AKADEMIA GÓRNICZO-HUTNICZA

im. Stanisława Staszica w Krakowie

Wydział Elektrotechniki, Automatyki, Informatyki i Elektroniki

mgr inż. Marcin Kapusta

Wykorzystanie sieci Kohonena do wizualizacji mowy patologicznej

rozprawa doktorska

promotor pracy

prof. zw. dr hab. inż. **RYSZARD TADEUSIEWICZ**

Kraków 2000

Chciałbym serdecznie podziękować za okazaną pomoc JM Rektorowi AGH prof. dr hab. inż. Ryszardowi Tadeusiewiczowi. Podczas powstawania niniejszej pracy zawsze służył dobrą radą, a Jego cenne wskazówki i uwagi były bardzo pomocne we właściwym ukierunkowaniu prowadzonych przeze mnie badań. Pragnę również podziękować dr inż. Wiesławowi Wszołkowi za udostępnienie mi materiału badawczego.

Marcin Kapusta

Spis treści

1. Wstęp

1.1. Cel pracy

Sygnał akustyczny, a w szczególności sygnał mowy, jest dynamicznym wielowymiarowym[1] zjawiskiem, trudnym do czytelnego przedstawienia w postaci graficznej. Tymczasem w dość licznych zastosowaniach zachodzi potrzeba analizy i **jakościowej** oceny tego sygnału przez człowieka – zwykle w celu oceny prawidłowości budowy i działania systemu generującego ten sygnał. Typowym obszarem, w którym potrzeba taka się pojawia jest obszar diagnostyki technicznej i medycznej. Zwłaszcza w odniesieniu do zagadnień biomedycznej natury możliwość wygodnej jakościowej oceny sygnału przez człowieka (lekarza) ma bardzo duże znaczenie. Celem niniejszej pracy jest rozwiązanie tego problemu poprzez znalezienie takiej projekcji sygnału mowy, która pozwoli unaocznić różnice między mową prawidłową i patologiczną. Taka projekcja pozwalająca wzrokowo oceniać (i dokumentować na papierze lub w komputerowej bazie danych o pacjentach, tworzonej dla celów porównawczych) stopień deformacji sygnału mowy patologicznej (np. u osób po zabiegach operacyjnych w obszarze krtani lub tzw. nasady – jamy ustnej, nosowej, podniebienia, szczęki, łuku zębowego itp.) ma bardzo duże znaczenie praktyczne na gruncie medycyny. Warto dodać, że zagadnienie to

[1] Pozornie można sądzić, że sygnał dźwiękowy jest sygnałem **jednowymiarowym**, gdyż daje się zarejestrować jako funkcja jednej zmiennej (na przykład przebieg zmiennego w czasie ciśnienia akustycznego). Sygnał ten można oczywiście przedstawić także w jednowymiarowej formie cyfrowej, na przykład w postaci serii wartości chwilowych napięcia próbkowanego na wyjściu mikrofonu. Jednak przytoczona wyżej uwaga o konieczności traktowania złożonych sygnałów akustycznych jako sygnałów wielowymiarowych jest też prawdziwa, jako że większość ocen i analiz odnoszących się do tego sygnału nie może być prowadzona w oparciu o cechy ich przebiegów czasowych, tworzących z reguły zupełnie nieczytelne zapisy, lecz musi odwoływać się do struktury czasowo-częstotliwościowej, która jest z natury dwu lub trójwymiarowa (w zależności od tego, czy uwzględnia się fazę sygnału, czy też się ją ignoruje), zaś w odniesieniu do sygnału mowy stosuje się nagminnie opisy odwołujące się do złożonych parametrów artykulacyjnych, takich jak formanty, momenty widmowe, współczynniki LPC czy struktury cepstrum sygnału, co w skrajnych przypadkach może zmuszać do rozważania sygnału mowy jako trajektorii układającej się w przestrzeni nawet kilkunastowymiarowej.

nie zostało jeszcze rozwiązane w dostępnej literaturze światowej, prezentowana praca wkracza więc na obszar, w którym brakuje punktów odniesienia i użytecznych wzorów.

Opracowanie takiego systemu automatycznej analizy, przetwarzania oraz (co jest głównym przedmiotem tej pracy) – wizualizacji mowy patologicznej jest bardzo ważnym i bardzo potrzebnym zagadnieniem naukowym i praktycznym. Człowiek (np. lekarz) potrafi bowiem głównie różnicować oceny pacjentów w zakresie wykrycia podstawowej różnicy między mową poprawną i patologiczną. Natomiast w przypadku, kiedy zachodzi potrzeba pogłębionej analizy porównawczej, ze wskazaniem stopnia występującej patologii albo z koniecznością śledzenia zmian (np. wynikających z postępu procesu leczenia lub rehabilitacji) – słuch człowieka okazuje się bezradny. Wynika to z faktu, że praktyka percepcji słuchowej formuje u ludzi liczne wzorce poprawnych wypowiedzi, natomiast brakuje analogicznych wzorców w zakresie wypowiedzi patologicznych, z którymi ludzie (nawet lekarze) spotykają się rzadziej, a które w dodatku mają bardzo nieregularny charakter (każdy przypadek mowy patologicznej jest w gruncie rzeczy unikatowy i niepodobny do innych przypadków, nawet przy tym samy rodzaju patologii). Jedną z przyczyn tego stanu rzeczy jest fakt, że obok uwarunkowań akustycznych, związanych ze zjawiskiem naturalnej lub jatrogennej deformacji narządów mowy, na strukturę fonetyczną mowy patologicznej bardzo duży wpływ mają także zmienne w czasie, subiektywne usiłowania pacjenta. Z obserwacji zachowań pacjentów cierpiących na przypadłość powodującą (między innymi) deformację sygnału mowy wiadomo, że każdy z nich zmagając się ze swoim kalectwem, stara się tak modyfikować i korygować proces artykułowania wydawanych dźwięków, żeby osiągnąć optymalny (jego zdaniem) efekt końcowy. Ten czynnik intencjonalnego (a często także emocjonalnego) korygowania dźwięków mowy patologicznej przez czynniki związane z unikatową osobowością każdego indywidualnego pacjenta powoduje, że mowa patologiczna każdego człowieka jest w istocie inna, niepowtarzalna, niemożliwa do „wytrenowania" w umyśle diagnosty czy terapeuty.

Tymczasem wiadomo, że skuteczność naszej percepcji (w każdej dziedzinie) jest zależna od tego, na ile skutecznie uformujemy w swoim umyśle wzorce kognitywne, współdziałające w procesie rozpoznawania i wartościowania sygnałów ze strumieniem wrażeń percepcyjnych, pochodzących od zmysłów. Obecność w umyśle lekarza takich wzorców kognitywnych dla (dobrze mu znanej) mowy prawidłowej oraz brak takich wzorców dla różnych – zwykle unikatowo odmiennych u każdego pacjenta – form mowy patologicznej powoduje, że jedyną różnicą, którą lekarz potrafi stosunkowo łatwo wykryć słuchem – jest (trywialna skądinąd) różnica między sygnałem mowy prawidłowej i patologią. Natomiast wszystkie próbki sygnału mowy patologicznej są tak

„egzotycznym" doznaniem percepcyjnym, że są one wszystkie kwalifikowane jednorodnie jako „nieprawidłowe" – jednak bez możliwości wartościowania czy różnicowania **stopnia** tej nieprawidłowości.

Tymczasem w wielu przypadkach (np. wybór właściwej formy terapii, selekcja zakresu i sposobu przeprowadzenia zabiegu operacyjnego, dobór optymalnej protezy, ocena postępu procesu rehabilitacji itp.) ocena stopnia deformacji mowy (chociażby na poziomie elementarnych ocen typu „lepiej"– „gorzej") jest czynnikiem o zasadniczym znaczeniu. Prezentowana praca ma na celu stworzenie informatycznego narzędzia, opartego na wykorzystaniu techniki sieci neuronowych Kohonena, pozwalającego na dokonywanie takich właśnie potrzebnych wartościowań w zakresie oceny stopnia deformacji sygnału mowy patologicznej. Podstawowym założeniem, na jakim oparto tę pracę, jest koncepcja wykorzystania do oceny form i stopnia deformacji mowy patologicznej ludzkiego analizatora wzrokowego, zamiast ograniczonego (z powodów wyżej wymienionych) analizatora słuchowego. Przy formułowaniu tego założenia brano pod uwagę następujące czynniki:

- system wzrokowy człowieka cechuje się znacząco większą pojemnością informacyjną, co powoduje, że można przy jego pomocy wykryć i ocenić bardziej subtelne różnice sygnału, niż te, które są wykrywane przy pomocy słuchu;
- nawyki i przyzwyczajenia kognitywne, ukierunkowujące percepcję słuchową na rozpoznawanie mowy prawidłowej, utrudniające prawidłową ocenę stopnia deformacji mowy patologicznej, nie stanowią żadnej przeszkody w przypadku prezentacji sygnału mowy w formie obrazowej, która dla każdej (zarówno prawidłowej, jak i patologicznej formy sygnału) jest równie egzotyczna i wymagająca stosowanego treningu w celu uzyskania możliwości poprawnej interpretacji;
- przy percepcji wizualnej człowiek potrafi przyjąć i wykorzystać równocześnie bardzo dużą ilość danych, prawidłowo dokonując ich jakościowej interpretacji zarówno w zakresie oceny wartości (i znaczenia) poszczególnych danych indywidualnych, jak i w zakresie poprawnej interpretacji współzależności i trendów dających się wykryć w rozważanych danych.

