

POLSKA AKADEMIA NAUK

BIOCYBERNETYKA I INŻYNIERIA BIOMEDYCZNA 2000

pod redakcją
MACIEJA NAŁĘCZA

TOM 2

BIOPOMIARY

redaktorzy tomu

WŁADYSŁAW TORBICZ,
LESZEK FILIPCZYŃSKI, ROMAN MANIEWSKI,
MACIEJ NAŁĘCZ, EDWARD STOLARSKI

Tom wydany przy współpracy Polskiego Towarzystwa
Techniki Sensorowej



Przedsiębiorstwo Badawczo-Produkcyjne
„OPTEL” Sp. z o.o.
52-429 Wrocław, ul. Morełowskiego 30
tel. (071) 329-68-53
NIP 8981047033 Regon 008375538

AKADEMICKA OFICyna WYDAWNICZA EXIT
WARSZAWA 2001

Rozdział 11

EFEKTYWNE WYKORZYSTANIE PSEUDOSZEPTU DO UZYSKANIA ZROZUMIAŁOŚCI MOWY OSÓB LARYNGEKTOMOWANYCH

Zygmunt Pawłowski*, Andrzej Owczarek†, Dariusz Kosz‡,
Wiesław Bicz‡

11.1. Wprowadzenie

11.1.1. Charakterystyka pseudoszeptu

Układ rezonansowy (nasada) osoby laryngektomowanej jest nieuszkodzony i sprawny czynnościowo (komory zmiennokształtne i stałokształtne). W związku z tym objętość powietrza zawarta w komorach rezonansowych nasady umożliwia powstawanie pseudoszeptu na skutek subtelnych minimalnych zmian położenia, wynikających z ruchów elementów anatomicznych narządu artykulacyjnego (język, podniebienie miękkie, policzki, wargi). Natężenie pseudoszeptu jest jednak zbyt małe dla jego łatwego zrozumienia, musi być więc wzmocnione drogą protezowania, tj. systemem elektroakustycznym.

* Katedra Foniatrii, Akademia Muzyczna im. Fryderyka Chopina, ul. Okólnik 2, 00-368 Warszawa.

† Instytut Podstaw Elektroniki, Politechnika Warszawska.

‡ Przedsiębiorstwo Badawczo-Produkcyjne Optel sp. z o.o., ul. Otwarta 10a, 50-212 Wrocław.

Istotą proponowanej w niniejszej pracy metody rehabilitacji jest wykorzystanie zjawiska pseudoszeptu w taki sposób, aby możliwe było uzyskanie jego zrozumiałości oraz na drodze ćwiczeń wyeliminowanie różnych czynników towarzyszących, które powodują zakłócenia mowy zastępczej.

11.1.2. Mowa gardłowa i przełykowa

Osoby laryngektomowane mogą porozumiewać się na skutek opanowania tzw. mowy gardłowej lub przełykowej. Jest to jednak proces trudny i nie każdy chory jest zdolny temu podołać. Mowa gardłowa charakteryzuje się tym, że pseudogłosnia znajduje się w gardle środkowym tj. pomiędzy tylną ścianą gardła a nasadą języka. Pseudogłosnia jest wprowadzana w drgania przez zbiornik powietrza zawarty w dolnym gardle. Zrozumiałość mowy przełykowej lub gardłowej znacznie ograniczają szумы oddechowe o dużym natężeniu, emitowane z tracheostomy.

Zasada tworzenia mowy przełykowej polega na dźwięcznym odbiciu (ructus) powietrza z przełyku, które pobudza do drgań usta przełyka, czyli pseudogłosnię. W tym momencie powstający dźwięk podstawowy jest modulowany przez komory rezonansowe nasady, a głównie przez komorę jamy ustnej.

Ustno-gardłowy pseudoszept charakteryzuje się niewystarczającą dźwięcznością, składa się przede wszystkim ze spółgłosek i tworzy się na skutek drgań objętości słupa powietrza zawartego w jamie ustnej i jej przedzionku oraz gardle środkowym i dolnym. Pseudoszept pozostaje do dyspozycji chorego pomimo usunięcia krtani.

