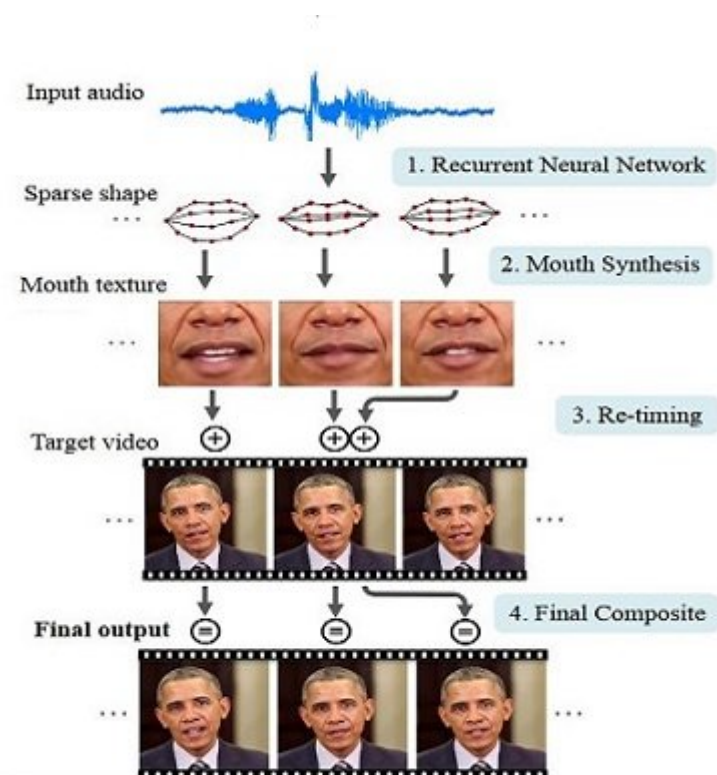




محققان دانشگاه واشینگتن موفق شدند با استفاده از شبکه‌های عصبی مصنوعی، سخنرانی‌های جعلی از اواما بسازند. آن‌ها برای این کار سامانه هوشمند را با میلیون‌ها فریم ویدئو از مصاحبه‌های وی آموزش دادند و سامانه موفق شد، بر اساس صدای گفتار اواما که به‌عنوان ورودی دریافت می‌کرد، حرکات لب‌های او را شبیه‌سازی کند. در واقع، این سامانه آموخته است تا با شنیدن گفتار، حرکات لب‌های مجازی را با آن گفتار هماهنگ کند. محققان در نهایت این لب‌های مجازی را با تصاویر ویدئویی وی ترکیب کردند، به طوری که این طور به نظر برسد او این کلمات را ادا می‌کند.



شکل 1 - سامانه بر اساس صدای ورودی لب‌ها را حرکت می‌دهد، بافت باورپذیری برای لب‌ها در نظر می‌گیرد و نتیجه

در محل مناسب روی ویدئوی هدف قرار داده می‌شود. پیش از ترکیب‌بندی نهایی، لب و ویدئو از نظر زمانی و مکانی هماهنگ می‌شوند و بدین ترتیب حرکات سر و چهره طبیعی‌تر و با صدای ورودی هماهنگ می‌شود.

از نظر این محققان، ویدئوهای اوپاما برای چنین تحقیقی بسیار ایده‌آل بوده است. دلیل اول اینکه ویدئوهای زیادی از وی به‌ویژه مصاحبه‌های هفتگی در یک بازه هشت ساله گرفته شده است و این ویدئوها به طور عمومی در دسترس هستند و منعی برای استفاده از آنها وجود ندارد و در نتیجه برای کارهای تحقیقاتی و دانشگاهی بسیار مناسب است. علاوه بر این کیفیت ویدئوها بالا است. چهره بخش زیادی از قاب تصویر را پر کرده است. هرچند شرایط نورپردازی و ترکیب‌بندی در هر مصاحبه هفتگی کمی متفاوت است و موقعیت سر هم تغییر می‌کند. با این حال، شات‌ها به طور نسبی کنترل شده هستند، به طوری که سوژه در مرکز است و به دوربین نگاه می‌کند. لحن او هم با توجه به اینکه در موقعیت یک مقام عالی کشوری صحبت می‌کند، قاطع و محکم و بدون تغییر است. همه این‌ها شرایط کنترل شده و استانداردی را فراهم می‌کند تا این ویدئوها برای چنین تحقیقی ایده‌آل باشند.

متحرک‌سازی لب‌ها بر اساس صدا یک چالش بسیار جدی است. نتیجه تلاش‌های گذشته در چنین حوزه‌ای، تصاویری بوده است که از باورپذیری کافی برخوردار نبوده‌اند. تولید چنین ویدئوهایی از گفتار علاوه بر ارزش علمی و تحقیقاتی، کاربردهای عملی قابل توجهی نیز دارد. به عنوان مثال، با کمک چنین روشی می‌توان تقاضای پهنای باند برای کدگذاری و ارسال ویدئو را به طور چشمگیر کاهش داد. چیزی که درصد زیادی از پهنای باند فعلی اینترنت به آن اختصاص یافته است. اگر در ویدئو کنفرانس‌ها به دلیل نقص شبکه دریافت تصاویر با اختلال همراه شد، می‌توان با استفاده از صدا بخش‌های از دست رفته ویدئو را به طور مجازی ایجاد کرد. از سوی دیگر، افراد ناشنوا یا کم‌شنوا قادر خواهند بود از سامانه تبدیل صوت به ویدئو برای لب‌خوانی استفاده کنند. در حوزه فیلم‌سازی و انیمیشن هم می‌توان با این روش شخصیت‌های دیجیتال باورپذیرتری ساخت. هدفی که سال‌ها است دنبال می‌شود.

## تاریخ انتشار:

<https://www.shabakeh-mag.com/information-feature/9243/%D8%AF%DB%8C%DA%AF%D8%B1-%D8%A8%D9%87-%D8%B3%D8%AE%D9%86%D8%B1%D8%A7%D9%86%DB%8C%E2%80%8C%D9%87%D8%A7%DB%8C-%D9%88%DB%8C%D8%AF%DB%8C%D9%88%DB%8C%DB%8C-%D8%A7%D8%B9%D8%AA%D9%85%D8%A7%D8%AF-%D9%86%DA%A9%D9%86%DB%8C%D8%AF>