Dla realizacji sformułowanego wyżej założenia należało opracować metodę przedstawiania złożonego procesu akustycznego, jakim jest sygnał mowy, w postaci odpowiedniego obrazu, nadającego się do wzrokowej interpretacji przez człowieka. Taka projekcja sygnału akustycznego do formy obrazowej nazywana jest zwykle procesem **wizualizacji mowy** – w związku z tym taki termin będzie używany w całej dalszej treści pracy. Analizując znane z literatury obecnie dostępne techniki wizualizacji mowy

(zarówno prawidłowej, jak i patologicznej) opierające się na prezentacji sygnału mowy w postaci czasowej, częstotliwościowej lub czasowo-częstotliwościowej a także parametrycznej, stwierdzono, że są one zbyt skomplikowane i zawierają zbyt wiele szczegółów, żeby mogły być użyteczne z punktu widzenia lekarza oceniającego np. wynik przeprowadzonej operacji. **W prezentowanej pracy skupiono więc uwagę na możliwościach, jakie w zakresie wizualizacji mowy niesie technika samouczących się sieci neuronowych typu Kohonena.** Sieci te pozwalają w toku procesu samouczenia znaleźć odwzorowanie (projekcję) wielowymiarowej przestrzeni sygnałów wejściowych do dwuwymiarowej przestrzeni wyjść kształtowanej przez warstwę konkurujących ze sobą neuronów.

1.2. Treść pracy na tle aktualnych osiągnięć w dziedzinie przetwarzania sygnału mowy

1.2.1. Ogólne problemy automatycznego rozpoznawania mowy

Język naturalny jest najpopularniejszym sposobem porozumiewania się ludzi między sobą. Stanowi on bardzo uniwersalny i efektywny środek komunikacji, pozwalający na przekazywanie różnorodnych informacji. Komunikacja głosowa ma tę zaletę, że nadawana wiadomość tworzona jest w sposób łatwy i naturalny do sformułowania przez nadawcę informacji, a przekazywany sygnał jest bezpośrednio zrozumiały dla odbiorców. Nawet małe dziecko nie ma kłopotów ze zrozumieniem tego, co się do niego mówi więc mogłoby się wydawać, że automatyczna analiza i rozpoznawanie wypowiedzi jest zadaniem prostym. W rzeczywistości jednak tak nie jest – sygnał mowy jest jednym z najbardziej skomplikowanych i złożonych sygnałów akustycznych. Jego automatyczna analiza i rozpoznawanie są procesem bardzo trudnym, a ich badanie i rozwijanie budzi przy okazji szacunek i uznanie dla ludzkich kompetencji językowych, dostępnych niemal każdemu człowiekowi, lecz w istocie opartych na niesłychanie złożonych procesach kognitywnych, zachodzących w ludzkim mózgu – tyle tylko, że nie penetrowanych introspektywnie przez człowieka, ponieważ zdecydowana ich większość zachodzi poniżej poziomu świadomości. Mimo tych trudności obecnie prawie na całym świecie trwają prace związane z przetwarzaniem i analizą sygnałów akustycznych, a w szczególności sygnału mowy, gdyż rozwiązanie problemów jego automatycznego roz-

poznawania otworzy bardzo cenne możliwości naturalnej (głosowej) komunikacji człowieka z różnymi maszynami. Zwłaszcza w czasie ostatniej dekady, kiedy to nastąpił gwałtowny rozwój technologiczny, umożliwiający wykorzystanie złożonych algorytmów także w sprzęcie elektronicznym pełniącym role pomocnicze (komunikacyjne), zaczęło się pojawiać coraz więcej prac dotyczących analizy sygnału mowy.

W chwili obecnej liczba publikacji związanych z omawianą problematyką sięga kilku tysięcy. Pierwsze prace dotyczące analizy i przetwarzania sygnału mowy związane były z poszukiwaniem metod zmniejszenia redundancji tego sygnału, która sięga około 90% [Bas88, Tad78] i jest szczególnie uciążliwa podczas jego transmisji, np. poprzez łącza telefoniczne [Sap66]. Obecnie proste problemy związane z kompresją sygnału mowy zostały już w zasadzie rozwiązane (stworzono nawet odpowiednie standardy oraz układy elektroniczne dokonujące on-line kompresji i dekompresji sygnału zgodnie z tymi standardami), więc we współczesnych pracach badawczych główny nacisk kładziony jest na stworzenie systemów automatycznego rozpoznawania mowy.

1.2.2. Rozpoznawanie treści wypowiedzi

Jednym z najbardziej popularnych i jednocześnie najprostszych zadań z zakresu techniki rozpoznawania mowy jest rozpoznawanie cyfr. Zadanie to jest ważne z praktycznego punktu widzenia – na przykład przy głosowym zestawianiu połączeń w systemach telekomutacji, a także w obszarze niektórych nowoczesnych usług– na przykład bankofonów. Nic więc dziwnego, że dla rozwiązania tego zadania użyto wielu nowoczesnych metod, związanych z wektorową analizą sygnałów, co zaowocowało tym, że uzyskano bardzo dobre rezultaty praktyczne – na przykład w przypadku języka angielskiego przy rozpoznawaniu, w sposób niezależny od mówcy, ciągów cyfr o znanej długości, wypowiadanych bez przerw, uzyskano współczynnik błędu zaledwie0,3 %.

Jeżeli chodzi o systemy bardziej złożone, to można wymienić w tym miejscu system rozpoznający około 1000 słów zwany *Resource Management* (RM), w którym wypowiedzi dotyczą zagadnień związanych z nawigacją statków na Pacyfiku; uzyskano w nim poziom błędu mniejszy od 4 %. Wśród systemów rozpoznających mowę spontaniczną warto wymienić system ATIS (*Air Travel Information Service*). Posługuje się on słownikiem złożonym z 2000 słów i jego współczynnik błędu wynosi mniej niż 3 %. Od 1992 roku tworzone są również bardzo skomplikowane systemy, korzystające z dużych słowników (ponad 20000 słów), niezależne od mówcy, rozpoznające mowę ciągłą. Najlepszy z tych systemów w 1994 roku uzyskał współczynnik błędu 7,3 % przy przetwarzaniu zdań z *North America Business News* [Col96, Rab93]. Warto również wspo-

mnieć o systemie ISADORA. Jest to system rozpoznawania mowy ciągłej w języku polskim [Fab99], jedno z pierwszych tego rodzaju osiągnięć opublikowanych dla naszego języka.

Wraz z rozwojem zastosowań następuje także stosowny rozwój metod rozpoznawania mowy. Ze względu na ogromną pomysłowość różnych badaczy trudno tu wyczerpująco omówić wszystkie stosowane techniki i metody, podane zostaną więc tylko wybrane uwagi na temat niektórych najczęściej stosowanych metod W systemach rozpoznawania mowy ciągłej przeważnie stosuje się ukryte modele Markowa HMM (*Hidden Markov Models*) [Gaj99a, Rab89] oraz sieci neuronowe [Mor92, Str94]. Jako przykład może posłużyć hybrydowy system łączący ukryte modele Markowa i sieci neuronowe [Rii97] oraz system rozpoznawania izolowanych słów (wykorzystujący sieci RBF oraz HMM) przedstawiony w pracy [Sin92].

Algorytmy rozpoznawania mowy są również stosowane we wszelkiego rodzaju aplikacjach wspomagających proces dyktowania tekstów (tzw. systemy *speech to text*). Obecnie istnieje kilka takich aplikacji wykorzystujących bardzo duże słowniki, jednakże większość z nich wymaga istnienia wyraźnych odstępów pomiędzy poszczególnymi wyrazami. Wydajność tych aplikacji można zwiększyć ograniczając dziedzinę zastosowania (np.: zbudowano bardzo wydajny system służący do dyktowania raportów medycznych [Kup96]). Wśród istniejących aplikacji tego typu warto wymienić programy: *Dragon Dictate*, *IBM VoiceType*, *Kurzweil Voice* [Kap98].

Z problemem automatycznego rozpoznawania mowy, związany jest także problem rozpoznawania mniejszych fragmentów wypowiedzi takich jak: pojedyncze wyrazy, głoski, czy wreszcie fonemy. Bardzo często stosowane są do tego celu sieci neuronowe. Jeżeli chodzi o zastosowanie sieci neuronowych do rozpoznawania mowy to należy wymienić takie prace jak: [Bri89, Bri90, LeC94, Ren92, Wai89a, Wai89b]. Przykładowo można przytoczyć, że w znanej pracy Le Crefa sieć MLP stosowana była jako kwantyzer wektorowy [LeC94], a jej zadaniem było rozpoznawanie wybranych grup fonemów języka flamandzkiego. W pracach [Wai89a, Ham90] omówiono sieci przeznaczone do rozpoznawania spółgłosek języka angielskiego, natomiast w pracy [Wai89b] Waibel przedstawia 7 sieci przeznaczonych do rozpoznawania różnych grup klas fonemów. W eksperymencie opisanym przez Robinsona [Rob94], rekurencyjna sieć neuronowa zastosowana została do rozpoznawania fonemów w mowie ciągłej wielu mówców, podobnie opisane w pracy [Han94] sieci rekurencyjne Elmana zostały zastosowane do rozpoznawania wybranych fonemów w układzie spółgłoska-samogłoska-spółgłoska. W pracy [Mor90] Morgan opisuje sieć MLP rozpoznającą fonemy w mowie ciągłej dla wielu mówców. Należy także wspomnieć o pracy przedsta-

wiającej grupę sieci neuronowych zastosowanych do rozpoznawania fonemów języka polskiego [Hał99].

Jak już wcześniej wspomniano, sieci neuronowe stanowią również grupę często stosowanych narzędzi rozpoznawania współpracujących z ukrytymi modelami Markowa [Ben92, LeC94, Pal92, Rig94, Zav94]. Przykładem może być rozpoznawanie fonemów w sieci rekurencyjnej współpracującej z modelami Markowa, opisane w pracy [Rob94].

