

Jolanta Zielińska

Wyższa Szkoła Pedagogiczna
im. Komisji Edukacji Narodowej
Kraków

Komputerowa animacja aktu mowy jako sposób wspomagania umiejętności czytania z ust

**Computerized Animation of Speech as a Means
of Support in Lipreading Skills**

Streszczenie

Komputerowa animacja aktu mowy, oznaczająca możliwość symulacji twarzy osoby mówiącej z dokładnością pozwalającą na czytanie z ust, jest nowoczesną i efektywną metodą, wykorzystującą wspomagająco-kompensacyjną rolę percepcji wzrokowej. Jest ona istotna, gdyż odczytywanie mowy z ust to często podstawowy sposób recepcji języka mówionego przez osoby z głębszą wadą słuchu. Synteza mowy na podstawie tekstu pisanego pozwala na opracowanie ćwiczeń-wzorców dla ludzi z wadą słuchu. Ćwiczenia te mają na celu wzmocnienie wizualnej percepcji mowy oraz kształtowanie umiejętności czytania z ust i poprawnej motoryki oralnej.

Summary

The authors present the current state of computer software developments in the area of speech animation. Speech animation is achieved by synthetic face construction and lips motion animation. Great attention has been drawn to realistic lip synchronization to make lip reading possible. Visual speech perception creates new means in lip reading training for deaf people. It is also a valuable tool for speech therapists.

I. ROZWAŻANIA OGÓLNE

Akt mowy wydaje się procesem prostym i łatwym, bo w sposób naturalny sprzęgniętym z codzienną egzystencją człowieka. Dopiero powstanie barier utrudniających lub wprost uniemożliwiających jego realizację, przykładowo

z powodu zaistnienia wady słuchu, powoduje zachwianie tej opinii. Analiza sygnału mowy, dokonana na gruncie osiągnięć nowoczesnej techniki, potwierdza fakt doskonałości i złożoności mowy. Odpowiednie wykorzystanie oraz interpretacja wyników tej analizy mogą przynieść znaczące zmiany w podejściu do rewalidacji dzieci z wadą słuchu.

Techniczne metody wizualizacji sygnału mowy możliwe do zastosowania w rewalidacji dzieci z wadą słuchu można podzielić na tradycyjne [Tadeusiewicz 1988], oparte na reprezentacji sygnału mowy w formie czasowej, częstotliwościowej, czasowo-częstotliwościowej i parametrycznej oraz metody specjalistyczne, oparte na pewnych nowych koncepcjach. Do tej ostatniej grupy metod należy komputerowa animacja aktu mowy i techniki multimedialne. Komputerowa animacja aktu mowy jest nowoczesną i efektywną metodą, wykorzystującą wspomagająco-kompensacyjną rolę percepcji wzrokowej i kształcącą umiejętność czytania z ust oraz prawidłową motorykę oralną dzieci z wadą słuchu [Waters, Levergood 1994].

Rozwiązania metodyczne dotyczące kształcenia językowego głuchych wykorzystują obecnie wszystkie możliwe sposoby komunikacji w ramach szeroko rozumianej komunikacji totalnej. Na komunikację totalną składają się: mowa ustna, system znaków migowych, czytanie i pisanie, alfabet palcowy oraz wszelkie inne środki służące skutecznemu przekazywaniu informacji w procesie komunikowania się, w czym istotną rolę odgrywa odczytywanie mowy z ust [Périer 1992]. Obserwacja mówienia innych osób jest bowiem naturalnym źródłem wzorców tej czynności dla dzieci.

Podstawowe sposoby nauczania prawidłowej wymowy opierają się na:

- wrażeniach wzrokowych — poprzez demonstrację prawidłowego układu narządów artykulacyjnych w przypadku tzw. głosek optycznych;
- wrażeniach słuchowych — poprzez uświadamianie prawidłowego brzmienia głosek;
- wrażeniach czuciowych — poprzez wyczuwanie siły wydychanego powietrza na dłoni ustawionej przy ustach oraz drgań strun głosowych podczas artykulacji głosek dźwięcznych czy drgań skrzydełek nosa przy wymawianiu głosek nosowych.