11.1.3. Protezy głosowe

Trudności nauczenia się mowy przełykowej lub gardłowej polegają, poza różnymi innymi czynnikami, na konieczności opanowania antyperystaltycznego ruchu przełyku, który przesuwa zawarte w nim powietrze w celu pokonania oporu pseudogłosni, np. mięśnia pierścienno-gardłowego.

Wytwarzane chirurgicznie przetoki głosowe w postaci kanału z błony śluzowej i mięśnia pomiędzy tchawicą a przełykiem lub dolnym gardłem mają na celu wprowadzenie powietrza z tchawicy do danego zbiornika powietrza, np. przełyku lub dolnego gardła. Jest to metoda, której zastosowanie zależy od rozległości nacieku nowotworowego i innych czynników. W Polsce jest mało stosowana.

W licznych przypadkach niemożności opanowania w ramach rehabilitacji foniatrycznej mowy przełykowej stosowane są różnorodne urządzenia, czyli protezy głosowe, tzw. sztuczne krtanie. Protezy te można podzielić na

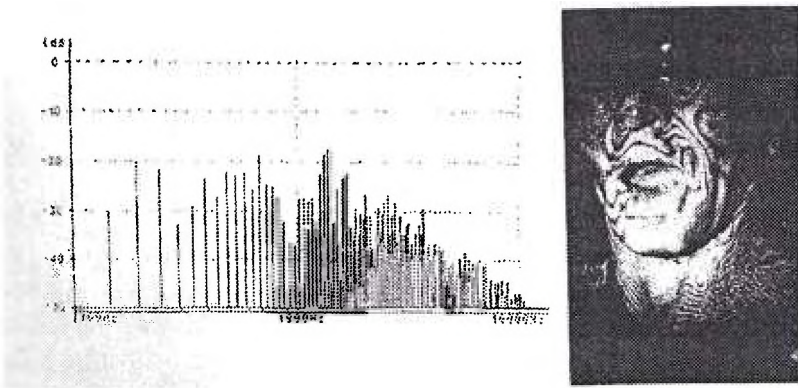
trzy rodzaje: pneumatyczne - wykorzystujące przepływ powietrza z tchawicy, ustne zewnętrzne - w kształcie cylindra, w którym wskutek vibracji membrany powstaje drgający słup powietrza. To urządzenie znane jest jako fajka Ticchioniego. Trzecim rodzajem protezy są - stosowane obecnie najczęściej - wibratory szyjne, których działanie zastępuje pseudogłosnię.

Powstającą w ten sposób mowę zastępczą można w pewnym stopniu porównać z monodią, czyli śpiewem na jednej wysokości, a w czasie jej produkcji słyszalny jest nieprzyjemny szmer vibracji membrany protezy, który znacznie ogranicza zrozumiałość mowy chorego.

Z danych statystycznych wynika, że około 20% laryngektomowanych nie potrafi opanować mowy przelykowej lub gardłowej. Zależy to od wielu czynników, np. rozwoju umysłowego chorego, rozmiaru zabiegu operacyjnego, chorób współistniejących, wieku i innych. W związku z tym, autorzy pracy, dla zapewnienia komunikowania się z otoczeniem osobom laryngektomowanym opracowali modele protez głosowych, dzięki którym wzmocniony (lub też przetworzony) zostaje pseudoszept.

Z naszych wcześniejszych badań [1] dotyczących analizy tworzenia barwy głosu i zrozumiałości mowy laryngektomowanych wynika, że niskie składowe spektralne głosu do 1kHz są spowodowane modulacją (interferencją) wysokich składowych przez układ rezonansowy nasady.

Podczas opracowywania założeń i konsultacji dotyczących modelu naszych protez głosowych, wzięto pod uwagę wyniki cytowanej pracy. W rezultacie opracowano koncepcję i przeprowadzono badania modelowe.



Rys. 11.1. Interferogram drgań skóry nasady i analiza spektralna głoski "a", wypowiedzianej przez mężczyznę po wykonanej laryngektomii (ze zbiorów autorów, Otol. Polska, Supl. Tom I, 168-172, 1986).

11.1.4. Podstawy rehabilitacji

Dotychczas stosowane w foniatryi metody rehabilitacji mowy zastępczej po usunięciu krtani są znane i w tym miejscu nie wymagają szczegółowej charakterystyki. Polegają one głównie na uzyskaniu dźwięcznego odbicia (ructus) powietrza przeważnie z przetyku.