1.2.3. Automatyczne rozpoznawanie mówcy

Oddzielną dziedziną związaną z przetwarzaniem sygnału mowy jest automatyczne rozpoznawanie głosów, należące do grupy biometrycznych metod identyfikacji osób. Technika automatycznego rozpoznawania mówcy pozwala na zweryfikowanie jego tożsamości na przykład w celu kontroli dostępu do różnego rodzaju usług, takich jak: głosowe karty telefoniczne, obsługa konta bankowego przez telefon, zakupy telefoniczne, dostęp do różnego rodzaju systemów informacyjnych, poczta głosowa, ochrona dostępu do poufnych baz danych, zdalny dostęp do komputerów oraz sterowanie urządzeń głosem przez osoby do tego uprawnione. Obecnie istnieje już wiele komercyjnych systemów automatycznego rozpoznawania mówcy. Wśród popularnych producentów takich systemów można wymienić firmy: *ITT, Lernout & Hauspie, T-NETIX, Veritel* oraz *Voice Control Systems*. W szczególności warto tu zwrócić uwagę na firmę *Sprint*, produkującą popularną kartę telefoniczną *Sprint's Voice FONCARD®* używaną w Stanach Zjednoczonych.

Zasadniczo istnieją dwa podejścia do rozpoznawania głosu [Cam97, Fur96]: rozpoznawanie ustalonej wypowiedzi pochodzącej z ograniczonego zbioru wypowiedzi (*Text Dependent Speaker Recognition*) bądź rozpoznawanie wypowiedzi swobodnej czyli dowolnej (*Text Independent Speaker Recognition*). Pewne połączenie tych dwóch podejść stanowi metoda rozpoznawania wypowiedzi wybieranej przez system w sposób losowy (z ustalonego zbioru wypowiedzi) i „podawanej" użytkownikowi wizualnie bądź drogą dźwiękową (*Text–Prompted Speaker Recognition*). Ma to zapobiec ewentualnym próbom oszukania systemu rozpoznawania poprzez odtworzenie zdobytego wcześniej nagrania pochodzącego od autoryzowanego użytkownika. Przykład takiego rozwiązania opisany jest np. w pracach [Hig91, Mat93].

1.2.4. Parametry na których opiera się proces rozpoznawania mówcy

Można stwierdzić, że w zadaniach związanych z rozpoznawaniem mówców najczęściej wykorzystywane są parametry cepstralne oraz parametry kodowania predykcyjnego. W przypadku rozpoznawania zależnego od tekstu stosuje się zazwyczaj jedno z dwóch podejść: technikę dopasowania opartą na programowaniu dynamicznym DTW (*Dynamic Time Warping*) [Fur81, Fur96] lub ukryte modele Markowa HMM [Che95, Mat95, Rey95, Tis91]. Natomiast metody rozpoznawania niezależnego od tekstu bazują na jednym z następujących podejść [Fur96]: długookresowych modelach statystycznych, technice kwantyzacji wektorowej, ergodycznych modelach Markowa, selekcji zdarzeń specyficznych, selekcji tych samych segmentów quasi-stacjonarnych (np. głoska „*a*"), sieciach neuronowych. Szerszy przegląd metod i problematyki związanej z automatycznym rozpoznawaniem mówcy można znaleźć np. w pracach: [Ata97, Bas93, Cam97, Dod97, Fur89, Fur96, Ros91, Ros97, Sho99].

Na podobnej zasadzie, aczkolwiek na nieco skromniejszą skalę, prowadzone są od lat badania w Polsce. W pracach autorów polskich koncentrowano się najczęściej na podstawowej problematyce, jaką jest dobór parametrów dyskryminujących głosy. Do takich prac należą m. in. wcześniejsze prace Gubrynowicza, Jassema, późniejsze publikacje m.in. Majewskiego, Basztury i innych [Bas79, Bas91, Gub81, Jas79, Maj79, Tad88a]. Ostatnie prace ujmowały szerzej problematykę automatycznego rozpoznawania mówcy, włączając w to badania wpływu algorytmu rozpoznawania [Bas87], funkcji podobieństwa [Bas91] oraz rzeczywistych warunków transmisji sygnału mowy [Bas81].

1.2.5. Specjalistyczne systemy rozpoznawania mowy

Prowadzone są również liczne badania nad możliwością zastosowania metod analizy i rozpoznawania sygnału mowy w bardziej specjalistycznych dziedzinach nauki i życia: w medycynie (możliwości protezowania i rehabilitacji funkcji narządu artykulacji mowy [Bie64, Com87, Ksi87a, Ksi87b, Wsz94]), sądownictwie i kryminalistyce (ekspertyzy, identyfikacja głosów [Bas95, Bas99]), sterowaniu procesami [Bas93, Tad75], itp. [Jas73, Bra99]. Kolejna grupa zagadnień (prezentowanych w wielu pracach) dotyczących wykorzystania analizy sygnałów dźwiękowych, to zagadnienia diagnostyki (w szerokim tego słowa znaczeniu) medycznej i technicznej. Próbę zastosowania metod akustycznych w diagnostyce medycznej narządu mowy przedstawiono m. in. w pracach [Bas96, Eng92, Eng94, Eng95, Gub80, Gub81, Izw99, Kac76, Kac79, Ksi87a, Ksi87b,

Now85, Tad83, Tad85a, Tad85b, Tad88a, Tad88b, Tad94. Tad97, Tad98a, Wsz91]. Próbę zastosowania metod automatycznego rozpoznawania obrazów dźwiękowych w systemach technicznych przedstawiono m. in. w pracach [Bas87, Bas88, Bas93, Bas96, Eng93, Pec85, Tad75, Tad91].

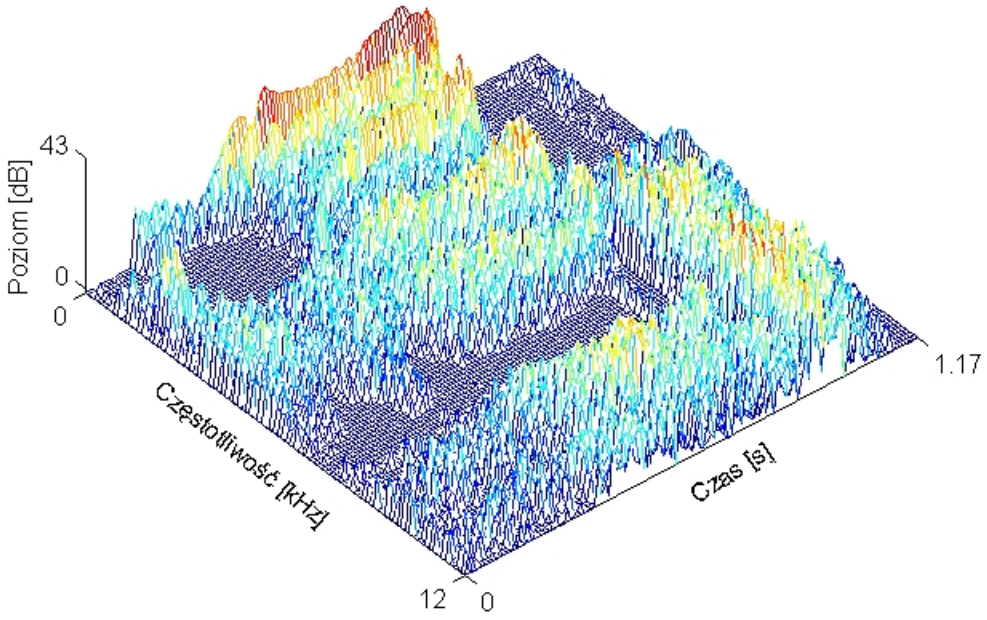
Obecnie badania nad automatycznym przetwarzaniem sygnału mowy są najbardziej zaawansowane w USA, Japonii oraz krajach Europy Zachodniej. Ocena wyników tych badań jest niezmiernie trudna z wielu powodów – dotyczą one bowiem różnych języków, prowadzone są przy użyciu odmiennych metod i przy wykorzystaniu różnych środków technicznych. Może warto wspomnieć na marginesie, że problem porównywania uzyskanych wyników oraz różnych podejść do zagadnienia rozpoznawania zarówno mowy jak i mówców jest na tyle istotny, iż zdecydowano się na tworzenie specjalistycznych baz danych przystosowanych do konkretnych założeń poszczególnych podejść do zadania rozpoznawania. Celem budowy tych baz jest przede wszystkim dostarczenie badaczom jednakowych materiałów doświadczalnych przystosowanych do obiektywnej oceny konkretnego podejścia i w pewnym stopniu narzucających warunki przeprowadzenia eksperymentów. Do czołowych producentów baz danych głosów należy koncern danych lingwistycznych Uniwersytetu Pensylwanii (*Linguistic Data Consortium* LDC – University of Pennsylvania [http://www.ldc.upenn.edu]), innymi przykładami takich baz mogą bazy: TIMIT [Lamel87] i DARPA [Price88].

1.2.6. Podsumowanie

Na podstawie badań literaturowych, których fragmenty przedstawiono, można stwierdzić, że wprawdzie istnieje najogólniej mówiąc bogactwo różnego rodzaju opracowań i metod pokrywających szeroki wachlarz problemów i potrzeb, to jednak pewne problemy nie zostały rozwiązane, między innymi problem wizualizacji sygnału mowy. Do chwili obecnej w zasadzie jedynym sposobem wizualizacji sygnału akustycznego, opisanego w dziedzinie częstotliwości i czasu (jest to jedna z najpopularniejszych form „nieparametrycznego" opisu sygnału), było przedstawienie takiego sygnału za pomocą wideogramu (zwanego też spektrogramem dynamicznym albo sonogramem). Wideogram jest wykresem trójwymiarowym, w którym jedna oś reprezentuje czas, druga częstotliwość, natomiast moduł widma sygnału reprezentuje wysokość wykresu.

