

**Piotra Łobacz, Waldemar Grygiel, Emilia Baranowska,
Katarzyna Francuzik**

Instytut Językoznawstwa, Uniwersytet im. Adama Mickiewicza, Poznań

Klasyfikacja samogłosek polskich za pomocą sieci neuronowych w wymowie dzieci niesłyszących

**Classification of Polish vowels spoken
by deaf children using artificial neural networks**

Słowa kluczowe: dzieci głuche, widma samogłosek, statystyczne sieci neuronowe.
Key words: deaf children, vowel spectra, statistical neural networks.

Streszczenie

W pracy wykorzystano probabilistyczne sztuczne sieci neuronowe do klasyfikacji samogłosek polskich, wypowiedzianych przez starsze dzieci z głębokim ubytkiem słuchu w wyrazowych wypowiedziach testu nazywania i czytania. Podstawę klasyfikacji stanowiły wartości częstotliwości pierwszych czterech formantów oraz szerokości ich wstęg. Dla bardzo zróżnicowanej pod względem sprawności artykulacyjnej grupy mówców uzyskano przeciętną poprawność klasyfikacji na poziomie 70%. Najlepiej rozpoznaną przez sztuczną sieć neuronową była samogłoska /u/, a najgorzej samogłoska /o/.

Summary

A probabilistic neural network has been used to classify Polish vowels spoken by older children with profound hearing loss, in balanced utterances of naming and reading tests. The classifying parameters are the frequencies and bandwidths of the first four formants. A general classification score of 70% was obtained for a group of speakers which was very diversified with respect to articulatory skill. The vowel which was best identified by the net was /u/. The lowest score was obtained for /o/.

I. UWAGI WSTĘPNE

Mowę osób niesłyszących analizuje się z wielu punktów widzenia. Najczęściej dokonuje się charakterystyki jej jakości na podstawie testów odsłuchowych. Rzadziej korzysta się z metod instrumentalnych. Jedną z nich jest analiza akustyczna. Stanowi ona istotną pomoc dla surdologopedów w celu bardziej szczegółowego opisu występujących w mowie głuchych odstępstw od wymowy normatywnej oraz cech dla tej mowy specyficznych. W 2002 r. przedstawiono dwa opracowania fonetyczno-akustyczne polskich wypowiedzi wyrazowych dzieci niesłyszących ze starszych klas szkoły podstawowej, które przeprowadzono dla grupy 37 osób. W pierwszym z nich [Łobacz 2002] przeprowadzono wstępną ogólną charakterystykę wymowy tych dzieci na podstawie oceny audytywnej, w drugim studium dokonano tylko szczegółowej analizy wymowy polskich samogłosek [Łobacz, Francuzik, Szalkowska 2002]. Porównano realizację samogłosek w wypowiedziach osób niesłyszących z wymową normatywną (dziecięcą i osób dorosłych), koncentrując się głównie na ustaleniu wspólnych całej grupie badanych osób mechanizmów artykulacyjnych, które manifestują się określonymi parametrami akustycznymi.

Przedmiotem obecnej pracy, wykorzystującej te same pomiary parametrów akustycznych, jest szczegółowa analiza indywidualnych umiejętności wymowy samogłosek przez poszczególne dzieci oraz ustalenie skali trudności artykulacyjnej związanej z realizacją określonych samogłosek¹. Zastosowana procedura badawcza w postaci szczegółowych pomiarów częstotliwości czterech pierwszych formantów samogłosek oraz szerokości wstęg wszystkich wyznaczonych formantów w trzech punktach czasowych przebiegu artykulacji każdej samogłoski stanowiła punkt wyjścia do stworzenia najpierw klasyfikacji, a potem rozpoznawania poszczególnych realizacji samogłosek u każdego mówcy. Do obu procedur wykorzystano probabilistyczną sieć neuronową przetwarzającą wartości liczbowe zawarte w utworzonej bazie danych, która obejmowała łącznie ok. 12 tysięcy danych. Stworzono także sieć neuronową przetwarzającą informacje wejściowe dla poszczególnych samogłosek². Uzyskane rezultaty porównano z wynikami analizy słuchowej. Dokonano również próby ustalenia, które czynniki w dotychczasowej edukacji i rehabilitacji poszczególnych dzieci miały istotny wpływ na jakość ich samogłoskowych artykulacji.

¹ Dla tej samej grupy dzieci dokonano także charakterystyki fonetyczno-akustycznej wybranych grup spółgłosek oraz określono poziom zrozumiałości na podstawie tego samego materiału językowego, co w obecnie referowanych badaniach [por. Kleśta 2002].

² Wykorzystywanie metod sztucznej inteligencji znajduje coraz szersze zastosowanie w badaniach wypowiedzi mowy zaburzonej [por. np. Izvorski, Wszolek 1999].

II. CHARAKTERYSTYKA BADANYCH OSÓB³

W badaniach uczestniczyło 37 dzieci (21 chłopców i 16 dziewcząt) w wieku od 10 do prawie 17 lat, uczęszczających do tej samej szkoły dla dzieci niesłyszących (klasy III-VIII). Wybór dzieci został dokonany przez dyrektorkę szkoły na podstawie ogólnej oceny sprawności mówienia i możliwości wykonania testu.

Wszystkie dzieci charakteryzują się głębokim upośledzeniem słuchu, dla znacznej większości uczestników badań stan słuchu nie zmienił się od momentu rozpoczęcia nauki w szkole. Badana grupa dzieci charakteryzuje się niejednorodnością z punktu widzenia przyczyn utraty słuchu. Co więcej, szkoła nie zawsze dysponowała kompletną dokumentacją w tym względzie. Stwierdzono występowanie następujących przyczyn utraty słuchu: głuchota o charakterze dziedzicznym, uszkodzenie wskutek różyczki, stosowania leków oraz zapalenia opon mózgowych, wcześniactwo i wodogłowie. Prawie połowa uczniów nie ma jednoznacznie ustalonej przyczyny głębokiego niedosłuchu. Jeszcze więcej niejasności (dla części badanych osób) wiąże się z określeniem dokładnego czasu utraty słuchu, co praktycznie uniemożliwia statystyczną analizę wpływu tego czynnika na jakość mówienia. Grupa była niejednorodna również pod względem czasu zaopatrzenia w protezy słuchowe. Ustalono, że żadne dziecko nie posiadało protezy w pierwszym roku życia, trzy osoby stosowały protezę od drugiego i trzeciego roku życia, pozostałe dzieci rozpoczęły używanie protez w wieku 4-6 lat, przy czym aż 13 dopiero w momencie pójścia do szkoły, a jedno dziecko powyżej siódmego roku życia. Gdyby przyjąć tzw. wiek zerowy jako m.in. początek względnie stałego noszenia protez słuchowych, to dla wieku metrykalnego zawierającego się w przedziale od 10 do prawie 17 lat obniży się on znacznie i będzie wynosił od 3 do 11 lat, co odpowiada wiekowi przedszkolnemu i wczesnoszkolnemu [por. Kurkowski 1996; Łobacz 2002]. Rehabilitacja dzieci miała bardzo zróżnicowany zakres. Troje uczęszczało przynajmniej rok do przedszkola dla niesłyszących, ośmioro korzystało ze stałej, trwającej od 1 roku do 3 lat, opieki poradni foniatrycznej, 12 dzieci uczęszczało prawie regularnie na terapię logopedyczną.

Badane dzieci różnią się także okresem uczęszczania do obecnej szkoły dla niesłyszących. Ma to wpływ nie tylko na ich zróżnicowanie pod względem ogólnego rozwoju poznawczego, ale także na stopień opanowania pierwszego języka – języka migowego. W ocenie nauczycieli nieco mniej niż połowa badanej grupy opanowała język migowy w stopniu przynajmniej dobrym, kilkanaścioro dzieci nadal się go uczy, przy czym dziewięcioro z nich opanowało go słabo, troje dzieci nie używa języka migowego w ogóle.

³ Szczegółowy opis badanej grupy mówców zamieszczono w: Łobacz, Baranowska, Francuzik [2002].

Grupa dzieci okazała się mniej zróżnicowana ze względu na status rodzinny. Tylko pięcioro dzieci ma niesłyszących rodziców, przeważają także dzieci posiadające słyszące rodzeństwo (28 osób), siedmioro dzieci ma niesłyszące rodzeństwo, cztery osoby mają rodzeństwo różne pod względem słuchu, a tylko troje dzieci to jedynacy.

Heterogeniczność badanej grupy dzieci znalazła odbicie również w ogólnej ocenie zrozumiałości ich mowy. Kleśta [2002] ustalił, że przeciętna zrozumiałość mowy wszystkich 37 dzieci, stwierdzona na podstawie testu wyrazowego, waha się w granicach 5-55%, a przeciętnie wynosi ok. 21%. Test wyrazowy jest stosunkowo ostrym kryterium oceny zrozumiałości mowy (w porównaniu z testem zdaniowym czy oceną słuchową próbek mowy spontanicznej), ale mimo to wynik ten należy uznać za dość niski. Obecne studium ma charakter przekrojowy, a wybór dzieci odzwierciedla autentyczną rzeczywistość szkolną – znaczne zróżnicowanie pod względem umiejętności mówienia występujące w tej samej klasie.