Zastosowania protez głosowych według naszych założeń w początkowej fazie, tj. po operacji nie jest takie proste, jak okularów lub aparatów poprawiających słuch. W tych przypadkach osoby podlegające protezowaniu nie wymagają czynnej współpracy z wyżej wymienionymi protezami. W naszym przypadku - z protezą głosową, dany pacjent zobowiązany jest do aktywnej "współpracy", która polega między innymi na ograniczeniu wykonywania zbędnych ruchów np. podniebienia miękkiego, unikając w ten sposób powstawania efektu głosowego tzw. "mlaskania", nadmiernego przyciskania języka do powierzchni podniebienia twardego itp.

11.2. Cyfrowe przetwarzanie sygnału pseudoszeptu

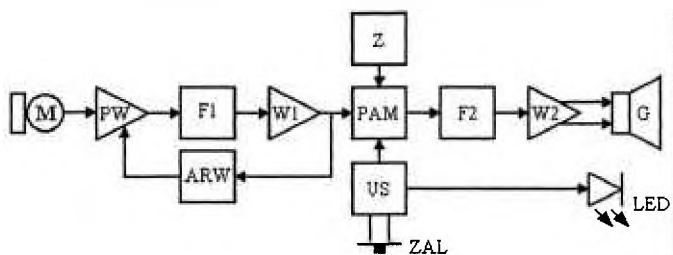
11.2.1. Wprowadzenie

Pseudoszept generowany przez osoby po operacji usunięcia krtani, choć jest różny u poszczególnych osób, ma zawsze jedną wspólną cechę: jest bezdźwięczny i o bardzo małym natężeniu, prawie niesłyszalny przez otoczenie. Stąd też pierwszym naturalnym sposobem poprawy jego zrozumiałości wydaje się być jego wzmocnienie wraz z numerycznym przetworzeniem do postaci mowy dźwięcznej.

11.2.2. Układ do nagrywania i odtwarzania mowy

Niestety, aby uzyskać odpowiednie wzmocnienie akustyczne pseudoszeptu nie wystarczy samo zastosowanie wzmacniacza, ponieważ występuje efekt sprzężenia zwrotnego z mikrofonem. W związku z tym opracowano i wykonano specjalne urządzenie do nagrywania pseudoszeptu a następnie głośnego odtwarzania z opóźnieniem. Na rys. 11.2 pokazano jego schemat.

Posługiwanie się nim jest bardzo łatwe i nie sprawia użytkownikom żadnych kłopotów. Nagrywanie głosu następuje po wciśnięciu małego guzika, przy czym jednocześnie zapala się czerwona dioda informująca otoczenie i samego użytkownika o rozpoczęciu nagrywania. Dioda została umieszczona w oprawie okularów, wraz z podwieszonym na pałąku małym mikrofonem tuż przed ustami pacjenta w odległości około 8cm. Nagranie trwa do momentu zwolnienia przez pacjenta guzika, nie dłużej jednak niż 12 sekund.

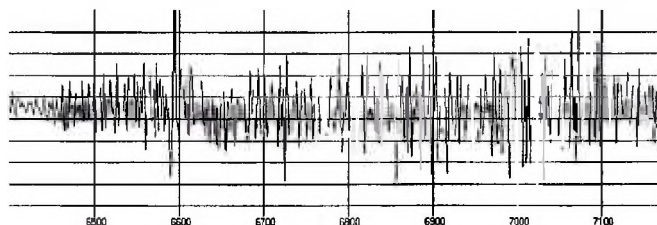


Rys. 11.2. Schemat blokowy układu do nagrywania i odtwarzania mowy: M – mikrofon, PW – przedwzmacniacz o regulowanym wzmacnieniu, F1 – filtr pasmowo przepustowy, W1 – wzmacniacz, ARW – układ automatycznej regulacji wzmacnienia, PAM – pamięć, Z – zegar próbkujący, US – układ sterujący, ZAL – przełącznik nagrywanie/odtwarzanie, F2 – filtr wygładzający, W2 – wzmacniacz, G – głośnik, LED – dioda sygnalizacyjna.