Na rysunku 1-1 przedstawiono wideogram sygnału mowy, reprezentujący wzorcową wypowiedź słowa „*pies*". Przedstawienie takiego wykresu na płaszczyźnie dwuwymiarowej jest trudne i nie zawsze czytelne (w większości przypadków uniemożli-

wiające analizę sygnału). Ponadto sama struktura wideogramu nie jest zbyt przejrzysta dla człowieka widzącego po raz pierwszy taki wykres.



Rys. 1-1. Wideogram wzorcowej wypowiedzi słowa „*pies*".

Jak już wzmiankowano wyżej, zadaniem niniejszej pracy jest stworzenie takiej metody wizualizacji sygnału mowy (w szczególności mowy patologicznej), która w sposób przejrzysty i nie wymagający złożonej analizy pozwoli na odróżnienie mowy patologicznej od mowy uznawanej za wzorcową, a także przyczyni się do możliwości różnicowania różnych poziomów patologii mowy. Poszukiwana będzie metoda taka, dzięki której można by było pewne procesy akustyczne oceniać nie tylko w kategoriach parametrycznych (tak jak do tej pory), lecz także wizualnych – na zasadzie podobieństwa obrazów. Różnica podejścia, które jest proponowane w tej pracy, w stosunku do klasycznych metod analizy sygnałów dźwiękowych, polega głównie na tym, że sygnał akustyczny nie jest tu przetwarzany w pewien zbiór parametrów opisujących ten sygnał, lecz na obraz graficzny, który to obraz może podlegać dalszej analizie lub bezpośredniej ocenie specjalisty.

Najbardziej wyrafinowanym narzędziem do przetwarzania sygnałów, w szczególności sygnałów mowy, w celu ich klasyfikacji i rozpoznawania, jest system słuchowy człowieka. Chociaż badania systemu słuchowego człowieka nie są wciąż zakończone, gdyż niemal każdy miesiąc przynosi publikacje informujące o nowych odkryciach fizjologii słyszenia oraz psychoakustyki, wiele problemów automatycznego rozpoznawa-

nia mowy zostało rozwiązanych właśnie po zapoznaniu się z naturalnym systemem słuchowym, traktowanym jako najdoskonalszy wzorzec systemu analizy i rozpoznawania mowy. Zatem również w trakcie tworzenia omawianego w tej pracy systemu wizualizacji mowy należy spróbować wykorzystać mechanizmy jakimi posługują się ludzie w trakcie rozpoznawania mowy i klasyfikacji różnego rodzaju dźwięków. Dlatego w następnym podrozdziale w skrócie zostanie zaprezentowana budowa i zasady funkcjonowania systemu słuchowego człowieka.

1.3. Podstawy budowy i funkcjonowania systemu słuchowego człowieka

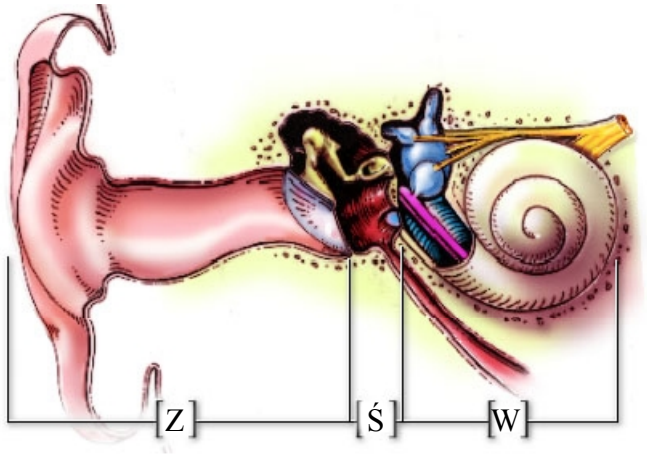
1.3.1. Związek budowy i funkcji niższych pięter systemu słuchowego

System słuchowy człowieka można wstępnie podzielić na dwie części: część mechaniczną oraz część nerwową [Tad88a, Moo99]. Punktem styku pomiędzy tymi dwoma częściami jest ślimak, a dokładniej mówiąc, znajdujący się w nim narząd Cortiego, przekazujący informacje o dźwięku do neuronów zwoju spiralnego. Inne podziały wskazują na ucho zewnętrzne, ucho środkowe i wewnętrzne (rys. 1-2). Ucho środkowe działa jak element sprzęgający, którego zadaniem jest dopasowanie impedancji akustycznych, mające z kolei na celu poprawę sprawności przepływu energii akustycznej pomiędzy powietrzem a płynami wypełniającymi ślimak.

Analiza dźwięku dokonywana jest w uchu wewnętrznym z wykorzystaniem złożonego systemu mechanicznego i neuronalnego. Wzdłuż ślimaka biegnie membrana, nazywana błoną podstawną, dzieląca ślimak na dwie części. Odbierane przez ucho dźwięki wywołują fale mechaniczne biegnące wzdłuż błony podstawnej. Dla każdej częstotliwości wejściowego dźwięku maksimum przebiegu wychylenia błony podstawnej występuje w innym jej miejscu. Małe częstotliwości powodują powstanie maksimum wychylenia błony w pobliżu wierzchołka ślimaka (w pobliżu przerwy wierzchołkowej zwanej *helicotrema*), a dla dużych częstotliwości maksimum drgań błony obserwuje się przy podstawie ślimaka. Zatem błona podstawna działa jak niedoskonały analizator częstotliwości lub jak układ filtrów pasmowych, rozkładających dźwięki złożone na ich składowe spektralne.

W przekazywaniu informacji akustycznej do systemu nerwowego uczestniczą komórki rzęskowe będące receptorami dźwięku. Charakterystycznym elementem ich budowy są subtelne rzęski (*sterocilia*) podlegające naprężeniom mechanicznym podczas drgań błony podstawnej. Ruch błony podstawnej wywołuje przemieszczenia końców rzęsek (utwierdzonych w tak zwanej błonie nakrywkowej) względem ich podstaw, znajdujących się na górnym końcu komórek rzęskowych, których dolne (bazalne) części unieruchomione są z kolei pomiędzy złożoną palisadą komórek Deitersa przytwierdzających je do powierzchni błony podstawnej. W ten sposób komórki rzęskowe (układające się wzdłuż kanału ślimaka w dwóch wyraźnie rozdzielonych grupach– jako komórki zewnętrzne oraz wewnętrzne w stosunku do osi ślimaka), które znajdują się wewnątrz organu Cortiego, stają się czujnikami drgań błony podstawnej,wywołanych falą dźwiękową. Fala dźwiękowa powoduje ruch błony, ta przemieszcza bazalne części komórek rzęskowych, a to powoduje powstawanie naprężeń ścinających w rzęskach. Prowadzi to do powstania tzw. potencjałów czynnościowych – najpierw w receptorach, a potem w neuronach zwoju spiralnego, którego aksony wchodzą w skład nerwu słuchowego.

Większość aferentnych (dośrodkowych) włókien nerwowych ma bezpośrednią styczność z **wewnętrznymi** komórkami rzęskowymi. Jest prawdopodobne, że zewnętrzne komórki rzęskowe aktywnie wpływają na drgania błony podstawnej, a co za tym idzie, na ostrość jej strojenia i czułość. Wydaje się, że zewnętrzne komórki rzęskowe są źródłem nieliniowości odpowiedzi ślimaka, a także wywołują emisje otoakustyczne. Za taką hipotezą przemawia fakt, że na działanie zewnętrznych komórek rzęskowych wpływają bezpośrednio neurony eferentne (odśrodkowe) doprowadzające do nich sygnały z pnia mózgu.



Rys. 1-2. Budowa systemu słuchowego człowieka: z) ucho zewnętrzne, ś) ucho środkowe, w) ucho wewnętrzne.

1.3.2. Sposoby wyznaczania właściwości dźwięku w systemie słuchowym człowieka

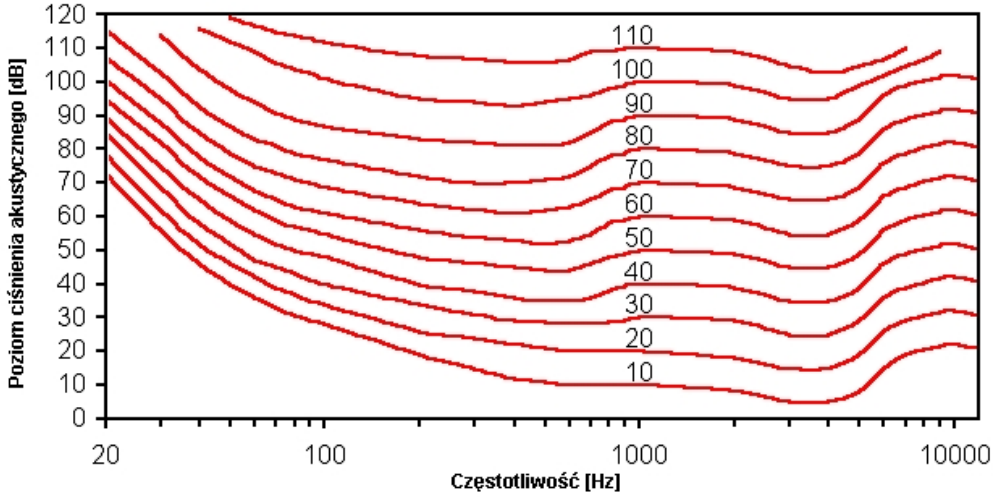
Neurony nerwu słuchowego w sytuacji braku dźwięku wykazują aktywność spontaniczną. Każde z włókien neuronowych charakteryzuje się przy tym określonym progiem pobudzenia i poziomem nasycenia (saturacji), powyżej którego wzrost natężenia dźwięku nie wywołuje już żadnych dalszych zmian odpowiedzi neuronu. Duże częstotliwości wyładowań spontanicznych są stowarzyszone z niskimi progami i małym zakresem dynamiki. Przeciwnie, przy wysokim progu pobudzenia i szerokim zakresie dynamiki działania aktywność spontaniczna (losowa) neuronów słuchowych jest pomijalnie mała.