III. METODA BADAWCZA

1. Wybór materiału językowego⁴ został zasugerowany przez pedagogów w szkole dla niesłyszących. Wybrano czterdziestowyrazowy test artykulacyjny Golanowskiej, opracowany pierwotnie do badania wymowy dzieci w wieku od dwóch do trzech lat⁵. Test artykulacyjny jest testem nazywania, zakłada więc określony zasób czynnego słownictwa u badanego dziecka i odpowiedni poziom jego wiedzy o świecie. Do badania dzieci niesłyszących test został specjalnie znormalizowany. Sposób zadawania pytań musiał być dostosowany do wieku dzieci, rzeczywistości szkolnej itp. Z wyjątkiem dwóch onomatopei ([hawhaw] oraz [mjawmjaw]) wszystkie wyrazy testu występują w czynnym słownictwie dzieci przedszkolnych, co ustalono na podstawie list frekwencyjnych Zgółkowej i Bułczyńskiej [1987]. Zastosowany test miał postać serii obrazków. Postać rysunkowa gwarantuje większą naturalność odpowiedzi, ale dopuszcza również jej brak. Poza testem nazywania wszystkie dzieci głuche proszono także o przeczytanie wyrazów, których dotyczyły obrazki. W budowie testu artykulacyjnego, zgodnie z przeznaczeniem do badania wymowy dzieci w wieku od 2 do 3 lat, znacznie więcej uwagi poświęcono doborowi odpowiednich spółgłosek niż wyrównaniu jednorodności otoczenia samogłosek.

2. Nagrania dokonano w szkolnym gabinecie logopedycznym. Dzieci uczestniczyły w nagraniach pojedynczo, zawsze w obecności dwóch eksperymentatorów.

⁴ Szczegółowy opis materiału badawczego oraz procedury pomiarów i nagrań zamieszczono w: Łobacz, Baranowska, Francuzik [2002].

⁵ Test wyrazowy Golanowskiej nie został jeszcze opublikowany.

Procedura eksperymentu była ujednolicona (test był standaryzowany). Eksperymentatorzy posługiwali się wyłącznie mową – badane dzieci czytały z warg. Dzieciom zdarzało się używać języka migowego, ale tylko wtedy, gdy nie znały odpowiedzi w języku mówionym. Nagrania z każdym mówcą przeprowadzano w ciągu jednego spotkania. Zrezygnowano z dodatkowych sesji, nawet jeśli nie udało się uzyskać niektórych wyrazów. Do nagrań wykorzystano mikrofon kierunkowy SHURE typ SM 48 i magnetofon cyfrowy DAT firmy Panasonic. Zarejestrowane wypowiedzi zostały przeniesione do komputera z częstotliwością próbkowania 20 kHz. Badania przeprowadzano w latach 1998-1999.

W trakcie przeprowadzania testu stosowano z wyboru metodę „dobrowolnego wymuszania” spontanicznych wypowiedzi jednowyrazowych, rzadziej posługiwano się metodą imitacji (dziecko głuche czytało z warg). W badaniu testem nazywania żadne dziecko nie udzieliło wszystkich odpowiedzi. Znajomość niektórych wyrazów przez dzieci niesłyszące okazała się zróżnicowana: mimo że teoretycznie powinny mieć w tzw. czynnym słownictwie wszystkie wyrazy z listy, w praktyce nie zawsze tak było.

Z nagranych 40 wyrazów testu do dalszych badań wybrano 18, tak by każda z sześciu ustnych samogłosek polskich (i, y, e, a, o, u) wystąpiła w trzech różnych kontekstach spółgłoskowych. Dobór wyrazów do dalszej analizy opierał się na czterech postulatach:

- samogłoska powinna wystąpić w sylabie akcentowanej – przeważnie w nagłosie;
- ze względu na miejsce akcentu w języku polskim wyrazy muszą zawierać przynajmniej dwie sylaby;
- kontekst spółgłoskowy nie powinien znacząco wpływać na jej barwę (należy unikać kontekstu spółgłosek miękkich i nosowych);
- powinno się uwzględnić te wyrazy, dla których w teście nazywania wystąpiło najmniej braków.

Okazało się, że postulaty powyższe nie zawsze mogły zostać spełnione. Do dalszej analizy wybrano wyrazy: *tata, misie, cukierki, ser, zupa, woda, kot, lew, żaba, ryba, szafa, łóżko, fotel, myje się, guziki, gitara, siedzi* oraz *czyta*. W teście czytania uzyskano wszystkie potrzebne realizacje samogłosek, natomiast w teście nazywania najwięcej braków dotyczyło samogłoski [i]. Niewiele było wyrazów zawierających tę samogłoskę w całym teście, poza tym występowała ona w wyrazach praktycznie nieznanymi dzieciom, jak np. *gitara*, w którym samogłoska /i/ znajduje się w sylabie przedakcentowej. Dla tego wyrazu stwierdzono łącznie 28 braków. Dla wyrazu *guziki* (por. wyżej) wystąpiło 26 braków. Trudnym dla dzieci wyrazem był także czasownik *czyta* (czasowniki w obrazkowym teście nazywania stwarzają na ogół więcej kłopotów).

3. Pomiarów częstotliwości formantowych samogłosek dokonano przy użyciu wbudowanego w system analizatora CSL50 firmy Kay Elemetrics. Łącznie wyko-

nano 1998 pomiarów dla testu czytanego (37 mówców $\times$ 18 wyrazów $\times$ 3 pomiary, w trzech różnych punktach czasowych najbardziej ustalonego stadium przebiegu formantu)⁶ oraz 1707 pomiarów dla testu nazywania (mniejsza liczba pomiarów dla testu nazywania spowodowana była głównie brakiem realizacji niektórych wyrazów).

IV. KLASYFIKACJA ORAZ IDENTYFIKACJA SAMOGŁOSEK W MOWIE DZIECI NIESŁYSZĄCYCH

Dla celów klasyfikacji statystycznej samogłosek wypowiedzianych przez dzieci niesłyszące wykorzystano sieci neuronowe. Metoda ta, jako narzędzie klasyfikacyjne oraz identyfikacyjne, została uprzednio zastosowana dla dorosłych mówców słyszących [por. Jassem, Grygiel – w druku].

1. Podstawowa baza danych

Wyniki pomiarów pierwotnie były zgromadzone w bazie danych zawierającej: (1) typ testu, (2) wyraz, (3) samogłoskę, (4) numer kodowy badanej osoby, (5) płeć badanej osoby, (6) częstotliwości pierwszych czterech formantów oraz ich szerokości wstęp z próbki pierwszej – frakcja ucząca, (7) szerokości wstęp pierwszych czterech formantów oraz ich szerokości wstęp z próbki drugiej – frakcja walidacyjna oraz (8) szerokości wstęp pierwszych czterech formantów oraz ich szerokości wstęp z próbki trzeciej – frakcja testowa. Na podstawie tych danych dla potrzeb klasyfikacji utworzono nową bazę danych.

Fragment nowej bazy danych przedstawiono w tab. 1. Zawiera ona następujące zmienne:

- *Zmienne ilościowe*, zawierające parametry fizyczne (częstotliwości pierwszych czterech formantów oraz szerokości wstęp pierwszych czterech formantów) samogłosek: F1, F2, F3, F4, B1, B2, B3, B4. Wszystkie te zmienne stanowią *parametry wejściowe* w zastosowanych probabilistycznych sieciach neuronowych.

- *Zmienna nominalna „samogłoska”*, mogąca przyjmować następujące wartości: a, e, i, o, u, y (w bazie wartości zmiennej zakodowane są za pomocą liczb 1 ... 6)⁷.

- *Zmienna nominalna „głos”*, której wartość to numer kodowy odpowiadający osobie wypowiadającej dane samogłoski. Zmienna ta przyjmuje 37 różnych wartości.

⁶ W rzeczywistości liczba ta jest mniejsza o 3 pomiary, ponieważ w jednym wyrazie nastąpiła elizja badanej samogłoski.

⁷ W niniejszej pracy dla wygody graficznej samogłoski zapisano ortograficznie oraz uporządkowano według alfabetu, a nie kryteriów artykulacyjnych.

- *Zmienna nominalna „replikacja”*, pełniąca funkcję selektywną (grupującą). Dzieli wszystkie przypadki (inaczej *wiersze* albo *rekordy*) w nowej bazie danych (por. fragment w tab. 1) na trzy grupy: A, B, C, ponieważ dla każdej wypowiedzianej samogłoski dokonano trzech pomiarów, opatrzonych literami: A, B, C. Wiersze opatrzone literą A wykorzystano do trenowania sieci i stanowią one tzw. *frakcję uczącą*. Przypadki opatrzone literą B należą do tzw. *frakcji walidacyjnej* i służą do sprawdzania procesu uczenia. Przypadki oznaczone literą C należą do frakcji testowej i są podstawą uogólnienia modelu.

Jak wspomniano we wstępie, dane zgromadzone w bazie rozpatrywane będą dwojako: (1) dla każdej osoby oddzielnie oraz (2) dla wszystkich osób łącznie.

Tab. 1. Zbiór danych. Zmienne wejściowe i wyjściowe sieci neuronowych (początkowy i końcowy fragment bazy danych)

Frakcja	Typ testu	Wyraz	Samogłoska	Głos	F1	F2	F3	F4	B1	B2	B3	B4
A	czyt.	1	a	1	1027	1620	3789	4287	200	149	398	224
A	spont.	1	a	1	1266	1702	3554	4541	464	545	365	522
B	czyt.	1	a	1	1007	1624	3740	4301	246	225	341	269
B	spont.	1	a	1	1240	1879	3591	4947	291	415	75	1062
C	czyt.	1	a	1	1017	1640	3711	4238	190	177	249	277
C	spont.	1	a	1	1219	1758	3603	4798	301	503	136	1061
.....												
A	czyt.	18	y	37	512	1265	2158	3329	115	333	650	107
A	spont.	18	y	37	465	1339	2567	3761	83	104	511	308
B	czyt.	18	y	37	516	1237	2151	3358	119	293	494	123
B	spont.	18	y	37	452	1301	2504	3608	66	81	391	374
C	czyt.	18	y	37	525	1315	2143	3379	143	306	786	308
C	spont.	18	y	37	436	1311	2625	3877	68	120	564	925

2. Mechanizmy klasyfikacji fonemów

Dysponując pełną informacją zawartą w rekordach typu A i frakcji walidacyjnej B, wybierano dowolny przypadek należący do frakcji testowej C. Na podstawie parametrów liczbowych: F1, F2, F3, F4, B1, B2, B3, B4 tego przypadku sieć neuronowa określa, która samogłoska została wypowiedziana. Do klasyfikacji wzorcowej obiektów (inaczej *przypadków* albo *wierszy* bazy danych) użyto probabilistycznej sieci neuronowej. W wyniku działania sieci neuronowej uzyskano zmienną wyjściową zwaną decyzją.

