد دیگری نیز عرضه شده‌اند که نشان می‌دهند صداهای گفتار صداهای خاصی هستند و دستگاه شنوایی ما برای ادراک یا پردازش این صداها از استراتژی خاصی استفاده می‌کند. این موقعی به خوبی آشکار می‌شود که سعی شود صداهای غیرگفتاری را جانشین صداهای گفتار کنند. یکی از این موارد، تلاش برای ساختن دستگاهی است که بتواند حروف چاپی را به صدا، ولی نه صداهای گفتار، تبدیل کند تا نابینایان به جای دیدنِ نوشته صدا بشنوند. بیش از هفتاد سال از تاریخ اولین تلاش برای ساختن چنین دستگاهی می‌گذرد، و در این مدت الفباهای صوتی متفاوتی مورد آزمایش قرار گرفته‌اند. ولی در عمل معلوم شده که هیچکدام از اینها به کارایی الفبای صوتی مورس نیستند. الفبای مورس، که در مخابره تلگرافی به کار می‌رود، صوتی است، ولی صداهای آن صداهای گفتار نیستند. به همین دلیل مشاهده می‌شود که دستگاه شنوایی انسان پس از سال‌ها آموزش و تمرین می‌تواند صداهای مورس را ادراک کند، آن هم با سرعتی که به سختی به یک‌دهم سرعت ادراک گفتار می‌رسد، در حالیکه صداهای زبان بدون هیچگونه

آموزش و تمرینی به سهولت ادراک می‌شوند.

مطالعاتی از اینگونه بعضی از دانشمندان گفتارشناس را بر آن داشته است که ادعا کنند دستگاه شنوایی ما پردازشگر خاصی برای صداهای گفتار دارد. این پردازشگر می‌تواند صداهای گفتار را حتی در شرایط آکوستیکی نامساعد تشخیص دهد و ارتباط گفتاری را امکان‌پذیر سازد. مطالعات آزمایشگاهی نشان داده است که اگر انرژی موجود در صداهای مخل به اندازهٔ انرژی موجود در گفتار باشد مکالمه به راحتی صورت می‌گیرد. اگر انرژی موجود در صداهای مخل بر انرژی موجود در گفتار بیشی گیرد باز هم گفتار قابل درک است؛ تنها موقعی ادراک گفتار با اشکال جدی مواجه می‌شود که نسبت انرژی صداهای مخل به انرژی گفتار به مرز چهار به یک برسد. دستکاری‌های آزمایشگاهی در خصوصیات آکوستیکی گفتار ممکن است طبیعی بودنِ گفتار را خدشه‌دار کنند یا به‌کلی از بین ببرند، ولی الزاماً به ادراک‌پذیری آن لطمه‌ای وارد نمی‌کنند. در واقع یکی از کشف‌های مهم این بوده است که طبیعی بودنِ گفتار و ادراک‌پذیری آن از یک مقوله نیستند. خصوصیات آکوستیکی که به گفتار حالت طبیعی می‌دهند متعدد و متنوع‌اند، ولی همهٔ آنها برای ادراکِ گفتار ضروری نیستند. مثلاً طیف‌نگاشت صوتی نشان می‌دهد که مصوت‌ها، سه و گاهی چهار فورمان دارند، ولی فقط فورمان اول و دوم برای ایجاد تمایز و ادراک مصوت کافی است. مصوتی که دارای سه فورمان باشد طبیعی‌تر به گوش می‌رسد، ولی حذف فورمان سوم تأثیری در ادراک مصوت ندارد. بنابراین، تا زمانی که مؤلفه‌های آکوستیکی اساسی در گفتار محفوظ بمانند، گفتار قابل درک خواهد بود، و این در حالی است که به علت حذف مؤلفه‌های جانبی، طنین آن ممکن است کاملاً غیرطبیعی شده باشد. همچنین مشاهده شده که قطع و وصل‌های سریع در زنجیرهٔ گفتار و نیز پژواک، گرچه آزارنده هستند، ولی در ادراک گفتار تأثیر چندانی ندارند. دوبرابرکردن یا نصف‌کردن سرعت گفتار نیز در ادراک آن بی‌تأثیر است. شدت یا بلندی گفتار فقط تا جایی ضروری است که گفتار را قابل شنیدن سازد؛ از این حد که بگذرد، افزایش شدت یا بلندی گفتار در ادراک آن بی‌تأثیر خواهد بود.