Próg pobudzenia danego neuronu jest najniższy dla częstotliwości zwanej częstotliwością charakterystyczną i wzrasta dla częstotliwości różnych od częstotliwości charakterystycznej. Wykres progu pobudzenia w zależności od częstotliwości nazywany jest krzywą pobudzenia lub krzywą strojenia. Krzywe strojenia pojedynczych włókien nerwowych są podobne do krzywych strojenia pojedynczych punktów błony podstawnej.

Przebieg czasowy wyładowań neuronu słuchowego zawiera całą dostępną dla mózgu informację o bodźcu słuchowym. Przy niskich częstotliwościach dźwięku wyładowania neuronów pojawiają się zazwyczaj dla określonej fazy bodźca akustycznego. Zjawisko to nazywa się synchronicznością fazową. Synchroniczność fazowa zanika dla częstotliwości większych od 4-5 kHz. Odpowiedź pojedynczego neuronu na ton może być niekiedy zmniejszona przez drugi ton. Zjawisko to może wystąpić nawet wówczas, gdy drugi z tych tonów (prezentowany osobno) nie wywołuje żadnej odpowiedzi analizowanego neuronu. Efekt ten nazywany jest dwutonową supresją i jest przykładem nieliniowego procesu przetwarzania sygnału, zachodzącego w peryferyjnym układzie słuchowym.

Odpowiedzi neuronów wyższych pięter układu słuchowego nie zostały jak dotąd zbadane tak dokładnie jak odpowiedzi neuronów nerwu słuchowego. Niektóre neurony płatu słuchowego kory mózgowej odpowiadają jedynie na bodźce złożone lub na bodźce o zmiennych w czasie charakterystykach [Pal95, Pic88], co wskazuje na to, że w korze słuchowej mogą występować pola gnostyczne i receptory złożonych wzorców sygnału, analogiczne do tych, jakie w korze wzrokowej (kota) wykryli Hubel i Wiesel (nagroda Nobla 1981).

Zarówno próg słyszalności, jak i subiektywnie odbierana głośność zależą od czasu trwania dźwięku. Dla czasów trwania tonów z przedziału ok. 15-120 ms ucho wydaje się sumować energię w czasie w celu detekcji właściwości sygnału. Dlatego też twierdzenie, że próg zależy od całkowitej energii niesionej przez sygnał, a nie od sposobu w jaki energia ta jest rozłożona w czasie, wydaje się w tym obszarze prawdziwe. Nie jest jednak ono słuszne dla długich czasów trwania sygnałów prawdopodobnie dlatego, że czas w którym ucho sumuje energię sygnału jest ograniczony (w przeciwnym razie przy dłuższej ekspozycji na bodziec dźwiękowy doszłoby do nieuchronnego „nasycenia"). Rozkład energii w szerokim paśmie częstotliwości może również ograniczać słyszalność krótkich sygnałów. Człowiek potrafi spostrzegać stosunkowo małe zmiany poziomu dźwięku (0,3-2,0 dB) dla szerokiego zakresu poziomów oraz dla bodźców różnego typu. Ogólnie dyskryminacja systemu słuchowego, wyrażona jako frakcja Webera ($\Delta I/I$) jest niezależna od poziomu dla pasm szumu, ale poprawia się dla tonów o wysokich poziomach.



Rys. 1-3. Krzywe jednakowej głośności dla różnych poziomów głośności.

1.3.3. Psychoakustyczne cechy percepcji dźwięków

Różne eksperymenty psychoakustyczne wskazują na to, że dla wysokich poziomów dźwięku informacje niesione przez neurony o częstotliwościach charakterystycznych większych i mniejszych od częstotliwości badanego sygnału wnoszą swój wkład w dyskryminację natężenia dźwięku, choć informacje te nie mają decydującego znaczenia. Innymi słowy poszerzenie zakresu pobudzenia nie jest **niezbędne** dla istnienia szerokiego zakresu dynamiki układu słuchowego. Podobnie szereg eksperymentów psychofi-

zycznych sugeruje również, że synchroniczność fazowa może odgrywać pewną rolę w dyskryminacji natężenia, choć jej znaczenie nie jest również decydujące [Pla95].

Obserwacje wyładowań pojedynczego neuronu nerwu słuchowego wskazują, że ilość informacji zawartych w częstotliwości wyładowań jest znacznie większa od ilości niezbędnej do wyjaśnienia dyskryminacji natężenia dźwięku, jaką obserwuje się w układzie słuchowym człowieka. W istocie informacje zawarte zaledwie w 100 z 30000 neuronów nerwu słuchowego byłyby w zasadzie wystarczające do pełnego opisania wejściowego wrażenia dźwiękowego. Dlatego też wydaje się, że dyskryminacja natężenia dźwięku nie jest ograniczana przez informacje źródłowe niesione przez nerw słuchowy, ale raczej przez sposób wykorzystania ich przez wyższe piętra układu słuchowego.

Ogólne właściwości analizy częstotliwościowej dokonywanej przez układ słuchowy można streścić wykorzystując koncepcję wstęg krytycznych (układu filtrów percepcyjnych): odpowiedzi słuchaczy na bodźce złożone różnią się w zależności od tego, czy składowe bodźców mieszczą się wewnątrz jednej wstęgi krytycznej, czy też przypadają na kilka wstęg krytycznych. Istnienie wstęg krytycznych potwierdziły eksperymenty dotyczące maskowania jednych dźwięków przez inne, subiektywnej oceny głośności, progu absolutnego, czułości na fazę oraz możliwości słyszenia składowych dźwięku złożonego jako oddzielnych tonów. Na podstawie licznych badań psychoakustycznych stwierdzono, że fizjologiczny filtr słuchowy ma zaokrąglony wierzchołek i opadające zbocza, nie jest więc wygodnym do analizy matematycznej filtrem prostokątnym. Podobnie ustalono, że wygodny parametr, jakim jest szerokość wstęgi krytycznej nie jest wystarczającym opisem filtru słuchowego. Pojęcie wstęgi krytycznej jest wygodnym uproszczeniem pełnej charakterystyki tego filtru, ograniczonym i zubożonym, ponieważ odpowiada ona zaledwie w przybliżeniu jego szerokości, nie wnosząc żadnych informacji o charakterze i kształcie obwiedni.

Pobudzenie neuronowe wywoływane danym dźwiękiem reprezentuje w wyższych piętrach systemu słuchowego dość złożony rozkład aktywności neuronów. Własności dźwięku można przy tym opisać jako wywołaną przez ten dźwięk funkcję częstotliwości charakterystycznej aktywności impulsowej wzbudzonych neuronów. W ujęciu psychofizycznym pobudzenie jest determinowane zależnością sygnałów wyjściowych wszystkich filtrów słuchowych odniesioną do ich częstotliwości środkowych. Pobudzenia wywołane dźwiękiem w obszarze neuronowym są z reguły niesymetryczne – są mniej strome po stronie dużych częstotliwości. Asymetria ta wzrasta ze wzrostem poziomu sygnału. Kształty pobudzenia poszczególnych pól percepcyjnych w mózgu są

przy tym podobne do krzywych maskowania wyznaczonych dla „*maskerów*" wąskopasmowych.

1.3.4. Kodowanie dźwięków w systemie słuchowym człowieka

Sposób kodowania informacji o sygnale dźwiękowym w neuronach wyższych pięter systemu słuchowego nie został jak dotychczas wyjaśniony. Zazwyczaj zakłada się, że ważnym czynnikiem jest w tym przypadku wartość aktywności neuronów o różnych częstotliwościach charakterystycznych, jednak przy bliższych badaniach ten sposób opisu wydaje się nadmiernie uproszczony. Inną możliwością jest to, że struktura czasowa aktywności neuronowej może zawierać informację o bodźcu, choć jest to tylko możliwe dla sygnałów o częstotliwościach mniejszych od 4-5 kHz [Plo86, Moo95a].

Istnieje wiele dowodów na to, że rozdzielczość częstotliwościowa ucha jest bardzo wrażliwa fizjologicznie. Udowodniono to zarówno w ramach badań neurofizjologicznych przeprowadzonych na zwierzętach, jak również w badaniach psychofizycznych ludzi z uszkodzeniami słuchu pochodzenia ślimakowego. Poszerzenie filtrów słuchowych u osób z uszkodzonym słuchem może wywoływać wiele trudności percepcyjnych, których nie można skompensować za pomocą konwencjonalnych aparatów słuchowych [Moo95b]. Aby scharakteryzować rozdzielczość czasową układu słuchowego należy wziąć pod uwagę filtrowanie sygnału zachodzące w peryferyjnym układzie słuchowym. Rozdzielczość czasowa zależy od dwóch zasadniczych procesów: analizy przebiegu czasowego wewnątrz każdego pasma częstotliwości oraz od porównania tych przebiegów w różnych pasmach.

Progi detekcji interwału ciszy w szumie szerokopasmowym zależą najprawdopodobniej od łączenia informacji pochodzących z różnych pasm częstotliwości. Natomiast spostrzeganie interwałów ciszy w dźwięku sinusoidalnym zależy zdecydowanie od fazy, dla której sygnał sinusoidalny jest wyłączany i następnie włączany po interwale ciszy. W niektórych przypadkach funkcje psychometryczne są wyraźnie niemonotoniczne, co można wyjaśnić opierając się na efekcie „wybrzmiewania" filtrów słuchowych. Rezultaty badań, uzyskane dla różnych sygnałów wąskopasmowych sugerują, że filtry słuchowe mogą mieć istotny wpływ na czasową zdolność rozdzielczą słuchu dla małych częstotliwości środkowych i mają relatywnie niewielki wpływ dla częstotliwości środkowych powyżej kilkuset herców [Vie93].