2.1. Ogólne informacje o budowie i działaniu sieci neuronowej

Sztuczną sieć neuronową można zdefiniować jako oparty na strukturze neuronu układ przetwarzający dane liczbowe, który: (a) przyjmuje informację poprzez wejścia (od kilku do znacznej ich liczby), (b) przetwarza przyjętą informację w sposób

równoległy, tj. równocześnie ze wszystkich wejść, z wykorzystaniem wybranych specjalnych algorytmów oraz (c) podaje na wyjściu pewną pożądaną informację.

Pomiędzy warstwą wejściową jednostek a warstwą wyjściową istnieje prawie we wszystkich faktycznie stosowanych sieciach przynajmniej jedna warstwa pośrednia, zwana ukrytą, ponieważ zachodzące w niej procesy nie są ujawniane. Liczba wejść w elementach przetwarzających warstwy ukrytej jest równa liczbie neuronów w warstwie poprzedniej. Ostatnia warstwa jest warstwą wyjściową. Na jej wejście podawane są wartości pochodzące z ostatniej warstwy ukrytej (w przypadku gdy nie ma warstw ukrytych, to z warstwy wejściowej). Pomiędzy kolejnymi warstwami neuronów poszczególne jednostki są gęsto z sobą wzajemnie połączone w taki sposób, by wyjście jednego neuronu połączone zostało z każdym neuronem warstwy następnej.

Poszczególne jednostki sieci jest funkcjonalnym odwzorowaniem jednego neuronu (komórki nerwowej mózgu) pod względem przyjmowania, „ważenia”, łączenia informacji z kilku lub wielu wejść, przetwarzania połączonej informacji i przesyłania jej do swojego wyjścia. Kompletna sieć neuronowa symuluje działanie mózgu pod względem paralelności przetwarzania oraz zdolności uczenia się na podstawie zadawanych („postrzeganych”) danych.

Wykorzystany program SNN (*Statistical Neural Network*) zawiera wiele typów neuronów: warstwy wejściowej, warstw ukrytych oraz warstw wyjściowych. Pierwsze z nich są bardzo proste w budowie, a każdy z nich ma tylko jedną wartość wyjściową, która w niezmiennionej postaci trafia do drugiej warstwy. Jak widać, w neuronach warstwy wejściowej nie dokonują się żadne obliczenia. Każdy neuron pozostałych warstw ma pewną liczbę połączeń wejściowych wychodzących z warstwy poprzedniej. Z każdym wejściem skojarzona jest pewna liczba zwana wagą. Oprócz tego z każdym neuronem skojarzona jest pewna wartość zwana wartością progową. Informacja wejściowa wprowadzana do neuronu ma postać numeryczną – jest liczbą rzeczywistą, również wagi są liczbami rzeczywistymi. W neuronie zachodzą dwa procesy: (1) agregacja danych wejściowych oraz (2) obliczenie wartości funkcji aktywacji (funkcji przejścia).

Agregacja realizowana jest przez funkcję *potencjału postsynaptycznego* (PSP). W przypadku referowanych klasyfikacji zastosowano dwa rodzaje funkcji PSP:

- *liniową* – agregacja dokonywana jest przez sumowanie ważonych wejść, a następnie przez odejmowanie od niej wartości progowej. Obliczona w ten sposób wartość stanowi argument funkcji aktywacji. W terminologii algebry liniowej obliczany jest iloczyn skalarny wektora wag przez wektor wartości wejściowych i od tego odejmuje się parametr progowy. (Neurony mające liniowe funkcje PSP przyjmują takie same wartości dla wektorów leżących na tych samych hiperpłaszczyznach);

- *radialną* – oblicza kwadrat odległości pomiędzy dwoma punktami w N-wymiarowej przestrzeni (gdzie N – jest liczbą wejść; w naszym przypadku N = 8).

Pierwszym z punktów jest wektor wartości wejściowych, a drugim wektor wag. Kwadrat odległości jest następnie mnożony przez wartość progową (wartość progowa określa rozrzut). Obliczona w ten sposób wartość stanowi wejście funkcji aktywacji. Neurony mające radialną funkcję PSP mają takie same wartości wyjściowe dla punktów leżących na sferze ze środkiem wyznaczonym przez wektor wag. Neurony te przeprowadzają klasyfikację, mierząc odległość punktu wartości wejściowych od przykładowego punktu. Przykładowe punkty są zapamiętane na wejściach do neuronów.

Innymi słowy, wielkości wyjściowe poprzedniej warstwy oraz wagi powiązane z połączeniami są wykorzystane do obliczenia wartości funkcji potencjału postsynaptycznego (do obliczenia pojedynczej wartości). Stanowi ona argument funkcji aktywacji.

Funkcja aktywacji oblicza wartość wyjściową dla neuronu. Wykorzystano następujące funkcje aktywacji dostępne w SNN:

- *liniową* $f(x) = x$ o wartościach w przedziale $(-\infty, +\infty)$;
- $f(x) = \exp(-x)$ – funkcja $\exp(-x)$ jest stosowana w neuronach z funkcją potencjałową radialną. Złożenie tych dwóch funkcji tworzy funkcję dzwonową, skoncentrowaną na wektorze wag. Parametr progowy neuronu decyduje o kształcie tego dzwonu;

- *SoftMax* $f(x) = \exp(x) / \text{Suma}(\exp(x_1); \exp(x_2); \dots; \exp(x_p))$ – jest to funkcja wykładnicza, której wartość jest tak znormalizowana, że suma wartości aktywacji z całej warstwy jest równa 1. Jest ona wykorzystywana w warstwach wyjściowych w problemach klasyfikacyjnych i pozwala interpretować wartości wyjściowe neuronów jako prawdopodobieństwa przynależności do klas;

- *UnitSum* – jest funkcją aktywacji, stosowaną w warstwie wyjściowej sieci probabilistycznej. Wektor wyjściowy ostatniej warstwy jest tak przeskalowany, aby suma wartości wektora dawała 1. To przekształcenie jest przydatne w przypadku, gdy wartości wyjściowe interpretuje się jako prawdopodobieństwa przynależności obiektu do poszczególnych klas.

2.2. Przetwarzanie informacji w sieci neuronowej

W niniejszym paragrafie omówione zostaną pokrótce trzy etapy przetwarzania informacji: (1) wstępne przetwarzanie danych oraz przetwarzanie danych wyjściowych, (2) klasyfikacja za pomocą neuronowej sieci probabilistycznej oraz (3) proces uczenia probabilistycznej sieci neuronowej.

1. Wstępne przetwarzanie danych oraz przetwarzanie danych wyjściowych (*pre-, postprocessing*)

Wymóg korzystania wyłącznie z informacji w postaci numerycznej pociąga za sobą konieczność stosowania w algorytmach etapu wstępnego przetwarzania

danych wejściowych przed ich podaniem na wejścia sieci (*preprocesing*) i procesu algorytmicznego przekształcenia oraz odpowiedniej interpretacji danych wyjściowych otrzymywanych z sieci (*postprocesing*). W konkretnym przypadku obecnych obliczeń na proces działania sieci złożyły się następujące etapy:

- *Konwersja zmiennych nominalnych* do postaci numerycznej (do reprezentacji służy skalar albo wektor). Dla dwóch jakościowych zmiennych nominalnych: „samogłoski” oraz „głosu” użyto w konwersji metody „jeden-z-N”, polegającej na wykorzystaniu kilku zmiennych numerycznych do reprezentacji tych zmiennych, czyli każda z wartości zmiennych nominalnych jest reprezentowana przez wektor, w którym liczba elementów jest równa liczbie wszystkich możliwych wartości. W wektorze tym jeden z elementów przyjmuje wartość 1, pozostałe zaś przyjmują wartość 0. W przypadku zmiennej „samogłoska” możemy mieć następujący sposób konwersji: wartości /a/ odpowiada wektor (1,0,0,0,0,0), wartości /e/ odpowiada wektor (0,1,0,0,0,0). Podobny sposób zastosowano do zmiennej „głos”.

- *Skalowanie zmiennych numerycznych*. Zmienne numeryczne: F1, F2, F3, F4, B1, B2, B3, B4 zostały tak przeskalowane, by ich wartości (dla zbioru uczącego) znajdowały się w przedziale [0,1]. Skalowanie odbywało się zgodnie z formułą:

Wartość przekształcona = Wartość pierwotna * Scale + Shift

Scale = 1/(Max-Min); Shift = -Min/(Max-Min); gdzie: Max, Min są wartościami policzonymi dla danej zmiennej wejściowej dla frakcji uczącej (dla każdej ze zmiennych zastosowano inne przekształcenie).

- *Wprowadzenie danych na wejścia sieci*.

- *Przetwarzanie przez sieć*.