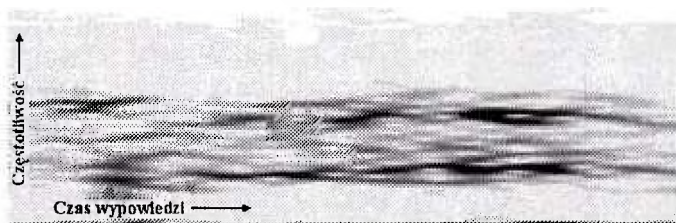
Po tym czasie dioda gaśnie i następuje automatyczne odtworzenie wzmacnionego głosu. Powyższą sekwencję czynności można oczywiście wykonywać dowolną ilość razy. Wykonanie tego układu pozwoliło ocenić funkcjonalność takiego rozwiązania jako bardzo ergonomiczną, idealnie nadającą się do zastosowania przez ludzi pozbawionych krtani.

11.2.3. Algorytm numeryczny udźwięczniania pseudoszeptu

Pomysł opracowania odpowiedniej metody numerycznej, która umożliwi zamianę szeptu w mowę dźwięczną wiąże się z faktem, że samo wzmacnienie dźwięku, choć istotne, nie daje jeszcze zadowalających efektów, gdyż pseudoszept, nawet bardzo głośny, jest nadal trudno zrozumiały. Niewątpliwie jedną z ważniejszych przyczyn takiego stanu rzeczy jest jego nienaturalny, szumowy charakter. Na rys. 11.3 przedstawiono przykładowy przebieg sygnału odpowiadający słowu “czy”, które wypowiedział jeden z pacjentów. Pomimo tego, że występuje tu głównie głoska dźwięczna “y”, obraz tej wypowiedzi jest typowym szumem całkowicie pozbawionym jakichkolwiek oznak pierwotnej periodyczności związanej z istnieniem tonu krtaniowego. Jeżeli chcemy “udźwięcznić” taki szumowy sygnał, to musimy tak go przetworzyć, by wzmacnić jego periodyczność a jednocześnie zachować przy tym te jego cechy, które warunkują lingwistyczny sens wypowiedzi. Problem ten najlepiej obrazuje spektrogram [2], na którym można dokładnie zobaczyć jak rozkłada się energia w poszczególnych przedziałach czasu i częstotliwości:



Rys. 11.3. Przebieg sygnału pseudoszeptu odpowiadający słowu “czy”.



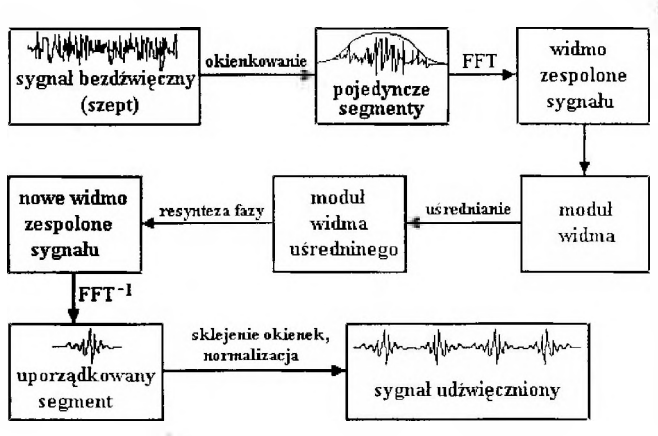
Rys. 11.4. Spektrogram sygnału pseudoszeptu odpowiadający słowu “czy”.

Widać tu wyraźnie, że pomimo pewnej losowości rozkładu energii (wielkość energii na obrazie jest proporcjonalna do stopnia jego zaczerwienienia), występuje także jej skupienie w obrębie dwóch częstości formantowych podobnie jak w przypadku prawidłowo wypowiedzianej głoski “y”. Potwierdza to znany fakt, że choć sygnał pseudoszeptu ma charakter szumowy, to często zawiera w sobie cechy artykulacyjne konkretnych głosek.