W kształceniu językowym dzieci z wadą słuchu kompensacja wzrokowo-dotykowo-ruchowa zaistniałej wady słuchu umożliwia komunikację z dzieckiem i przekazanie mu, przesuniętych w czasie w stosunku do sygnału mowy, praktycznie pełnych o nim wiadomości. Podstawowy problem kompensacji wzrokowo-dotykowo-ruchowej polega jednak na niezgodności zaistnienia tych procesów w czasie z aktem mowy. Występuje ona bowiem po nim, a nie równocześnie, co znacząco zmniejsza siłę i możliwości jej oddziaływania.

Odczytywanie mowy z ust, zwłaszcza przy głębszej wadzie słuchu, bywa jednym z podstawowych sposobów recepcji języka mówionego. Warunkiem

spostrzegania mowy jest konieczność ciągłego patrzenia na usta osoby mówiącej. Odczytywanie mowy z ust jest bardzo trudne, wyklucza możliwość wiązania odbieranych informacji z równoczesnym działaniem i wymaga [Rakowska 1992]:

- intensywnej aktywności wzroku,
- koncentracji uwagi,
- wysiłku myślowego związanego ze zrozumieniem znaczeń odbieranych wzrokiem,
- stałej pozycji ciała.

Istnieje ścisła korelacja pomiędzy sprawnością odczytywania mowy z ust a postępami w nauce szkolnej dziecka z wadą słuchu. Natomiast na zdolność do odczytywania mowy z ust nie ma wpływu ani ogólna inteligencja, ani tempo wypowiedzi.

Funkcje wzrokowe odgrywają istotną rolę w kształtowaniu werbalnego i niewerbalnego zachowania osób z wadą słuchu. Jeden aspekt problemu to ewentualny rozwój sprawności kompensacyjnych w zakresie nie uszkodzonych zmysłów w sytuacji uszkodzenia jednego (lub więcej) układów zmysłowych. Drugi — to zastępowanie układu słuchowego przez układ wzrokowy w celu odbioru i zrozumienia mowy innych osób [Löwe 1990]. Należy jednak zaprzeczyć twierdzeniu, że istnieje naturalna, fizjologiczna kompensacja ze strony zdrowego układu zmysłowego przy uszkodzonym innym układzie (np. kompensacja zmysłem wzroku uszkodzonego zmysłu słuchu), powodująca dodatkowy, pozytywny rozwój „zastępczego” zmysłu. Fakt, iż u osób z wadą słuchu większą rolę aniżeli u słyszących pełnią zmysły wzroku, dotyku i wibracji, wynika z konieczności życiowej oraz intensywnego treningu [Grzegorzewska 1964].

Najbardziej efektywna wydaje się stymulacja bisensoryczna rozpoznawania mowy słownej poprzez udział w tym procesie jednocześnie dwóch zmysłów, np. słuchu i wzroku. Słuch, chociaż nie funkcjonuje u osób z wadą słuchu prawidłowo, w połączeniu ze wzrokiem poprawia jakość percepcji mowy. Jeżeli bowiem percepcja mowy ustnej odbywa się za pomocą dwóch lub więcej zmysłów, to zarówno u osób z uszkodzonym słuchem, jak i u osób poprawnie słyszących jest ona pełniejsza i pozwala odebrać więcej informacji aniżeli przy udziale jednego tylko zmysłu.