- *Uzyskanie wartości wyjściowych sieci*. Wartości wyjściowe ostatniej warstwy można interpretować jako prawdopodobieństwa przynależności obiektów do poszczególnych klas. W przypadku klasyfikacji na większą liczbę klas wartości wyjściowe sieci mogą być interpretowane jako prawdopodobieństwa przynależności do poszczególnych klas, jeżeli neurony warstwy wyjściowej będą mieć funkcję aktywacji typu UnitSum albo Softmax.

- *Zamiana wartości numerycznych do postaci wyrażonej w skali nominalnej*. Na wyjściach neuronów ostatniej warstwy uzyskujemy liczby, mimo że klasy reprezentowane są przez zmienne nominalne. W zmiennej nominalnej „samogłoska” każda z możliwych wartości reprezentuje jedną klasę.

Decyzję o przynależności obiektu do danej klasy podjęto na podstawie porównania wartości generowanej na wyjściu neuronu z ustalonymi poziomami ufności, do których zalicza się poziom zaakceptowania oraz poziom odrzucenia. Sposób interpretacji uzależniony jest od liczby neuronów w warstwie wyjściowej:

- *Jeden neuron w warstwie wyjściowej* – klasyfikacja na dwie klasy. Jeżeli wartość wyjściowa neuronu jest wyższa od poziomu zaakceptowania, obiekt jest zaliczany do klasy „1”, natomiast jeżeli wartość wyjściowa jest niższa od poziomu odrzucenia, obiekt jest klasyfikowany do klasy „0”. Jeżeli nie są spełnione powyższe warunki, wynik klasyfikacji nie jest ustalony.

- *Przynajmniej dwa neurony w warstwie wyjściowej* – klasyfikacja na większą liczbę klas. Jeżeli wartość wyjściowa dla jednego neuronu jest wyższa od poziomu zaakceptowania i dla wszystkich pozostałych jest niższa od poziomu odrzucenia, to obiekt jest zaliczany do klasy, dla której przekroczony został poziom zaakceptowania; w każdym innym przypadku sieć nie jest w stanie zaklasyfikować obiektu. Zawsze poziom zaakceptowania powinien być wyższy od poziomu odrzucenia. Jeżeli poziom odrzucenia jest wyższy od poziomu zaakceptowania, to program ignoruje wartość poziomu odrzucenia i traktuje go tak, jakby był równy poziomowi zaakceptowania.

2. Probabilistyczna sieć neuronowa (PNN) w zadaniach klasyfikacyjnych

W konkretnym przypadku obecnej klasyfikacji PNN składa się z trzech warstw i ma następującą architekturę:

- *Warstwa wejściowa*. Każdy neuron wchodzący w skład warstwy wejściowej ma jedno wejście. Wewnątrz neuronów warstwy wejściowej może następować liniowe przekształcanie wprowadzanych danych, ale bardzo często przekształcenie to ma charakter tożsamościowy. Wartości otrzymywane na wyjściach poszczególnych neuronów przekazywane są do każdego neuronu następnej warstwy radialnej.

- *Warstwa ukryta składająca się z neuronów radialnych*. Działanie neuronu radialnego polega w pierwszej kolejności na obliczeniu wartości funkcji PSP (Post Synaptic Potential Function) – obliczany jest kwadrat odległości pomiędzy wektorem wejściowym a wektorem wag. Wynik funkcji PSP przemnożony przez stałą progową P jest argumentem funkcji aktywacji (Act fn). Parametr threshold = $1/(2 * \text{smoothing}^2)$ jest wyliczany na podstawie parametru wygładzania zadanego w algorytmie uczącym. W naszym przypadku funkcją aktywacji w neuronach radialnych jest funkcja w kształcie dzwonu $\exp(-x)$. Wartość wyjściowa neuronu radialnego jest tym większa, im wektor wartości wejściowych jest bliższy wektorowi wag. Jako wag stosuje się wektory wejściowe służące do nauki sieci.

- *Warstwa wyjściowa klasyfikująca*. Kolejne neurony reprezentują wyróżnione klasy (określone przez zmienną nominalną). Neurony warstwy wyjściowej sumują wartości wyjściowe tych neuronów radialnych, których wagi (równe elementom uczącym) należą do danej klasy. W ten sposób obliczana jest wartość, która jest proporcjonalna do prawdopodobieństwa przynależności obiektu wyjściowego do danej klasy. Wagi neuronów klasyfikujących mogą zawierać informacje dotyczące prawdopodobieństw *a priori* przynależności obiektów do kolejnych klas. Prawdopodobieństwa te podawane są jako parametry procedury uczącej. Funkcja potencjału postsynaptycznego w warstwie wyjściowej jest zatem funkcją liniową. Stosując funkcję aktywacji *UnitSum* albo *SoftMax*, można spowodować, aby wartości warstwy klasyfikującej sumowały się do 1 i aby można było interpretować wartości wyjściowe tej warstwy jako prawdopodobieństwa przynależności do danej klasy.

3. Proces uczenia probabilistycznej sieci neuronowej

Działanie sieci neuronowej uzależnione jest od wartości wag i progów, które wyznacza się w trakcie nauki sieci neuronowej. Naukę probabilistycznej sieci neuronowej przeprowadza się, nadając wartość parametrowi *wygładzanie*, na podstawie którego obliczana jest wartość progowa dla wszystkich neuronów warstwy ukrytej, natomiast jako wagi przyjmowane są wartości wejściowe obiektów z ciągu uczącego. Takie próby przeprowadza się dla różnych parametrów *wygładzania* i wybiera się ten, dla którego wartość „błędu” jest najmniejsza dla frakcji uczącej i walidacyjnej.

Funkcja błędu służy do określenia, jak dobrze funkcjonuje sieć podczas treningu i testowania. Wykorzystano funkcję błędu, będącą sumą kwadratów różnic pomiędzy wartościami zmiennych wyjściowych a wartościami otrzymanymi z warstwy wyjściowej.

3. Wyniki klasyfikacji

W rezultacie procedury omówionej w par. IV.2 otrzymano łącznie 39 tabel klasyfikacyjnych polskich samogłosek w wypowiedziach starszych dzieci niesłyszących. 37 tabel zawiera wyniki kolejno dla każdego mówcy. Dwie tabele przedstawiają klasyfikacje, w których uwzględniono wyniki wszystkich mówców. Poniżej przedstawiono przykładowo trzy klasyfikacje: dla pojedynczego mówcy (tab. 2), dla samogłosek wypowiedzianych przez wszystkich mówców (tab. 3) oraz dla samogłosek wypowiedzianych przez wszystkich mówców z uwzględnieniem dodatkowej zmiennej określonej jako głos (tab. 4).

Dla 33 dzieci uzyskano trenowanie sieci na poziomie 100%. Dla pozostałych czterech osób poprawność identyfikacji samogłosek frakcji uczącej wynosiła od 94 do 97%. Biorąc pod uwagę niezbyt dużą liczbę danych wejściowych, można przyjąć, że sieć bardzo dobrze nauczyła się identyfikowania 6 samogłosek. Zadaniem frakcji walidacyjnej było ustalenie, czy sieć nie działa na zasadzie przypadkowego doboru przykładów, czy dane we frakcji uczącej nie były obciążone. Dla 8 osób uzyskano poprawność identyfikacji na poziomie 50-70%, ale dla połowy badanych dzieci poprawność identyfikacji wynosi w tej frakcji od 81 do 100%. Uzyskany procent poprawnych identyfikacji we frakcji walidacyjnej oznacza, że sieć nie dla wszystkich mówców w pełni spełniła swoje zadanie. W warstwie testowej zakres poprawnych rozpoznań samogłosek dla poszczególnych dzieci waha się w granicach 34,3-91,4%, co oznacza, że poziom opanowania poprawnej artykulacji samogłosek przez dzieci niesłyszące jest bardzo zróżnicowany. Najlepsze wyniki rozpoznawania przez sieć uzyskano dla samogłoski wysokiej i tylnej /u/ – prawie 73% poprawnych identyfikacji. Rozpoznawanie pozostałych samogłosek oscylowało na poziomie ok. 50%, przy czym dla wysokiej i przedniej samogłoski /i/ uzyskano nieco wyższy procent poprawnych klasyfikacji, tj. 54,5%, a dla /o/ najniższy – 47,9%.

Tab. 2. Przykład klasyfikacji samogłosek dla jednego mówcy – 06 (chłopca w wieku 13 lat)

Frakcja	Samogłoska	Razem	Poprawnie	Błędnie	Nieokreślone	a	e	i	o	u	y	Poprawnie (%)	Poprawnie (%)
Ucząca	a	6	6	0	0	6	0	0	0	0	0	100%	100%
	e	6	6	0	0	0	6	0	0	0	0	100%	
	i	6	6	0	0	0	0	6	0	0	0	100%	
	o	6	6	0	0	0	0	0	6	0	0	100%	
	u	5	5	0	0	0	0	0	0	5	0	100%	
	y	6	6	0	0	0	0	0	0	0	6	100%	
Walidacyjna	a	6	6	0	0	6	0	0	0	0	0	100%	94,29%
	e	6	5	0	1	0	5	0	0	0	0	83,33%	
	i	6	6	0	0	0	0	6	0	0	0	100%	
	o	6	6	0	0	0	0	0	6	0	0	100%	
	u	5	5	0	0	0	0	0	0	5	0	100%	
	y	6	5	1	0	0	0	1	0	0	5	83,33%	
Testowa	a	6	3	3	0	3	2	0	1	0	0	50,00%	68,57%
	e	6	4	1	1	1	4	0	0	0	0	66,67%	
	i	6	4	0	2	0	0	4	0	0	0	66,67%	
	o	6	4	1	1	1	0	0	4	0	0	66,67%	
	u	5	3	2	0	0	0	0	2	3	0	60,00%	
	y	6	6	0	0	0	0	0	0	0	6	100%	