W wyniku przeprowadzonych doświadczeń udało się stworzyć matematyczną metodę przetwarzającą szept (mowę bezdźwięczną) do postaci mowy dźwięcznej. Oznacza to, że istnieje w tej chwili bardzo efektywny, czyli szybki i stosunkowo nieskomplikowany algorytm, który potrafi “udźwięcznić” dowolne frazy wypowiedziane szeptem przez zdrowego człowieka. Okazuje się, że metoda ta pozwala również poprawić zrozumiałość pseudo-szeptu generowanego przez ludzi pozbawionych krtani, choć wiadomo, że w ogólnym przypadku nie zawsze się to udaje.

Idea udźwięcznienia opiera się na nieco uproszczonym założeniu, że widmo mocy w krótkim segmencie czasu dla sygnału wypowiedzianego szeptem jest podobne do widma pełnej, dźwięcznej mowy, natomiast różnice w sygnałach wynikają głównie z różnic fazowych. To właśnie odpowiedni przebieg fazy sygnału zapewnia równomierny rozkład energii, odbierany później jako “dźwięczność”. Idąc dalej tym tropem można stwierdzić, że jeżeli uporząd-

kuje się fazę sygnału, oraz nieco uśredni się widmo po czasie, to w efekcie powstanie przebieg, który będzie miał periodyczny charakter przy zachowaniu podstawowych własności widmowych oryginału. Schemat na rys. 11.5 pokazuje sekwencję przetwarzania sygnału w celu jego udźwięcznienia:



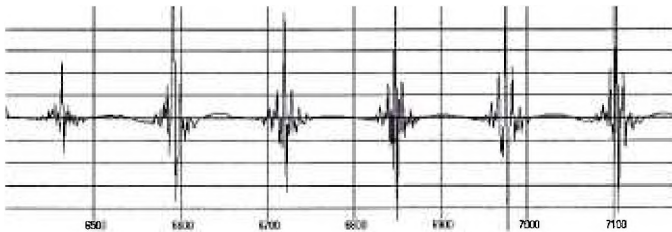
Rys. 11.5. Schemat przetwarzania sygnału w celu jego udźwięcznienia.

Proces uśredniania modułu widma dokonywany jest zazwyczaj w obrębie kolejnych trzech segmentów. Najprostszą metodą re-syntezy nowej fazy jest założenie, że przyjmuje ona wartość stałą. Stanie się tak jeżeli przykładowo: dokonamy podstawienia: $Re(w) = Im(w) = abs[Fu(jw)]$, czyli wypełnimy część urojoną i rzeczywistą transformaty uśrednioną wartością modułu widma.

Nie jest to jedyna możliwość uporządkowania fazy. Gdybyśmy dokładniej wiedzieli jaka głoska jest udźwięczniona, możliwe byłoby przyjęcie pewnego ustalonego (dla tej konkretnie głoski) rozkładu fazy $Ph(w)$, a następnie wyliczenie na nowo składowej rzeczywistej i urojonej transformaty jako:

$$\begin{aligned}
 Re(w) &= abs[Fu(jw)] * \cos(Ph(w)) \\
 Im(w) &= abs[Fu(jw)] * \sin(Ph(w))
 \end{aligned}
 \tag{11.1}$$

Powyższe wzory stanowią definicję resyntezy zespolonego widma sygnału na podstawie uśrednionego widma mocy $Fu(w)$ i przy pewnej założonej fazie $Ph(w)$. Jak działa taki algorytm dla najprostszego przypadku, gdy $Ph(w) = 1$ pokazuje rys. 11.6.



Rys. 11.6. Przebieg sygnału z rys. 11.3 po zastosowaniu procedury udźwięcznienia.

Widać, że zastosowanie algorytmu ze schematu na rys. 11.5 spowodowało koncentrację energii sygnału w krótkich paczkach, co akustycznie objawia się wrażeniem dźwięczności. W tym momencie możliwe jest również sztuczne dodanie niskoczęstotliwościowych składowych dowolnego tonu krtaniowego o częstotliwości podstawowej odpowiadającej częstotliwości okienkowania sygnału. Zabieg taki dodatkowo wzbogaci brzmienie powstałej w ten sposób “pseudomowy”.