یکی از زمینه‌هایی که به فهم ما از ادراک گفتار کمک کرده است تلاش برای مجهزکردن کامپیوتر به نرم‌افزاری است که بتواند گفتار انسان را درک کند. در کامپیوترهای معمولی که فعلاً در دسترس هستند و مورد استفاده قرار می‌گیرند، تماس انسان با کامپیوتر از راه صفحه کلید است. این بدان معنا است که درون‌داد از طریق نوشتار یا علائم نوشتاری به کامپیوتر داده می‌شود و برون‌داد یا پاسخِ کامپیوتر نیز از طریق نوشتار یا علائم نوشتاری از کامپیوتر گرفته می‌شود. هدف ایده‌آل این است که صفحه کلید از میان برداشته شود و انسان بتواند ارتباط گفتاری مستقیم با کامپیوتر برقرار کند. این ارتباطِ گفتاری را نباید با تماس تلفنی اشتباه کرد. در تماس تلفنی حداقل دو انسان شرکت دارند و خط تلفن و دستگاه‌های وابسته به آن مجرای انتقال پیام هستند، ولی در ارتباط گفتاری مستقیم با کامپیوتر، هدف این است که کامپیوتر نقش یکی از آن دو انسان را به عهده بگیرد. ارتباط گفتاری مستقیم با کامپیوتر بدین معنا است که کاربر، به جای استفاده

از صفحه کلید، مستقیماً در میکروفنی که به کامپیوتری متصل است صحبت کند و پاسخ خود را نیز بی‌واسطه از بلندگوی کامپیوتر بشنود. البته این مانع از آن نیست که یکی از دو قطب، گفتاری و دیگری نوشتاری باشد. مثلاً کاربر در میکروفن کامپیوتر صحبت کند و از کامپیوتر بخواهد گفته‌های او را تایپ کند، که در این صورت درون‌داد گفتاری و برون‌داد نوشتاری است؛ یا برعکس متنی نوشته به کامپیوتر داده شود و از آن خواسته شود که متن را بخواند، که در این صورت درون‌داد نوشتاری و برون‌داد گفتاری است. به هر حال، ارتباط گفتاری با کامپیوتر پای دو فرایند بسیار پیچیده را به میان می‌کشد: یکی شناخت گفتار (speech recognition) و دیگری ترکیب گفتار (speech synthesis). شناخت گفتار مستلزم این است که کامپیوتر به نرم‌افزاری مجهز باشد که به آن امکان دهد عناصر سازنده گفتار را شناسایی کند و پیام را از آنها بیرون بکشد، در حالیکه ترکیب گفتار مستلزم آن است که نرم‌افزار طوری ساخته شده باشد که کامپیوتر بتواند مؤلفه‌های آکوستیکی لازم را با هم ترکیب کند و پاسخ خود را به صورت گفتار مصنوعی ارائه دهد. در اینجا مجال آن نیست که به جنبه‌های فنّی این فرایندها پرداخته شود، ناچار فقط به نکاتی اشاره خواهد شد که بتوانند پرتو تازه‌ای روی بحث اصلی ما، یعنی ادراک گفتار به‌وسیلهٔ انسان، بیندازد.

از این دو فرایند، ترکیب گفتار ساده‌تر از شناخت گفتار از آب درآمده است، به‌طوری که امروز کامپیوتر می‌تواند به صورت برون‌داد نوعی گفتار مصنوعی عرضه کند که به گفتار طبیعی بسیار نزدیک است و به همین دلیل از لحاظ تجاری مورد قبول قرار گرفته و کاربردهایی پیدا کرده است، ولی در شناخت گفتار، که قرینهٔ ادراک گفتار در انسان است، پیشرفت چندانی حاصل نشده است. یکی از علل مهم این ناکامی این است که زنجیرهٔ گفتار را نمی‌توان به راحتی به صداهای ترکیب‌کنندهٔ آن تقطیع کرد و واج‌های آن را شناسایی نمود، زیرا همانگونه که قبلاً گفته شد، در هنگام تلفظ، صداهای زبان درهم تنیده می‌شوند به‌طوری که هر قطعه کوچکی از زنجیرهٔ گفتار، حاوی اطلاعات آکوستیکی دربارهٔ دو و گاهی سه واج است که درهم فرو رفته و با هم همپوشانی یافته‌اند. علاوه بر این، تغییرات واج‌ها در بافت‌های آوایی متفاوت، هویت آنها را دستخوش نوسان می‌کند، و همین امر، که برای دستگاه شنوایی انسان بسیار بی‌اهمیت و غیرقابل اعتنا است، یکی از موانع بزرگ در راه شناخت گفتار به‌وسیلهٔ کامپیوتر است. نخستین سیستمی که براساس تقطیع واجی زنجیرهٔ گفتار برای کامپیوتر طراحی شده در دههٔ ۱۹۵۰ به‌وجود آمد، ولی ناتوانی این سیستم به زودی آشکار شد و به دست فراموشی سپرده شد. از آن زمان به بعد، این رویکرد نسبت به شناخت گفتار به‌کلی کنار گذارده شده است.