Zbiór danych: nowa baza danych dla głosu 06;
 Zmienne wejściowe: F1, F2, F3, F4, B1, B2, B3, B4;
 Zmienna wyjściowa: **samogłoska**;
 Sieć neuronowa: sieć probabilistyczna (PNN);
 Funkcja błędu: suma kwadratów;
 Warstwa 1. Funkcja aktywacji: liniowa; funkcja potencjałowa liniowa;
 Warstwa 2. Funkcja aktywacji: $f(x) = \exp(-x)$; funkcja potencjałowa radialna;
 Warstwa 3. Funkcja aktywacji: *UnitSum*; funkcja potencjałowa liniowa;
Pre-post processing
 Poziom akceptacji: 0.85;
 Poziom odrzucenia: 0.2;
 Parametry uczenia sieci: *wygładzanie* = 0.04, próg = 312.5

Średnia poprawność klasyfikacji przez sieć uczoną tylko samogłoskowymi częstotliwościami formantowymi i szerokościami ich wstęg wyniosła 54,4%. Wśród błędnych klasyfikacji największy odsetek stanowiły klasyfikacje nieokreślone (od 9,8% dla /u/ do 16,2% dla /e/ i /o/), tzn. SNN nie potrafiła na podstawie danych z frakcji uczącej określić przynależności danego widma z frakcji testującej do któregoś z polskich fonemów samogłoskowych. Znaczna liczba samogłosek nieokreślonych może oznaczać, że dzieci niesłyszące popełniając błędy w wymowie samogłosek najczęściej nie dokonują substytucji polegających na wstawianiu dźwięku należącego do innego fonemu języka polskiego, lecz realizują dźwięki pośrednie, np. w przypadku /o/ jest to dźwięk pomiędzy polskim [a] i [o]

[por. też: Kurkowski 1996]. W wielu przypadkach struktura akustyczna była bardzo niestabilna i stwarzała trudności w wyznaczeniu przedziału 30 ms, służącego do zdjęcia w nim trzech widm chwilowych danego dźwięku ze względnie wyraźnymi częstotliwościami formantowymi.

Tab. 3. Klasyfikacja samogłosek na podstawie wypowiedzi wszystkich badanych dzieci

Frakcja	Samogłoska	Razem	Poprawnie	Błędnie	Nieokreślone	a	e	i	o	u	y	Poprawnie (%)	Poprawnie (%)
Ucząca	a	222	221	0	1	221	0	0	0	0	0	100%	98,87%
	e	216	214	0	2	0	214	0	0	0	0	99,1%	
	i	160	158	0	2	0	0	158	0	0	0	98,8%	
	o	217	211	1	5	0	0	0	211	1	0	97,2%	
	u	213	212	1	0	0	0	0	0	212	1	100%	
	y	206	204	0	2	0	0	0	0	0	204	99,0%	
Walidacyjna	a	222	136	55	31	136	14	6	16	10	9	61,26%	64,96%
	e	216	122	59	35	25	122	9	13	2	10	56,48%	
	i	159	105	32	22	4	8	105	6	2	12	66,04%	
	o	217	141	51	25	25	9	2	141	10	5	64,98%	
	u	213	171	28	14	6	4	2	9	171	7	80,28%	
	y	206	126	52	28	8	13	11	8	12	126	61,17%	
Testowa	a	222	112	79	31	112	25	4	29	11	10	50,45%	54,42%
	e	216	109	72	35	21	109	4	24	6	17	50,46%	
	i	160	87	49	24	2	15	87	4	10	18	54,38%	
	o	217	104	78	35	26	17	3	104	19	13	47,93%	
	u	213	155	37	21	1	5	4	18	155	9	72,77%	
	y	205	104	72	29	10	16	25	9	12	104	50,73%	

Zbiór danych: nowa baza danych;
 Zmienne wejściowe: F1, F2, F3, F4, B1, B2, B3, B4;
 Zmienna wyjściowa: **samogłoska**;
 Sieć neuronowa: sieć probabilistyczna (PNN);
 Funkcja błędu: suma kwadratów;
 Warstwa 1. Funkcja aktywacji: liniowa; funkcja potencjałowa liniowa;
 Warstwa 2. Funkcja aktywacji: $f(x) = \exp(-x)$; funkcja potencjałowa radialna;
 Warstwa 3. Funkcja aktywacji: *UnitSum*; funkcja potencjałowa liniowa;
Pre-post processing
 Poziom akceptacji: 0.39;
 Poziom odrzucenia: 0.39;
 Parametry uczenia sieci: *smoothing* = 0.04, *threshold* = 312.5

Tab. 4. Klasyfikacja samogłosek wymówionych przez wszystkie dzieci z uwzględnieniem czynnika mówcy (dodatkowa zmienna wejściowa: głos)

Frakcja	Samogłoska	Razem	Poprawnie	Błędnie	Nieokreślone	a	e	i	o	u	y	Poprawnie (%)	Poprawnie (%)
Ucząca	a	222	221	0	1	221	0	0	0	0	0	99,55%	99,35%
	e	216	215	0	1	0	215	0	0	0	0	99,54%	
	i	160	158	0	2	0	0	158	0	0	0	98,75%	
	o	217	216	0	1	0	0	0	216	0	0	99,54%	
	u	213	212	0	1	0	0	0	0	212	0	99,53%	
	y	206	204	0	2	0	0	0	0	0	204	99,03%	
Walidacyjna	a	222	173	37	12	173	10	0	18	3	6	77,93%	80,05%
	e	216	171	36	9	12	171	5	9	1	9	79,17%	
	i	159	138	15	6	3	4	138	2	1	5	86,79%	
	o	217	165	44	8	19	8	4	165	7	6	76,04%	
	u	213	189	20	4	3	3	2	9	189	3	88,73%	
	y	206	151	37	18	8	8	11	5	5	151	73,30%	
Testowa	a	222	156	49	17	156	16	2	23	3	5	70,27%	70,15%
	e	216	136	64	16	28	136	5	13	8	10	62,96%	
	i	160	120	27	13	2	13	120	2	2	8	75,00%	
	o	217	144	58	15	18	18	6	144	10	6	66,36%	
	u	213	175	29	9	3	1	5	14	175	6	82,16%	
	y	205	134	52	19	6	15	15	7	9	134	65,37%	

Zbiór danych: nowa baza danych;
 Zmienne wejściowe: F1, F2, F3, F4, B1, B2, B3, B4 oraz dodatkowa zmienna głos;
 Zmienna wyjściowa: **samogłoska**;
 Sieć neuronowa: sieć probabilistyczna (PNN);
 Funkcja błędu: suma kwadratów;
 Warstwa 1. Funkcja aktywacji: liniowa; funkcja potencjałowa liniowa;
 Warstwa 2. Funkcja aktywacji: $f(x) = \exp(-x)$; funkcja potencjałowa radialna;
 Warstwa 3. Funkcja aktywacji: *UnitSum*; funkcja potencjałowa liniowa;
Pre-post processing
 Poziom akceptacji: 0.48;
 Poziom odrzucenia: 0.48;
 Parametry uczenia sieci: *smoothing* = 0.3, *threshold* = 200

Dane zawarte w tab. 4 przedstawiają klasyfikację samogłosek dla wszystkich mówców z uwzględnieniem dodatkowej zmiennej określonej jako głos, co w istotny sposób zwiększyło procent poprawnych rozpoznań – ogółem o 16% (do 70,15%). Najbardziej spadła liczba nieokreślonych klasyfikacji – do poziomu kilku procent, szczególnie dla /e/ (z 16,2% do 7,4%) i dla /o/ (z 16% do 6,9%), czyli dla samogłosek klasyfikowanych najgorzej. Najbardziej wzrosła poprawność klasyfikacji dla /o/ (o 18,43%) i /a/ (o 20%).

W przypadku mówców słyszących zastosowanie dodatkowej zmiennej polepszyło ostateczny poziom rozpoznawania samogłosek zaledwie o 2% (z 91% do

93%) [por. Jassem, Grygiel – w druku]. Jest to dalszy dowód na to, że wymowa samogłosek przez dzieci niesłyszące jest znacznie mniej ustabilizowana niż przez dzieci w tym samym wieku z normalnym słuchem.

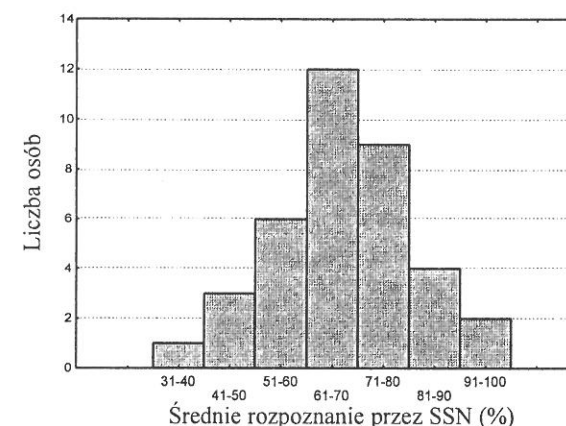
V. INTERPRETACJA UZYSKANYCH WYNIKÓW

Stworzenie sztucznej sieci neuronowej (SSN) dla każdego mówcy z osobną wiązało się z ryzykiem niedostatecznego treningu sieci ze względu na zbyt małą liczbę danych, ale pozwoliło wstępnie na:

- 1) ustalenie rankingu mówców ze względu na uzyskaną przez sieć poprawność rozpoznania zrealizowanych przez nich samogłosek w różnym kontekście spółgłoskowym;
- 2) zbadanie związku pomiędzy wynikami rozpoznania samogłosek dla poszczególnego dziecka a oceną zrozumiałości jego mowy na podstawie wyrazowego testu zrozumiałości;
- 3) analizę czynników w dotychczasowej rehabilitacji i edukacji dzieci, które mogłyby mieć wpływ na jakość wytwarzanych przez nie samogłosek;
- 4) ogólną charakterystykę artykulacji samogłosek przez wszystkie dzieci niesłyszące, które wzięły udział w badaniach.