11.2.4. Podsumowanie

Na podstawie analizy materiału dźwiękowego - zarówno odsłuchu jak i obserwacji graficznych przebiegów sygnału - wiadomo, że jakość generowanego pseudoszeptu jest nie tylko różna dla poszczególnych osób, ale także dla tej samej osoby zależy od frazy jaka jest wypowiedziana. Niewątpliwie najlepszym materiałem do “przetworzenia przez komputer” jest pseudo-szept może nawet cichy i szumiący, ale wyraźny, ponieważ wtedy właśnie zastosowanie nowej metody daje najlepsze rezultaty. Dotyczy to w szczególności samogłosek, które często występują w postaci zanikającego szumu o barwie zbliżonej do głosek naturalnych, a to oznacza, że w wyniku udźwięcznienia i wzmocnienia mogą stać się bardzo dobrze zrozumiałe.

Jeżeli chodzi o spółgłoski, to ich rekonstrukcja różnie wygląda w zależności od tego, czy pierwotnie były dźwięczne czy bezdźwięczne, nosowe, trące czy wybuchowe. Prawdopodobnie każda z tych kategorii głosek wymaga trochę innego podejścia i pewnym problemem może być ich wstępna klasyfikacja przed ich rekonstrukcją. Przykładowo, głoski takie jak “t” i “d” często są zbyt “wybuchowe” co objawia się przesterowaniem sygnału w miejscu, gdzie aktualnie występują. Zaskóca to również proporcje głośności w stosunku do pozostałych dźwięków. Rozwiązania tego akurat problemu można szukać zarówno w doborze innego mikrofonu jak i odpowiedniej

metody filtracji numerycznej [3], która wykrywałaby ten akurat rodzaj głoski jeszcze przed podjęciem procedury udźwięczniania. Na szczęście, w większości pozostałych przypadków udźwięcznienie głoski, która normalnie jest bezdźwięczna nie pogarsza znacząco jej zrozumiałości. Znacznie większym problemem jest przetwarzanie pseudoszeptu u tych osób, które po usunięciu krtani praktycznie nie są w stanie nic sensownego wyartykułować.

11.3. Aktywne próbkowanie kanału głosowego przy pomocy ultradźwięków

11.3.1. Opis metody

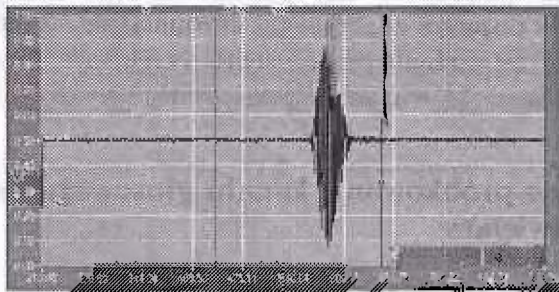
W trakcie rozważania możliwości wykorzystania pseudoszeptu, do którego generacji zdolni są ludzie pozbawieni krtani powstała idea nieco odległa od klasycznych, znanych dotychczas. Można stwierdzić, że polega ona niejako na odwróceniu procesu mówienia: W warunkach fizjologicznych wytwarzania głosu dźwięk krtaniowy jest produkowany przez fałdy głosowe i następnie kształtowany w komorach rezonansowych nasady np. jamy ustnej, komór gardła, komór zatok bocznych nosa itd.

W proponowanej idei dźwięk jest emitowany przez umieszczony w niewielkiej odległości przed otworem wylotowym ust nadajnik i propaguje się w odwrotnym kierunku. Aby nie był on przykry dla użytkownika i otoczenia, wykorzystuje się niesłyszalne ultradźwięki. Wysyłany jest krótki impuls ultradźwiękowy, a odbierana dłużej trwająca odpowiedź - dźwięk odbity od warg, języka gardła i innych elementów kanału głosowego.