Istnieją dwa główne sposoby kodowania częstotliwości dźwięku: poprzez rozkład aktywności neuronów słuchowych (kodowanie miejscowe lub lokalizacyjne) oraz po-

przez rozkład wyładowań w czasie w poszczególnych neuronach i we wszystkich neuronach danego obszaru percepcyjnego. Wykorzystywanie obydwu wymienionych typów kodowania informacji (zależnie od okoliczności) jest wielce prawdopodobne, ale ich względne znaczenie jest zdecydowanie różne dla różnych zakresów częstotliwości i dla różnych rodzajów dźwięków.

Badania neurofizjologiczne zwierząt wykazały, że zsynchronizowanie impulsów nerwowych z określoną fazą fali pobudzającej zanika dla częstotliwości większych od 4-5 kHz. Równocześnie dla tych właśnie częstotliwości maleje zdolność rozróżniania zmian częstotliwości tonów i zanika poczucie wysokości muzycznej. Zanik ten prawdopodobnie odzwierciedla fakt wykorzystywania przez mózg informacji czasowej jedynie dla częstotliwości mniejszych od 4-5 kHz. Synchroniczność fazowa jest najprawdopodobniej także wykorzystywana w systemie słuchowym człowieka do określenia częstotliwości impulsów tonu oraz detekcji modulacji częstotliwościowej dla bardzo małych częstotliwości modulacji, gdy częstotliwość nośna jest mniejsza od 4-5 kHz. Jednak dla częstotliwości modulacji większych lub równych 10 kHz detekcja modulacji częstotliwościowej prawdopodobnie zależy już przede wszystkim od zmian pobudzenia, czyli od mechanizmu lokalizacyjnego, zatem zależności fazowe nie mają tu żadnego znaczenia.

1.3.5. Dwie teorie percepcji dźwięków

Podsumowując podane rozważania można stwierdzić, że natura percepcji słuchowej, jest złożona, ale może być opisana z pomocą dwóch teorii. Teorie czasowe sugerują, że wysokość dźwięków złożonych (jako kategoria psychoakustyczna) tworzona jest na podstawie odstępów czasowych pomiędzy kolejnymi wyładowaniami neuronów odpowiadających drganiom tego punktu błony podstawnej, w którym nakładają się na siebie sąsiadujące składowe. Z kolei teorie analizy pobudzenia sugerują, że wysokość (jako parametr psychologiczny) tworzona jest w centralnym procesorze, przetwarzającym sygnały neuronowe odpowiadające poszczególnym składowym obecnym w dźwięku złożonym. Według teorii obu typów na percepcję wysokości mogą wpływać tony kombinacyjne o częstotliwościach mniejszych od najmniejszej częstotliwości składowej.

Teorię analizy pobudzenia potwierdza odkrycie dominującej roli niskich harmonicznych tj. takich, które można rozseparować. Innym argumentem na korzyść tej teorii jest to, że wysokość odpowiadająca nieobecnej częstotliwości podstawowej może być percypowana nawet wtedy, gdy nie istnieje możliwość interakcji składowych na błonie

podstawnej. Zademonstrowano to w ramach dychotycznej prezentacji składowych, po jednej do każdego ucha lub przez ich prezentację kolejno w czasie. Z kolei na korzyść teorii czasowej przemawia odkrycie, że (nikłe) wysokości dźwięku mogą być percypowane nawet wtedy, gdy harmoniczne leżą zbyt blisko siebie w dziedzinie częstotliwości, aby mogły być rozseparowane na błonie podstawnej ucha wewnętrznego oraz wtedy, gdy bodźce nie posiadają dobrze zdefiniowanej struktury widmowej (np. przerywany szum) [Hou95, Gil96].

1.3.6. Identyfikacja różnych źródeł dźwięku

Identyfikacja określonego źródła dźwięku zależy od rozpoznania jego barwy. Dla sygnałów ustalonych w czasie barwa zależy przede wszystkim od rozkładu energii w dziedzinie częstotliwości. Jeśli dźwięki analizuje się przy zastosowaniu filtrów o szerokości równej w przybliżeniu szerokości jednej wstęgi krytycznej, to percepcyjne różnice dźwięków są odnoszone do różnic ich widm. Struktura czasowa dźwięku może również wpływać na barwę; szczególnie istotnymi parametrami są w tym przypadku transjent początkowy i obwiednia sygnału.

W celu osiągnięcia percepcyjnej segregacji składowych pochodzących z różnych źródeł może być wykorzystanych wiele różnych wielkości fizycznych związanych z sygnałem. Są wśród nich różnice częstotliwości podstawowej, różnice włączenia, kontrast względem wcześniejszych dźwięków, zmiany częstotliwości i natężenia oraz położenie dźwięku. Jeśli bierze się pod uwagę wyłącznie jeden z tych czynników, to żaden z nich nie jest efektywny. Jeśli jednak uwzględnia się je wszystkie razem, to dostarczają one doskonałych podstaw do rozłożenia wejściowego bodźca akustycznego na strumienie percepcyjne [Ber90]. Jak z tego wynika – identyfikacja źródła dźwięku wymaga stosowania wektorowej reprezentacji cech badanego dźwięku.

1.3.7. Naturalna percepcja sygnału mowy

Mowa jest bodźcem wielowymiarowym, który zmienia się w złożony sposób w dziedzinie częstotliwości i czasu. Chociaż sygnał mowy można opisać wyłącznie za pomocą zmian amplitudy w czasie, to z punktu widzenia układu słuchowego taki sposób opisu nie jest najodpowiedniejszy. Badania pokazują, że opis sygnału mowy za pomocą uśrednionego w czasie widma jest również niezadowalający. Aby osiągnąć opis sygnału mowy zgodny z poznanymi dotychczas zasadami funkcjonowania układu słuchowego człowieka, należy **jednocześnie** przedstawić zmiany sygnału mowy w dziedzinie częstotliwości, natężenia i czasu.

Podstawowym zagadnieniem rozważań percepcji mowy jest powiązanie właściwości dźwięku mowy z poszczególnymi cząstkami językowymi. Nie do końca wyjaśniony jest problem, czy podstawową cząstką percepcji mowy przez człowieka jest sylaba, fonem czy jakaś inna wielkość percepcyjna, taka jak np. ustalona cecha fonetyczna. Wiele dowodów psychoakustycznych świadczy o tym, że dla szybkiej mowy ciągłej przebiegi akustyczne sygnalizujące obecność w wypowiedzi określonych cech fonetycznych lub fonemów pojawiają się za szybko, by mogły być oddzielnie spostrzegane przez mózg człowieka w odpowiedniej kolejności. Dlatego też bardziej prawdopodobna jest hipoteza, że w mózgu zachodzi (jako elementarny składnik percepcji) rozpoznawanie całego przebiegu dźwięku odpowiadającemu dłuższemu segmentowi mowy, takiemu jak np. sylaba.

W tej pracy założono jednak, że podstawowymi cząstkami percepcji mowy są najprawdopodobniej fonemy lub cechy fonetyczne, ale informacja o kilku fonemach lub cechach fonetycznych jest ekstrahowana jednocześnie w stosunkowo długich przedziałach czasowych dźwięku.

Drugim zagadnieniem wymagającym wyjaśnienia, jest znalezienie takich czynników charakteryzujących sygnał akustyczny, które wskazują jednoznacznie na obecność określonej cząstki lingwistycznej. W przypadku pojedynczych słów lub sylab wypowiedzianych oddzielnie i wyraźnie, możliwe jest znalezienie wielu względnie niezależnych czynników, umożliwiających identyfikację każdego wchodzącego w ich skład fonemu bądź też pewnych cech fonetycznych. Czynniki te jednak nie są wystarczająco skuteczne, jeśli weźmie się pod uwagę szybką mowę ciągłą. W wielu przypadkach fonem może być poprawnie zidentyfikowany dopiero na podstawie informacji pochodzącej od całej sylaby lubnawet od grupy kilku słów. Złożone związki pomiędzy sygnałem mowy, a cząstkami językowymi mowy są jednym z najważniejszych zagadnień percepcji mowy i wywierają znamienny wpływ na rozwój teorii percepcji mowy. W tej pracy zagadnienie to nie będzie jednakoddzielnie analizowane.

W licznych badaniach percepcyjnych wykazano ponad wszelką wątpliwość, że mowa jest dla ludzkiego systemu słuchowego szczególnym rodzajem bodźców akustycznych, a sygnały mowy są percypowane i przetwarzane w inny sposób niż bodźce dźwiękowe nie będące mową. Postuluje się w związku z tym istnienie w mózgu specjalnego „modu mowy", zajmującego się wybiórczo analizą i percepcją wyłącznie tego właśnie specyficznego sygnału. Wśród dowodów potwierdzających istnienie takiego modułu można podać następujące fakty:

- stwierdzono, że istnieje tzw. percepcja kategorialna gdyż zaobserwowano w badaniach psychoakustycznych, że istnieje efekt lepszej dyskryminacji tych dźwięków, które językowo identyfikowane są jako różne;
- wykazano, że występuje znamienna asymetria funkcji mózgu– pewne obszary mózgu specjalizują się wyłącznie w przetwarzaniu mowy i występują one tylko w tzw. półkuli dominującej (u większości ludzi jest to lewa półkula);
- wykryto, że występuje tzw. percepcja dualna – pojedynczy przebieg akustyczny, taki jak np. przejście formantu, może być słyszany (zależnie od okoliczności) zarówno jako część dźwięku mowy, jak i oddzielny dźwięk nie będący sygnałem mowy;
- wykazano równoważność czynników percepcji mowy, umożliwiających identyfikację dźwięku mowy i innych dźwięków, to znaczy wykazano, że określony parametr akustyczny ma różne właściwości psychologiczne, zależne od tego, czy sygnały percypowane są jako mowa czy też jako sygnały nie będące mową;
- wykonano prace dotyczące audiowizualnego łączenia informacji pokazujące, że na percypowaną tożsamość sygnału wpływ ma zarówno to co właściwie słyszymy, jak i to co potrafimy „odczytać" z twarzy mówcy;
- dźwięki mowopodobne są percypowane (zależnie od okoliczności) jako mowa lub jako sygnał niejęzykowy, przykładem takiego fenomenu jest percepcja „mowy sinusoidalnej" (człowiek słucha i rozumie sygnał mowy, w którym pierwsze trzy formanty zamienione na sygnały sinusoidalne o częstotliwościach odpowiadających tym formantom). Zależnie od okoliczności sztuczny sygnał „mowy sinusoidalnej" słyszany jest jako mowa, albo jako sztuczne dźwięki nie będące mową.