1. Rozpoznawanie samogłosek przez SNN dla poszczególnych dzieci

Jak stwierdzono w rozdz. IV.3, rozpoznania samogłosek dla poszczególnych mówców różnią się znacznie. Na ryc. 1 przedstawiono histogram średniej poprawności rozpoznania wszystkich 6 samogłosek przez poszczególnych mówców. Cztery osoby osiągnęły wynik poniżej 50% poprawnych klasyfikacji, natomiast dla ponad jednej trzeciej uczestników badania uzyskano poprawność w zakresie 61-70%. Jak wspomniano powyżej, dwie osoby przekroczyły próg 90% i zbliżyły się do rezultatów uzyskanych dla dorosłych osób z normalnym słuchem [por. Jassem, Grygiel – w druku]. Uśredniony wynik dla wszystkich 6 samogłosek nie oznacza oczywiście, że wszystkie one zostały dla poszczególnych mówców rozpoznawane w takim samym stopniu. Rozpiętość wyników jest znaczna. Dla 14 osób żadna z samogłosek nie została rozpoznana w 100%, dla 12 innych dzieci była tylko jedna taka głoska (najczęściej /u/ – cztery osoby, /i/ oraz /y/ – po trzy osoby, /e/ oraz /a/ – po jednej osobie; żadna z tych osób nie uzyskała 100% poprawności rozpoznania samogłoski /o/).



Ryc. 1. Rozkład średnich poprawnych rozpoznań wszystkich samogłosek przez SSN dla wszystkich 37 badanych dzieci

W przypadku najlepszego mówcy (07 – chłopca w wieku 12 lat, poprawność klasyfikacji 91,4%) błędy wystąpiły w przypadku samogłoski /i/ oraz /o/. Dziecko oznaczone symbolem 16, dla którego poprawność rozpoznania wynosiła zaledwie 34,3%, wymawiało samogłoskę /o/ w taki sposób, że sieć nie rozpoznała jej ani razu poprawnie. Jest to jedyny wynik 0% w całym materiale. Pozostałe samogłoski dla tego dziecka rozpoznane zostały poprawnie w następującym stopniu: /u/ – 83%, /a/ – 50%, /i/ – 40% oraz /y, e/ – 16%.

Tab. 5. Klasyfikacja samogłosek przez SSN poniżej progu 50% u poszczególnych mówców

Samogłoska	Kod dziecka
a	12, 21, 33, 34
e	10, 12, 16, 19, 24, 34,
i	9, 12, 16
o	1, 10, 12, 16, 22, 35
u	8, 19, 21, 24
y	15, 16, 27, 30, 35, 37

Dane zawarte w tab. 5 stanowią dalszy dowód na to, że samogłoska /o/ jest względnie trudna dla dzieci niesłyszących, podobnie jak /e/ czy /y/, natomiast lepiej są artykułowane samogłoski skrajne: przednia – wysoka /i/, tylna – wysoka /u/ oraz nieco gorzej środkowa – niska /a/. Jednakże różnice międzysobnicze są znaczne.

2. Rozpoznawanie samogłosek przez SSN a test zrozumiałości wyrazowej dla poszczególnych mówców

Dla badanych dzieci przeprowadzono wyrazowy test zrozumiałości mowy [por. Kleśta 2002]. Zadaniem testu było określenie poziomu zrozumiałości poszczególnych 37 mówców. Badanie poprawnego rozpoznania wyrazów bez kontekstu zdaniowego i sytuacyjnego jest ostrym kryterium ustalania zrozumiałości mowy, zwłaszcza że jeden niepoprawnie zidentyfikowany segment (fonem) powoduje błędne odebranie całego wyrazu. Jak stwierdzono w par. II, średnia zrozumiałość badanej grupy dzieci była niska. Niewątpliwie na taki wynik decydujący wpływ miała przede wszystkim wadliwa artykulacja spółgłosek, ponieważ z punktu widzenia teorii informacji to one są najbardziej istotnym elementem kształtu dźwiękowego wyrazu w procesie jego rozpoznawania. Spółgłoski przenoszą znaczenie leksykalne wyrazu i nie wystarcza prawidłowa artykulacja samogłosek, aby być dobrze rozumianym. Już w latach czterdziestych XX w. Hudgins i Numbers [za: Monsen 1974] stwierdzili dla języka angielskiego, że osoby niesłyszące popełniają dwa razy mniej błędów w produkcji samogłosek niż spółgłosek. Należy jednakże podkreślić znaczącą rolę samogłosek w wypowiedziach jako najbardziej podstawowych dźwięków mowy, które pokrywają ok. 40% tekstu [Ryalls 1996] oraz stanowią podstawowy „nośnik” cech prozodycznych. Istotny jest także samogłoskowy „substrat” w produkcji większości spółgłosek.

W przypadku badanych dzieci na całkowity kształt artykulacyjny wyrazów złożyło się wiele zjawisk fonetycznych: uproszczenia i elizje sylab lub głosek, wstawienia sylab lub pojedynczych głosek oraz substytucje. Wszystkie te zjawiska występowały w nagłosie, śródgłosie i wygłosie wyrazów [por. Łobacz 2002], co znacznie utrudniało prawidłowe ich zrozumienie. Celem porównania obecnie uzyskanych wyników rozpoznawania samogłosek z ogólnym poziomem zrozumiałości mowy poszczególnych dzieci było sprawdzenie, czy istnieje bezpośredni związek pomiędzy poziomem artykulacji samogłosek a ogólną zrozumiałością ich wyrazowych wypowiedzi. W tab. 6 przedstawiono dla każdego mówcy procent poprawnych rozpoznań samogłosek przez sieć oraz poziom zrozumiałości mowy, także wyrażony w procentach. Dane testu zrozumiałości zostały zaczerpnięte z pracy Kleśty [2002] i stanowią uśrednione wartości dla obu testów: nazywania i czytania.

Na podstawie danych przedstawionych w tab. 6 stwierdzono, że istnieje tylko słaba korelacja pomiędzy poprawnością klasyfikacji samogłosek przez sieć a ogólną zrozumiałością mowy poszczególnych mówców (współczynnik korelacji wynosi 0,39). Jedynie osoby o najbardziej poprawnej klasyfikacji samogłosek charakteryzują się względnie wysoką zrozumiałością (por. np. mówca 07 – o najlepszej klasyfikacji samogłosek przez sieć – uzyskał drugi wynik w teście zrozumiałości mowy; mówca 14, który także osiągnął ponad 90-procentowe rozpo-

Tab. 6. Porównanie klasyfikacji samogłosek przez SNN z wynikami testu percepcyjnego zrozumiałości mowy dla każdego dziecka

MÓWCA	POPRAWNA KLASYFIKACJA SAMOGŁOSEK PRZEZ SIEĆ (%)	POPRAWNE ROZPOZNANIA WYRAZÓW PRZEZ SŁYSZĄCYCH (%)
7	91,4	47,8
14	91,2	38,3
13	86,1	50,6
17	85,7	36,1
36	82,3	23,6
11	82,3	27,3
29	79,4	8,8
37	75,8	9,5
26	75,8	7,1
3	75,8	9,6
33	74,3	18,5
20	74,3	9,9
30	73,5	44,8
9	72,7	32,5
23	70,6	11,9
6	68,6	42,3
32	67,6	27,7
19	66,7	34,2
18	66,7	7,3
15	66,7	24,2
4	64,7	13,9
22	64,5	7,4
2	64,5	17,5
5	64,3	11
25	63,6	8
8	63,6	21,6
1	63,6	26,7
28	60	17,7
35	59,4	12,5
27	59,4	8,3
31	56,2	20,3
34	54,8	17,4
24	51,5	16
21	50	2,4
10	50	36,5
12	44,4	28,7
16	34,3	11,5

nawanie samogłosek, pod względem ogólnej zrozumiałości zajął pozycję piątą, a osoba 13 – rozpoznawanie samogłosek przez SNN na poziomie 86% – w rankingu zrozumiałości uzyskała najlepszy wynik). Dla pozostałych badanych osób analizowana zależność zachodzi w znacznie mniejszym stopniu. Mówca 36 uzyskał poprawność klasyfikacji samogłosek na poziomie powyżej 82%, ale w rankingu ogólnej zrozumiałości zajmuje dopiero 15 pozycję. Najgorszy wynik zrozumiałości mowy, wynoszący zaledwie 2,4% (dziecko 21), potwierdzony został również niską klasyfikacją samogłosek przez SNN, która w tym przypadku wyniosła 50%, ale z kolei najgorszej klasyfikacji samogłosek, wynoszącej 34% dla mówcy 16, odpowiada zrozumiałość na poziomie prawie 12%.

3. Wpływ niektórych czynników fizycznych, rozwojowych i procesu rehabilitacji na poprawność artykulacyjną samogłosek

Jak wykazano w par. II, badana grupa dzieci niesłyszących była niejednorodna pod względem wielu czynników, a istotnym powodem wzięcia udziału w obecnych badaniach była ustalona przez pedagogów wystarczająca sprawność posługiwania się mową. Mimo trudności z ustaleniem pełnej dokumentacji badanych dzieci starano się określić zależność pomiędzy poprawnością rozpoznawania samogłosek a następującymi czynnikami: wiekiem, płcią, przyczyną utraty słuchu, momentem utraty słuchu, poziomem znajomości języka migowego, przebiegiem rehabilitacji oraz momentem zaopatrzenia w protezy słuchowe (jest to jedyny czynnik rehabilitacyjny, dla którego udało się zgromadzić pełną dokumentację).