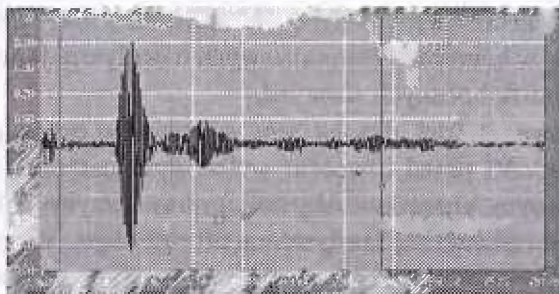
11.3.2. Sposób realizacji

Wstępne eksperymenty z wykorzystaniem fal ultradźwiękowych o częstotliwości rzędu 40 oraz 300kHz wykazały, że już same ruchy warg pozwalają na uzyskanie znacznej zmienności sygnału, co pozwala na wygenerowanie wystarczającej do porozumiewania się różnorodności reakcji. Zapewne jest też możliwe, aby urządzenie tego typu mogło naśladować naturalną mowę człowieka, ale na pewno prostsze do wykonania jest urządzenie funkcjonujące na zasadzie wykonywania wyuczonych rozkazów, a mianowicie: Osoba używająca je uczy urządzenie, że określone ułożenie warg, języka i ust odpowiada określonym głoskom lub wyrazom. W ten sposób przekazywane są do komputera informacje, umożliwiające komunikowanie się z otoczeniem. Możliwe jest też wykonywanie wszelkich innych operacji. Na rys. 11.7 i 11.8 przedstawione są odpowiednio: Impuls ultradźwiękowy o częstości

300kHz odbity od powierzchni płaskiej i od ust człowieka, wymawiającego jakąś głoskę.



Rys. 11.7. Impuls ultradźwiękowy odbity od płaskiej powierzchni (na osi poziomej przedstawiony jest czas, na pionowej amplituda impulsu).



Rys. 11.8. Impuls ultradźwiękowy odbity od ust człowieka, wymawiającego jakąś głoskę. (na osi poziomej przedstawiony jest czas, na pionowej amplituda impulsu).

Sygnal przedstawiony na rysunku 11.8 jest oczywiście zależny od tego, jaką głoskę się wymawia i zmienia się w trakcie jej wymawiania. Niestety występuje też zależność od położenia nadajnika i odbiornika względem ust i musi ona być uwzględniona w obróbce sygnału.

11.3.3. Konstrukcja urządzenia

W jego skład wchodzi dwa elementy:

- a) Zestaw nadawczo-odbiorczy, wysyłający krótkie impulsy ultradźwiękowe i odbierający odbite echa, który może mieć formę taką,

jak mocowane do głowy mikrofony, popularne dzisiaj wśród artystów estradowych oraz:

- b) Skrzynka, zawierająca konieczną do nadawania i odbierania ultradźwięków elektronikę oraz komputer, dokonujący analizy odebranego sygnału, a także jego zamiany na zrozumiałe wyrazy lub na inne rozkazy wykonawcze.

Wygląda na to, że po odpowiednim zminiaturyzowaniu, wymiary zestawu nadawczo-odbiorczego nie przekroczą wielkości pudełka od zapalek. Założenia dotyczące realizacji opisanego urządzenia są realne z uwagi na opracowany i wykonany wstępnie model laboratoryjny. W chwili oddawania artykułu do druku nie był jeszcze - co prawda - gotowy program analizujący sygnał i pozwalający na sterowanie komputerem przy pomocy ruchów ust, wygląda jednak na to, że taki demonstracyjny program niedługo już powstanie.

11.4. Podsumowanie

Postawiony cel efektywnego wykorzystania pseudoszeptu do uzyskania zrozumiałości mowy osób laryngektomowanych wydaje się być do osiągnięcia przy pomocy zaproponowanych metod. Prace nad ich urzeczywistnieniem doprowadzone zostały do etapu sprawdzenia ich realizowalności i pozwalają na wydanie oceny, że dalsze działania mają duży sens i dają szanse stworzenia działających i dających się zastosować w praktyce urządzeń. Trudno jednak dzisiaj już powiedzieć, jak dużo wysiłku trzeba włożyć w ich stworzenie.

Bibliografia

- [1] Pawłowski Z., Kraska Z., Pawluczyk R., Muras A., Maniecka, B.: *Holograficzna wizualizacja drgań nasady u osób laryngektomowanych*. — *Otolaryngologia polska*, 1986, Supl. Tom I, pp.168-172.
- [2] Basztura Cz.: *Źródła, sygnały i obrazy akustyczne*. — WKŁ, 1988, Warszawa.
- [3] Rutkowski L.: *Filtry adaptacyjne i adaptacyjne przetwarzanie sygnałów*. — WNT, 1994, Warszawa.