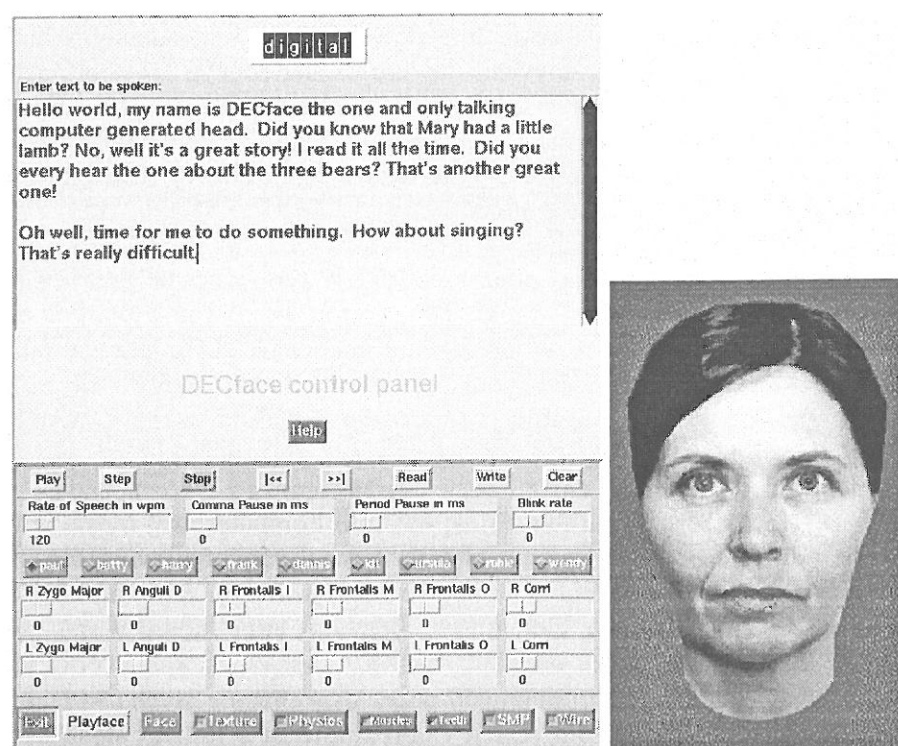
Rzeczywistość nowoczesnej, multimedialnej techniki komputerowej niesie w tym zakresie nowe możliwości, mogące znacząco zwiększyć efektywność działań rewalidacyjnych. Wykorzystanie komputerowej animacji aktu mowy, jako narzędzia wspomagającego umiejętność odczytywania mowy z ust przez osoby z wadą słuchu, może odegrać znaczącą rolę w ich kształceniu językowym, w tym w sferze utrudnionych przez uszkodzenie słuchu kontaktów społecznych, postrzegania informacji i porozumienia z ludźmi słyszącymi.

II. KOMPUTEROWA ANIMACJA AKTU MOWY

W laboratorium firmy DEC w Cambridge USA przeprowadzono w 1993 r. badania dotyczące syntezy aktu mowy na podstawie tekstu. Obejmowały one automatyczną generację sygnału mowy oraz twarzy mówiącego z dokładnością pozwalającą na czytanie z ust. Zasadniczą część tych badań dotyczyła syntezy ruchu warg sztucznej twarzy. Wynikiem tych prac było opracowanie programu DECface, pozwalającego na bezpośrednią animację wymawiania zadanego tekstu [Waters, Levergood 1994].

Postać ekranu sterującego tego programu przedstawiono na rys. 1. Program ten wykonuje następujące sekwencje operacji:

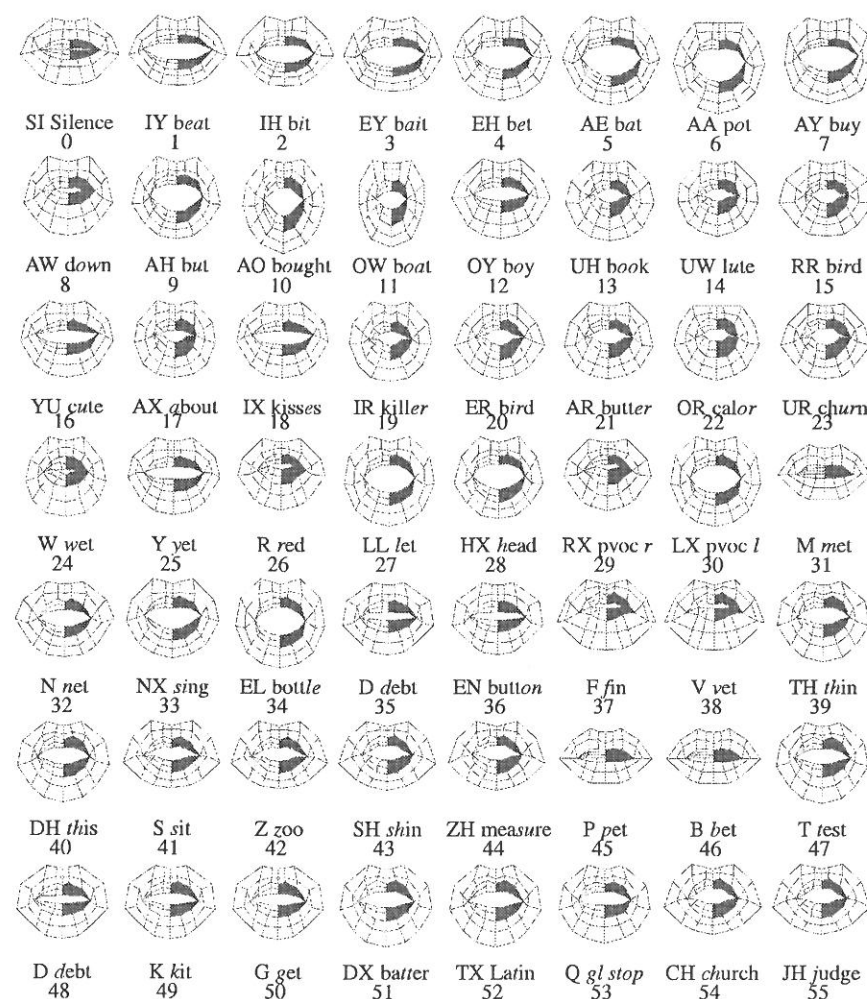
- 1) wprowadzenie tekstu w formie ASCII,
- 2) utworzenie transkrypcji fonetycznej tekstu,
- 3) generację próbek syntezywanego sygnału mowy,
- 4) wyszukiwanie odpowiednich fonemów w bazie danych,
- 5) obliczenie kształtu warg i przebiegu trajektorii ich zmian,
- 6) odtworzenie próbek syntezywanego sygnału mowy w urządzeniu akustycznym (głośniku) oraz ich synchronizacja z ruchami warg sztucznej twarzy na ekranie komputera.



Rys. 1. Wygląd ekranu sterującego programem DECface [Waters, Levergood 1994]

Kroki: 1, 2 i 3 stanowią integralną część rozważanego algorytmu i mogą być widziane jako fazy przetwarzania wstępnego. Zarówno transkrypcja fonetyczna, jak i próbki mowy mogą być generowane wcześniej i przechowywane na dysku. Pozostałe kroki algorytmu są związane z adaptacyjną synchronizacją sygnału akustycznego oraz obrazu i powtarzane aż do momentu, gdy wszystkie próbki mowy zostaną odtworzone.

Czytanie z ust w języku angielskim opiera się na 55 fonemach oraz związanych z nimi ułożeniach warg pokazanych na rys. 2. Tradycyjnie czytanie z ust jest traktowane jako proces całkowicie wizualny, rozwinięty przez osoby z tak dużą wadą słuchu, że słuch nie odgrywa u nich żadnej roli w odbiorze mowy. W wizualnej percepcji mowy można wyróżnić trzy

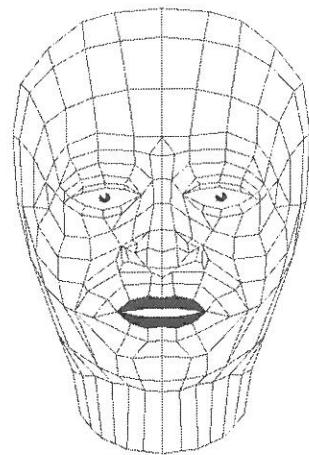


Rys. 2. Ułożenie warg związane z poszczególnymi fonemami języka angielskiego [Waters, Levergood 1994]

mechanizmy: audytoryjny, wizualny i audiowizualny. Osoby z wadą słuchu koncentrują się na audiowizualnej percepcji mowy, kładąc nacisk na obserwację kontekstu, w którym słowa są wypowiedzane, oraz na postawę i mimikę twarzy mówiącego.