امروز سیستم‌هایی که برای شناسایی گفتار طراحی می‌شوند، از الگو (template) استفاده می‌کنند، بدین معنی که ویژگی‌های اکوستیکی کل کلمه، و نه واج‌های سازنده آن، را به زبان ریاضی در حافظهٔ کامپیوتر نگهداری می‌کنند. بعضی از سیستم‌ها نیز هجا را به عنوان الگو انتخاب کرده‌اند. غرض ما از طرح این مسأله تشریح چگونگی شناخت گفتار نیست، بلکه می‌خواهیم

نتیجه‌ای بگیریم که گفته‌های قبلی در این مقاله را تأیید می‌کند: مهندسان کامپیوتر نیز به این نتیجه رسیده‌اند که واج‌ها یا صداهای منفرد گفتار، واحد ادراک نیستند، بلکه کوچکترین واحد ادراکِ گفتار احتمالاً هجا است.

اکنون که سخن به اینجا رسید بد نیست مطلب دیگری را نیز یادآور شویم. امروز در میان زبانشناسان و روانشناسان قبول عام یافته که گفتار صورت طبیعیِ ارتباط در انسان است، در حالی که نوشتار امری ثانوی و عرضی است. به همین دلیل، تولید و ادراک گفتار برای انسان بسیار ساده است و فراگیری آن نیازی به آموزش ندارد. اما خط و نگارش را باید یاد گرفت و یادگیری آن احتیاج به سال‌ها ممارست و تمرین دارد. اما وقتی به نحوهٔ کار کامپیوتر نگاه می‌کنیم وضع را کاملاً برعکس می‌بینیم: شناخت نوشتار و کار کردن با نوشتار برای کامپیوتر بسیار آسان‌تر از گفتار است. شاید ذکر مثالی این موضوع را روشن‌تر کند. نگهداری صورت نوشتاری کلمه‌ای که از پنج حرف ترکیب شده است، مثلاً کلمهٔ «میزان»، پنج بایت یا چهل بیت حافظهٔ کامپیوتر را اشغال می‌کند، در حالیکه نگهداری الگوی آوایی این کلمه با کیفیّت متوسط به بیش از ۲۰/۰۰۰ بیت حافظه نیاز دارد، و اگر قرار باشد با کیفیّت آکوستیکی عالی نگهداری شود به بیش از ۴۰/۰۰۰ بیت حافظه نیاز خواهد داشت. تفاوت بین ۴۰ و ۴۰/۰۰۰ بیت رقمی نیست که به سادگی قابل اغماض باشد. البته برای صرفه‌جویی در حافظه باید این انبوه اطلاعات را به نحوی فشرده کنند. این کار با روشی که ال‌پی‌سی (LPC= linear predictive coding) نامیده می‌شود انجام می‌گیرد.

خلاصه بحث

همانگونه که اندام‌های گویایی و، هماهنگ با آنها، دستگاه عصبی ما برای تولید صداهای زبان تکامل یافته‌اند، همانگونه نیز دستگاه شنوایی ما برای دریافت و تشخیص صداهای زبان تکامل یافته و مجهز شده است. شواهد موجود بعضی از دانشمندان را به این نظریه سوق داده است که دستگاه شنوایی ما صداهای زبان را به نحوی متفاوت از صداهای دیگر پردازش می‌کند؛ یا به بیان دقیق‌تر، بخشی از مغز انسان بالغ برای پردازش صداهای زبان تخصصی شده است. هنگام تلفظ، یا تولید گفتار، صداهای زبان منفک و مجزا از یکدیگر تلفظ نمی‌شوند، بلکه در زنجیرهٔ گفتار روی خواص آکوستیکی یکدیگر اثر می‌گذارند. شواهدی ارائه شد که بر این دلالت داشت که دستگاه شنوایی ما طوری سازمان یافته است که به این ردّ صوتی که صداهای زبان بر یکدیگر می‌گذارند بسیار حساس است و همین حساسیت باعث سهولت و سرعت در پردازش زنجیرهٔ گفتار می‌گردد، در حالی که همین ویژگی، یعنی درهم تنیده شدن صداها در هنگام تولید، یکی از موانع بزرگ در راه «شناخت گفتار» به‌وسیلهٔ کامپیوتراست. همچنین شواهدی ارائه شد که بر این دلالت داشت که هجا کوچکترین واحدِ ادراک گفتار است.