Dla tak zróżnicowanej grupy nie stwierdzono żadnej statystycznie istotnej zależności pomiędzy uzyskanymi wynikami rozpoznawania samogłosek a którymkolwiek z analizowanych czynników, wobec czego ewentualne zależności zostaną zilustrowane na przykładzie wybranych dzieci. Szczególnie interesujące są dwa skrajne przypadki. Mówca 16, którego samogłoski zostały rozpoznane przez sieć najgorzej – średnio na poziomie 34% poprawnych klasyfikacji, w tym 0% dla samogłoski /o/ – jest niesłyszący od urodzenia (przyczyna nieznana), otrzymał najpóźniej ze wszystkich dzieci protezę słuchową (w ósmym roku życia), przed rozpoczęciem nauki w szkole nie korzystał z żadnych form rehabilitacji słuchu, a jego wyniki testu psychologiczno-rozwojowego były w dolnych granicach normy. Natomiast mówca 07, zajmujący pierwsze miejsce w rankingu, utracił słuch w 13 miesiącu życia na skutek zapalenia opon mózgowych, po ukończeniu czwartego roku życia został zaopatrzony w protezę słuchową i równocześnie rozpoczął rehabilitację w przedszkolu dla niesłyszących oraz stale współpracował z logopedą. Obaj chłopcy mają ubytek słuchu powyżej 90 dB (podobnie jak wszystkie inne badane dzieci), znają dobrze język migowy, mają słyszących rodziców i rodzeństwo z normalnym słuchem. Dwie dziewczynki w wieku 13 lat, dla których otrzymano klasyfikacje na poziomie 91,2% oraz 86,1% (osoby: 14 i 13), utraciły słuch

we wczesnym dzieciństwie na skutek nie sprecyzowanej w dokumentacji choroby, mają słyszące rodzeństwo, zaopatrzone zostały w protezy słuchowe w 4 i 5 roku życia, nie korzystały w wieku przedszkolnym z żadnej rehabilitacji logopedycznej, a tylko jedna z nich posługuje się dobrze językiem migowym. Z kolei osoby 10 i 12 o niskiej poprawności klasyfikacji (50% i 44%) to dziewczynka w wieku 14 lat oraz chłopiec w wieku 13 lat, dzieci, które zaopatrzone zostały w protezy słuchowe względnie wcześnie, bo w 3 i 4 roku życia, język migowy znają słabo i jeszcze się go uczą, mają słyszące rodzeństwo. Osoba 12 utraciła słuch we wczesnym dzieciństwie, a osoba 10 jest niesłysząca od urodzenia. Obie osoby nie były w wieku przedszkolnym poddane żadnej specjalistycznej rehabilitacji i korzystały tylko z badań okresowych w poradni foniatrycznej.

Z powyższych opisów wybranych przypadków charakteryzujących się najlepszą oraz najłabszą umiejętnością artykulacyjną w zakresie samogłosek wyłaniają się pewne regularności: decydujący wpływ na wysoką sprawność mówienia ma na ogół utrata słuchu dopiero w kilka miesięcy po urodzeniu oraz prawidłowa rehabilitacja (por. osoba 07). Należy jeszcze uznać za wysoce istotny czynnik osobniczy, polegający na chęci posługiwania się mową. Niektóre badane dzieci ten sposób komunikacji wykorzystywały głównie na lekcjach, ponieważ było to obowiązkowe, podczas gdy na przerwach najczęściej posługiwały się językiem migowym. Wiele z nich mieszkało w internacie, w którym ten ostatni rodzaj komunikacji był zdecydowanie przeważający. Im bardziej aktywnie dziecko uczestniczyło w przebiegu lekcji, tym lepsze uzyskało rezultaty w opanowaniu sprawności artykulacyjnej.

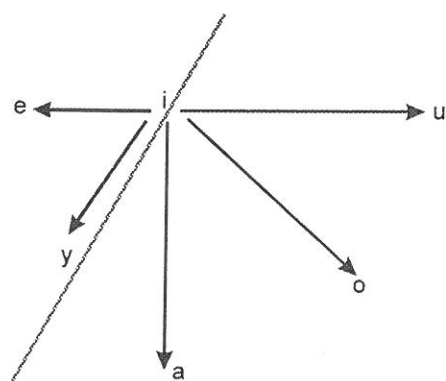
Ponieważ badane dzieci niewątpliwie nie stanowią wzorcowej grupy pod względem procesu rehabilitacji, szczególnie prowadzonej w wieku przed pójściem do szkoły, uzyskane wyniki należy uznać za zadowalające.

4. Ogólna charakterystyka artykulacji samogłosek dla wszystkich badanych dzieci

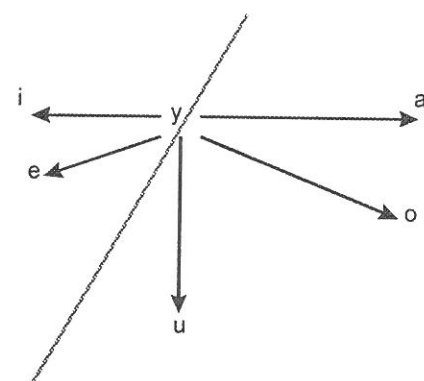
Szczegółowy opis charakterystyki artykulacyjnej samogłosek badanych dzieci niesłyszących zaprezentowano w pracy Łobacz (i in.) [2002]. Przedmiotem niniejszych rozważań jest ustalenie analogii pomiędzy klasyfikacjami samogłosek dla mówców słyszących a obecnie uzyskanymi wynikami. Dla samogłosek języka polskiego wykonano wiele różnych metodologicznie klasyfikacji, począwszy od tradycyjnego opisu artykulacyjnego w postaci trójkąta samogłoskowego [por. np. Wierzchowska 1965], poprzez statystyczne klasyfikacje wykorzystujące podstawowe parametry akustyczne (np. częstotliwości formantowe [por. Jassem (i in.) 2001]) aż do klasyfikacji percepcyjnych, opartych na subiektywnym podobieństwie [por. np. Łobacz 1991]. Niezależnie od analizowanych parametrów każda klasyfikacja musi uwzględnić wiele dodatkowych czynników, wynikających ze specyfiki sygnału mowy. Do najistotniejszych należą uwarunkowania kontekstowe i osob-

nicze. Rolę tych ostatnich omówiono w par. V.2 oraz V.3. Czynniki różnic kontekstowych można oczywiście pominąć albo poprzez analizowanie samogłosek w izolacji, albo poprzez zastosowanie jednorodnego kontekstu (np. bity, byty, bety, baty, boty, buty). Z punktu widzenia procesu komunikacji najciekawsze są jednak te prace, które zróżnicowanie kontekstowe uwzględniają. Tak było właśnie w przypadku obecnych badań.

Poza ustaleniem ogólnej poprawności uzyskanych klasyfikacji szczególnie interesujące są tzw. matryce błędów rozpoznawania – por. tab. 2-4, par. IV.3. Pozwalają one ustalić, jakie elementy są z sobą wzajemnie mylone. W rozpoznawaniu dźwięków mowy opartym na widmowych cechach charakterystycznych najczęściej błędnie rozpoznane są głoski niedostatecznie różne artykulacyjnie. W przypadku mówców niesłyszących analiza błędów jest szczególnie istotna.



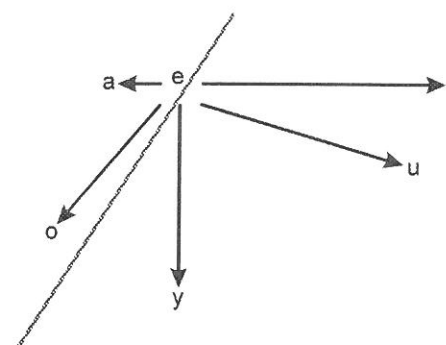
Ryc. 2. Podobieństwo samogłosek wyrażone długością wektorów dla rozpoznania samogłoski /i/



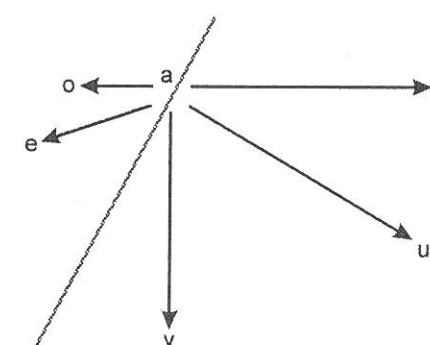
Ryc. 3. Podobieństwo samogłosek wyrażone długością wektorów dla rozpoznania samogłoski /y/

Na ryc. 2-7 przedstawiono w postaci graficznej błędy w klasyfikacjach dla poszczególnych samogłosek. Długość wektora symbolizuje podobieństwo artykulacyjne, ponieważ im krótszy wektor, tym więcej stwierdzono błędów klasyfikacyjnych dla połączonych nim samogłosek. Samogłoski, które charakteryzują się większym podobieństwem, oddzielono na rycinach linią tnącą. Samogłoska przednia – wysoka /i/ (zob. ryc. 2) była najczęściej mylona z samogłoską przednią – średnią /e/, mimo że wartości częstotliwości pierwszych czterech formantów /i/ są bardziej zbliżone do odpowiednich wartości samogłoski /y/ [zob. tab. 7 i 8 w: Łobacz (i in.) 2002]. SNN działała poprawnie, ponieważ w warstwie walidacyjnej najczęściej /i/ było mylone z /y/. W przypadku samogłoski /y/ (przedniej – wysokiej – cofniętej) jej podobieństwo do pozostałych przednich samogłosek jest takie samo (zob. ryc. 3). Cofnięcie artykulacji spowodowało więcej błędów klasyfikacyjnych z samogłoską wysoką – tylną /u/ niż w przypadku /i/.