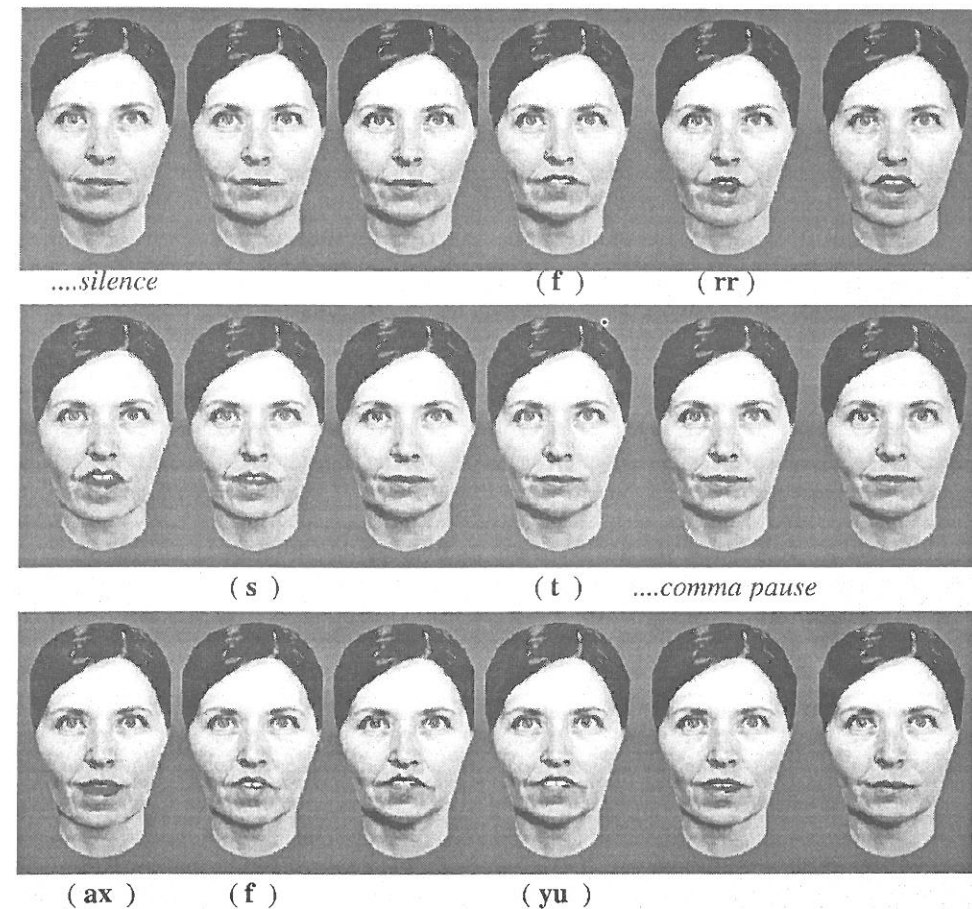
Automatyczna synteza mowy z tekstu opisuje proces generacji sygnału akustycznego odpowiadającego zadanemu ciągowi liter. Systemy te wykorzystują słowniki wymowy określonych wyrazów, uzupełnione zbiorem zasad konwersji tekstu na fonemy, uwzględniające również znaki leksykalne wpływające na rozłożenie akcentu. Syntetyzatory otrzymują zazwyczaj na wejściu ciągi fonemów, które łączone są w wypowiedzi według odpowiednich reguł. Istnieją trzy zasadnicze grupy tych reguł, oparte odpowiednio na: formantach, zasadach artykulacji i konkatencji. Zadaniem syntetyzatora mowy jest generacja sekwencji złożonej z par [fonem, czas]. Parametr czasu określa czas trwania w milisekundach danego fonemu. Sekwencja ta jest tłumaczona na podstawowy ciąg ułożeń ust zgodnie z tabelą przedstawioną na rys. 2.