1.3.8. Wyjaśnienie fenomenów związanych z percepcją mowy

Jest raczej oczywiste, że percepcja mowy jest procesem zachodzącym na różnych poziomach systemu słuchowego. Na każdym z tych poziomów rozstrzyganie dwuznaczności i kompensowanie błędów (będących wynikiem przetwarzania sygnału na innych poziomach) dokonywane jest na podstawie innej porcji informacji zawartych w sygnale. Wstępna analiza sygnałów mowy pod kątem cech fonetycznych, fonemów lub sylab może być potwierdzona i skorygowana na wyższych piętrach systemu na podstawie znajomości wpływu jednego dźwięku na inne dźwięki. Nasza wiedza syntaktyczna (czyli wiedza na temat reguł gramatycznych) i semantyczna (czyli wiedza dotycząca

znaczenia słów) umożliwiają dalsze dopasowanie i korektę wstępnej, czysto akustycznej i fonetycznej identyfikacji sygnału, podczas gdy pewne czynniki percepcji wynikające z okoliczności słuchania, takie jak np. tożsamość mówcy czy kontekst rozmowy, dostarczają jeszcze dodatkowych informacji, pozwalających „odgadnąć" brakujące dane.

Wydaje się, że przetwarzanie sygnału mowy w systemie słuchowym człowieka nie ma charakteru hierarchicznego, tj. rozpoznawane elementy nie przechodzą kolejno z jednego poziomu do drugiego, ale jest wiele rozgałęzionych połączeń (w tym zwłaszcza sprzężeń zwrotnych) pomiędzy wszystkimi poziomami tej analizy. Informacja dostępna na określonym poziomie może być zawsze przeanalizowana powtórnie przez wcześniejszy poziom na podstawie dodatkowej informacji (np. kontekstowej) pochodzącej z innych poziomów. Obserwacje psychoakustyczne wskazują na to, że wiele szczegółów dotyczących struktury sygnału mowy może być przechowywanych w pamięci przez krótki czas, tak że możliwa jest ponowna analiza sygnału w razie pojawienia się trudności z jego interpretacją [Nyg95].

1.3.9. Dodatkowe informacje zawarte w sygnale mowy

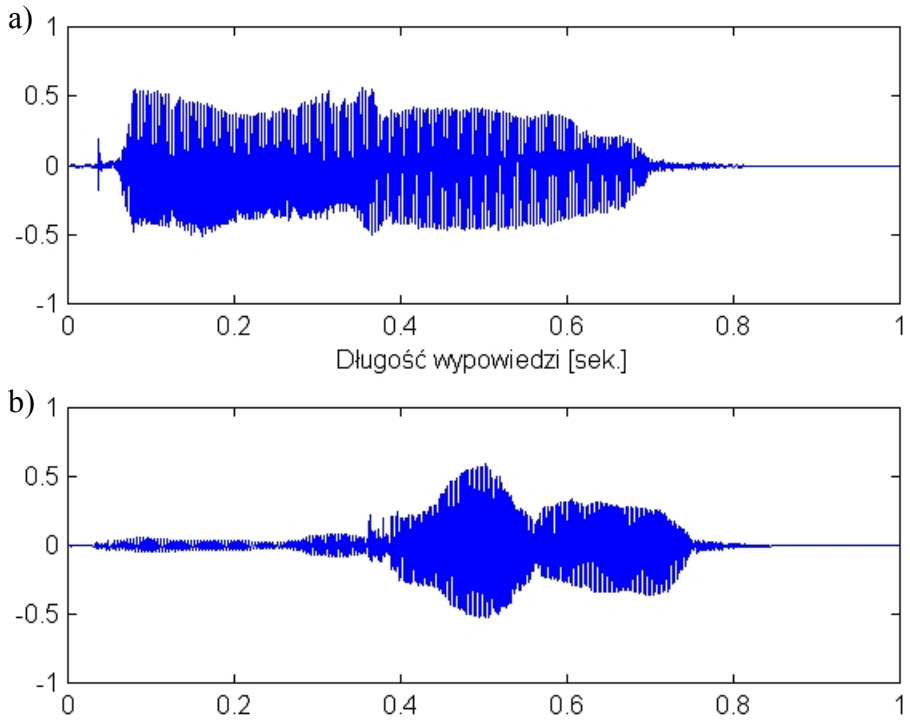
Wielowymiarowa natura sygnałów mowy oraz różnorodność niezależnych informacji dostępnych na każdym poziomie jej przetwarzania sugeruje, że sygnał mowy zawiera w sobie znacznie więcej informacji, niż jest to niezbędne dla poprawnej jego identyfikacji w samej tylko warstwie semantycznej. Fakt ten znajduje swe odbicie w niewielkim wpływie nawet bardzo silnych zakłóceń na zrozumiałość mowy. Mowa może być poprawnie rozumiana w obecności szumu o znacznym poziomie a także wówczas, gdy jest w znaczący sposób zmieniona przez jej filtrowanie lub przez zawężanie zakresu zmian amplitudy sygnału. Dlatego też mowa jest wysoce wydajnym, niezawodnym, nawet w trudnych warunkach, narzędziem porozumiewania się.

1.4. Trudności związane z wizualizacją mowy patologicznej

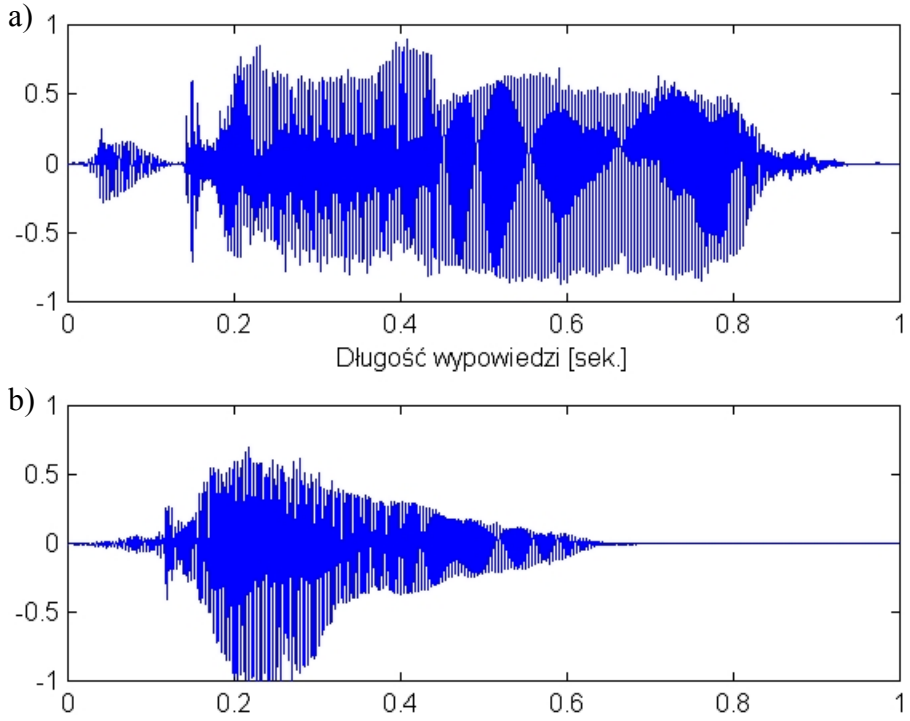
W niniejszej pracy zaproponowano konkretną metodę wizualizacji mowy patologicznej, dzięki której wypowiedzi człowieka mogą zostać przekształcone w nadające się do interpretacji obrazy graficzne. Na podstawie wcześniejszych badań [Tad99] ustalono ponad wszelką wątpliwość, że obrazy takie mogą być pomocne na przykład dla lekarzy

specjalistów, decydujących o tym czy pewna operacja dotycząca narządów mowy została poprawnie wykonana, albo ustalających, czy pacjent wymaga albo nie wymaga określonej rehabilitacji. Dzięki takim obrazom mowy zwizualizowanej dotychczasowe subiektywne czy intuicyjne przyjmowane rozstrzygnięcia (w szczególności medyczne) mogą stać się lepiej umotywowane i bardziej obiektywne.

Należy podkreślić, iż subiektywna wrażliwość ucha ludzkiego i całego systemu percepcji słuchowej człowieka na rozmaite nieprawidłowości w strukturze i przebiegu analizowanego sygnału jest bardzo zróżnicowana. Niekiedy duże różnice w formie sygnału mogą być subiektywnie niedostrzegane lub akceptowane jako granice normy, w innym przypadku nawet drobne rozbieżności pomiędzy sygnałami stanowią kryterium do ich rozróżniania i są podstawą do dyskwalifikacji jednego z nich jako formy patologicznej.