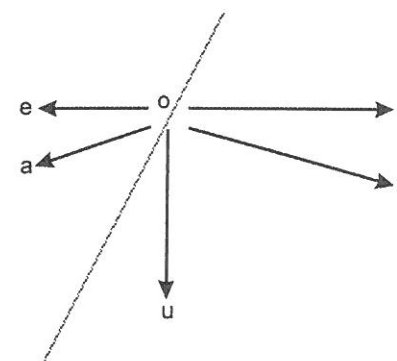
Samogłoska /e/ była najczęściej mylona z samogłoską /a/ (zob. ryc. 4), natomiast samogłoska /a/ z samogłoską /o/ (zob. ryc. 5). Wymowa dzieci niesłyszących charakteryzowała się silną centralizacją, objawiającą się podwyższeniem formantu pierwszego i obniżeniem formantu drugiego, wobec czego wartości podstawowych parametrów akustycznych tych trzech samogłosek uległy znacznemu zbliżeniu.



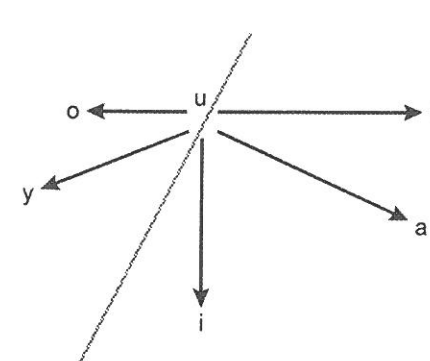
Ryc. 4. Podobieństwo samogłosek wyrażone długością wektorów dla rozpoznania samogłoski /e/



Ryc. 5. Podobieństwo samogłosek wyrażone długością wektorów dla rozpoznania samogłoski /a/



Ryc. 6. Podobieństwo samogłosek wyrażone długością wektorów dla rozpoznania samogłoski /o/



Ryc. 7. Podobieństwo samogłosek wyrażone długością wektorów dla rozpoznania samogłoski /u/

Znaczne podobieństwo artykulacyjne tych trzech samogłosek bardzo dobrze ilustruje graf dla błędnych rozpoznania samogłoski /o/, najtrudniejszej artykulacyjnie dla badanej grupy dzieci. Błędne rozpoznania wskazujące samogłoskę przednią /e/ świadczą o tym, że dla samogłosek średnich opozycja przedniej i tylnej artykulacji nie jest zbyt wyrazista, natomiast w przypadku samogłosek wysokich opozycja ta jest przez dzieci niesłyszące dobrze realizowana. Samogłoska wysoka – tylna /u/

najczęściej była mylona z samogłoską /o/, prawdopodobnie ze względu na wysokie wartości częstotliwości formantu drugiego.

VI. UWAGI KOŃCOWE

Badana grupa dzieci nie potrafiła całkiem poprawnie realizować wszystkich swoich zamierzeń artykulacyjnych w zakresie samogłosek, co najczęściej sprowadzało się do bardziej centralnego ułożenia masy języka. Znacznie łatwiejsze okazały się samogłoski wysokie /i, u/ o wyraźnie skrajnej przedniej bądź tylnej artykulacji.

Wyniki obecnej klasyfikacji przy użyciu sieci neuronowej są zgodne z rezultatami badań Kurkowskiego [1996], który oceniał poprawność realizacji samogłosek w wypowiedziach niesłyszących sześciolletnich dzieci innymi metodami. Samogłoska /u/, dla której uzyskano w obecnym eksperymencie najlepszą poprawność klasyfikacji, tj. 72,77% w pierwszej części eksperymentu i 82,16% w drugiej, jest, według Kurkowskiego [1996], najlepiej realizowana przez niesłyszących. Autor odnotował 84,7% jej prawidłowych realizacji i zauważył, że jeśli jest ona zastępowana, to najczęściej przez [o]. Samogłoska [u] ma najbardziej wyrazistą wzrokowo artykulację ze względu na wyraźne zaokrąglenie i wysunięcie warg do przodu i to prawdopodobnie ułatwia jej nabywanie przez niesłyszących.

Gorsze wyniki klasyfikacji widm samogłosek średnich i niskiej [a] można przypisać także zaburzeniu ich struktury formantowej przez proces silnej nazalizacji, będący efektem niedostatecznej kontroli podniebienia miękkiego przez niesłyszących [Hudgins 1934; Chen 1995]. Zaobserwowano, że samogłoski [a] i [e] poddają się nazalizacji najczęściej, a samogłoski [i] i [u] najrzadziej, ponieważ niska pozycja języka umożliwia, a wysoka przeszkadza obniżeniu się podniebienia miękkiego i otwarciu jamy nosowej [por. Łobacz (i in.) 2002].

Na podstawie przedstawionej klasyfikacji można sporządzić hierarchię trudności artykulacyjnych w prawidłowym nabywaniu polskich samogłosek ustnych przez dzieci niesłyszące. Najmniej wyrazista dla dzieci jest opozycja średniej i niskiej artykulacji, przy czym dla samogłosek średnich trudna jest także opozycja: przedniość/tylność. Najłatwiejsze jest wysokie położenie języka, i to zarówno jego wysunięcie do przodu jak i wysklepienie w kierunku podniebienia miękkiego.

Ogólny wynik 70% poprawnych rozpoznania przez SNN jest niższy o ok. 20% od rozpoznawania przez tego typu sieć samogłosek dorosłych mówców słyszących. Dla tej samej badanej grupy dzieci uzyskano także, za pomocą innych narzędzi statystycznych, wynik klasyfikacji o ok. 20% niższy dla spółgłosek trących i afrykat [Kleśta 2002]. Należy jednakże zdawać sobie sprawę z faktu, że pomiary parametrów akustycznych były wykonywane dla mówców o niskiej ogólnej zrozumiałości. Kolejnym krokiem referowanych obecnie badań jest zastosowanie procedur klasyfikacyjnych dla mówców tworzących bardziej jednorodną grupę pod

względem poziomu rehabilitacji i metod kształcenia (obecnie prowadzone są badania w ośrodku szkolno-wychowawczym dzieci słabosłyszących i niesłyszących w Kaliszu, w którym dzieci rehabilitowane są metodą fonogestową).

Zastosowana procedura klasyfikacyjna wykazała, że możliwe jest opracowanie programów komputerowych w postaci np. interaktywnych podręczników i gier dla osób niesłyszących posługujących się mową niezależnie od konstruowania programów rozpoznających elementy języka migowego. Umożliwienie aktywnego uczestnictwa osobom niesłyszącym w dialogu człowiek-komputer za pomocą mowy pomoże im w wielu sytuacjach nowoczesnego zdobywania wiedzy.

Bibliografia

- Chen M. Y. [1995]. Acoustic parameters of nasalized vowels in hearing-impaired and normal-hearing speakers. „Journal of the Acoustic Society of America” 95, (5), 2443-2453.
- Hudgins C. V. [1934]. A comparative study of speech coordination of deaf and normal subjects. „Journal of the Genetic Psychology” 64, 3-48.
- Izworski W., Wszolek W. [1999]. Wykorzystanie metod sztucznej inteligencji w diagnostyce i przetwarzaniu patologicznych sygnałów mowy. W: W. Jassem, Cz. Basztura, G. Demenko, K. Jassem (red.). Technologia mowy i języka. T. 3. Poznań: Polskie Towarzystwo Fonetyczne s. 299-319.
- Jassem W., Grygiel W. [w druku]. Classification of Polish Vowel Spectra Using Artificial Neural Network. „Journal of International Phonetic Association” 33 (2).
- Jassem W., Krzyśko M., Wołyński W. [2001]. Reduction of speaker-related variability in Polish vowel spectra. „Journal of the International Phonetic Association” 31 (2), 187-202.
- Kleśta J. [2002]. Charakterystyka akustyczna cech segmentalnych mowy dzieci niesłyszących. Nie opublikowana rozprawa doktorska. Poznań: UAM.
- Kurkowski Z. M. [1996]. Mowa sześciolletnich dzieci z uszkodzonym narządem słuchu. Lublin: Wydawnictwo UMCS.
- Łobacz P. [1991]. Psychometric classification of phonemes – some methodological considerations. „Lingua Posnaniensis” 32-33, 147-168.
- Łobacz P. [2002]. Wstępna charakterystyka fonetyczna mowy wybranej grupy dzieci niesłyszących. W: S. Puppel (red.). Scripta Neophilologica Posnaniensia 4. Poznań: Dziekan Wydziału Neofilologii i Motivex s. 29-50.
- Łobacz P., Francuzik K., Szalkowska E. [2002]. Acoustic-phonetic description of hearing-impaired children's vowels in Polish. „Psychology of Language and Communication” 6 (2), 3-31.
- Monsen R. B. [1974]. Durational aspects of vowel production in the speech of deaf children. „Journal of Speech and Hearing Research” 17, 386-398.
- Ryalls J. [1996]. A basic introduction to speech perception. San Diego-London: Singular Publishing Group, Inc.
- Wierzchowska B. [1965]. Wymowa polska. Warszawa: PZWS.
- Zgólkowa H., Bułczyńska K. [1987]. Słownictwo dzieci w wieku przedszkolnym. Poznań: Wydawnictwa UAM.

Adres do korespondencji:

Piotra Łobacz
Zakład Psycholingwistyki
Uniwersytet im. A. Mickiewicza
ul. Międzychodzka 5
60-371 Poznań