Płynna zmiana ułożenia ust pomiędzy kolejnymi elementami tej sekwencji wyliczana jest na podstawie modelu fizycznego opartego na dynamice Newtona. Model ten umożliwia obliczenie szybkości przemieszczeń poszczególnych elementów sztucznej twarzy pokazanej na rys. 3 i uwzględnia w sposób parametryczny szybkość wypowiedzi oraz masę poszczególnych elementów twarzy. Wygenerowany w ten sposób ciąg obrazów wizualizowany jest w formie filmu na ekranie komputera.



Rys. 3. Sztuczna twarz [Waters, Levergood 1994]

Przykład wypowiedzi uzyskanej w ramach programu DECface pokazano na rys. 4 w formie wybranych klatek filmu. Przedstawiony system został zimplementowany na stacji roboczej DECAAlpha 3000/500 (150 MHz) i pozwalał na generację 15 klatek na sekundę.



Rys. 4. Przykład wypowiedzi sztucznej twarzy [Waters, Levergood 1994]

III. IMPLIKACJE I WNIOSKI

Najnowsze osiągnięcia nowoczesnej techniki komputerowej stwarzają nowe możliwości jej zastosowania zarówno w analizie sygnału mowy, jak i w syntezie oraz animacji aktu mowy, obejmującego cały kontekst wizualnej możliwości percepcji. Oznacza to m.in. możliwość symulacji twarzy osoby mówiącej z dokładnością pozwalającą na czytanie z ust. Wymaga jednak użycia dużych mocy obliczeniowych i dlatego, niestety, jeszcze obecnie nie można uznać metod rewalidacji opartych na tej koncepcji jako możliwych do powszechnego zastosowania w polskim szkolnictwie specjalnym.

Celem artykułu było pokazanie stanu zaawansowania techniki komputerowej w zakresie animacji aktu mowy oraz możliwości, jakie w perspektywie stwarza niewątpliwie to nowe narzędzie rewalidacji osób z wadą słuchu.

Synteza wypowiedzi na podstawie tekstu pozwala bowiem na opracowanie testów do ćwiczeń-wzorców dla ludzi z wadą słuchu, mających na celu wzmocnienie wizualnej percepcji mowy, kształtowanie umiejętności czytania z ust oraz poprawnej motoryki oralnej, głównie pracy warg. Prezentowane narzędzie jest możliwe do zastosowania zarówno przez nauczyciela-rewalidatora, jak i samodzielnie przez rewalidowaną osobę, przykładowo poprzez wykorzystanie kamery zainstalowanej przy komputerze i wyświetleniu na jego ekranie aktualnego kształtu ust osoby ćwiczącej obok widocznego wzorca.

Sposób wykorzystania komputerowej animacji aktu mowy w kształceniu językowym osób z uszkodzonym słuchem stanowi problem badawczy, wymagający m.in. opracowania wersji oprogramowania przystosowanej do języka polskiego.

Bibliografia

- Grzegorzewska M., 1964: *Pedagogika specjalna*, Warszawa: PiPS.
- Löwe A., 1990: *Wybór pism*, Warszawa-Szczecin.
- Périer O., 1992: *Dziecko z uszkodzonym narządem słuchu*. Tłum. T. Gałkowski, Warszawa: WSiP.
- Rakowska A., 1992: *Rozwój systemu gramatycznego u dzieci głuchych*, Kraków: WN WSP.
- Tadeusiewicz R., 1988: *Sygnal mowy*, Warszawa: WKiŁ.
- Waters K., Levergood T. M., 1994: *DECface: An Automatic Lip-Synchronization Algorithm for Synthetic Faces*, DEC Cambridge Research Lab.