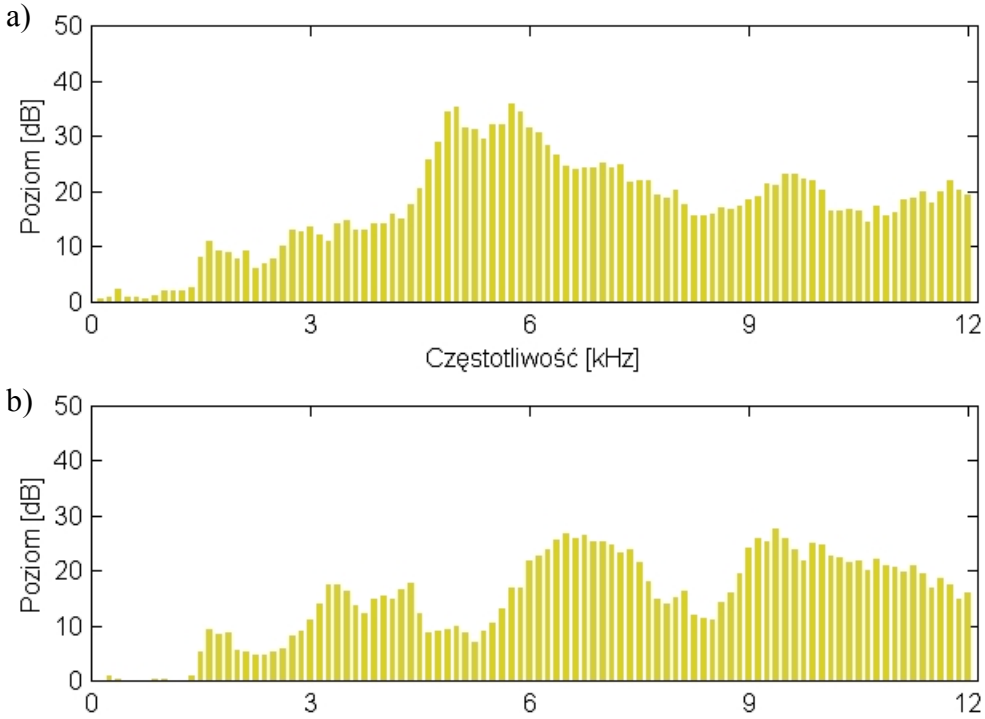
Rys. 1-4. Przebiegi czasowe wzorcowych wypowiedzi słowa „*goni*”. Oba przebiegi są subiektywnie oceniane jako poprawne mimo dużych różnic widocznych w ich strukturze czasowej.



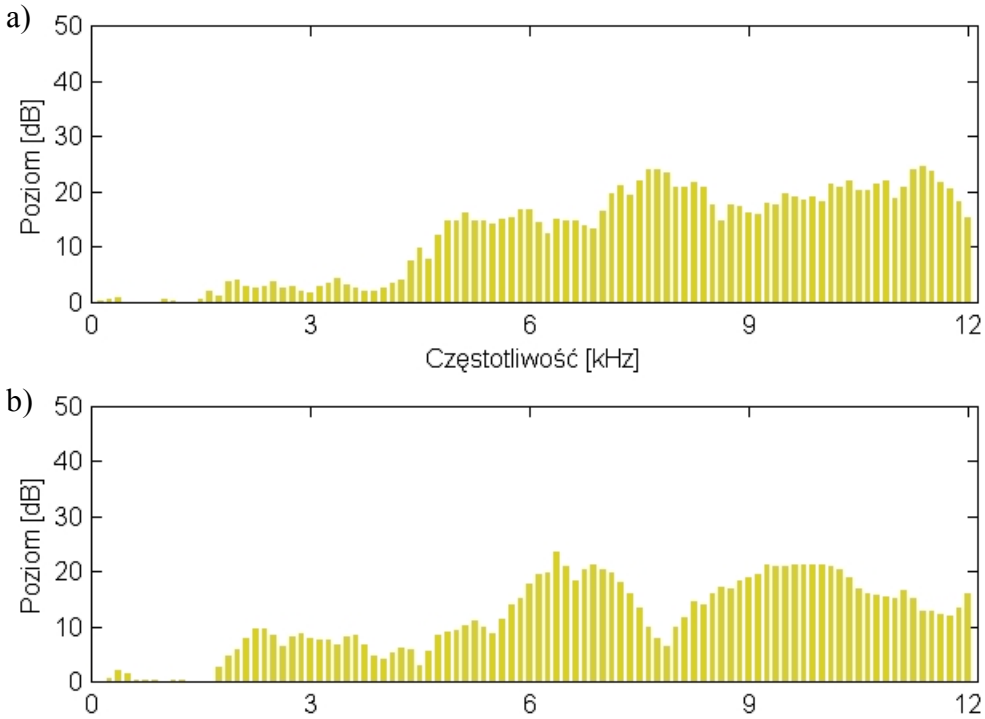
Rys. 1-5. Przebiegi czasowe patologicznych (rozszczep podniebienia wtórnego częściowy) wypowiedzi słowa „*goni*”. Widoczna jest odmienność sygnału w stosunku do próbek mowy prawidłowej (z poprzedniego rysunku), jednak różnice te (wobec dużej zmienności indywidualnej) są trudne do konkretnego nazwania.

Na rys. 1-4 przedstawiono dwa przebiegi czasowe prawidłowej wypowiedzi słowa „*goni*”. Dla porównania rys. 1-5 prezentuje dwa przebiegi czasowe wypowiedzi tego samego słowa przez pacjentów z częściowym rozszczepem podniebienia wtórnego (dla wygody porównania obydwa rysunki przedstawione są w tej samej skali). Na wykresach tych ciężko jest znaleźć jakieś cechy wspólne dla każdej z tych dwóch kategorii wypowiedzi, a równocześnie skutecznie różnicujące te wypowiedzi w stosunku do wypowiedzi z drugiej grupy (tzn. różnicę „dowolnej” wypowiedzi prawidłowej i „dowolnej” wypowiedzi patologicznej). Spostrzeżenie to praktycznie dyskwalifikuje możliwość zastosowania tego typu wykresów do analizy i wzrokowej oceny poprawności wypowiedzi.

Również przedstawienie wypowiedzi tylko w dziedzinie częstotliwości nie jest wystarczające do poprawnego rozróżnienia wypowiedzi wzorcowych i patologicznych. To stwierdzenie potwierdzają obrazy uśrednionych widm głoski „*s*” zaprezentowane na rys. 1-6 i 1-7.

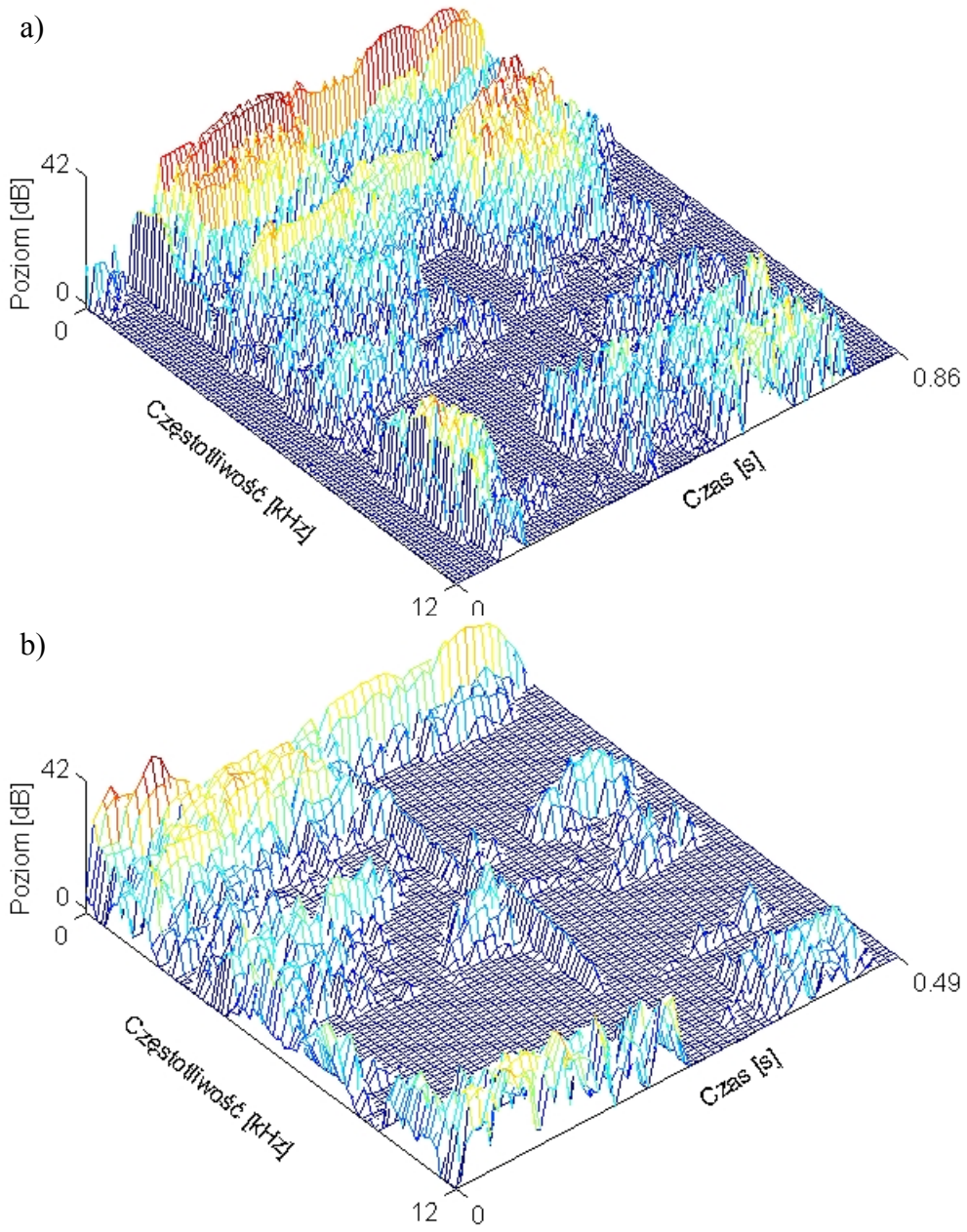


Rys. 1-6. Uśrednione widmo głoski „*s*" – wypowiedź wzorcowa.

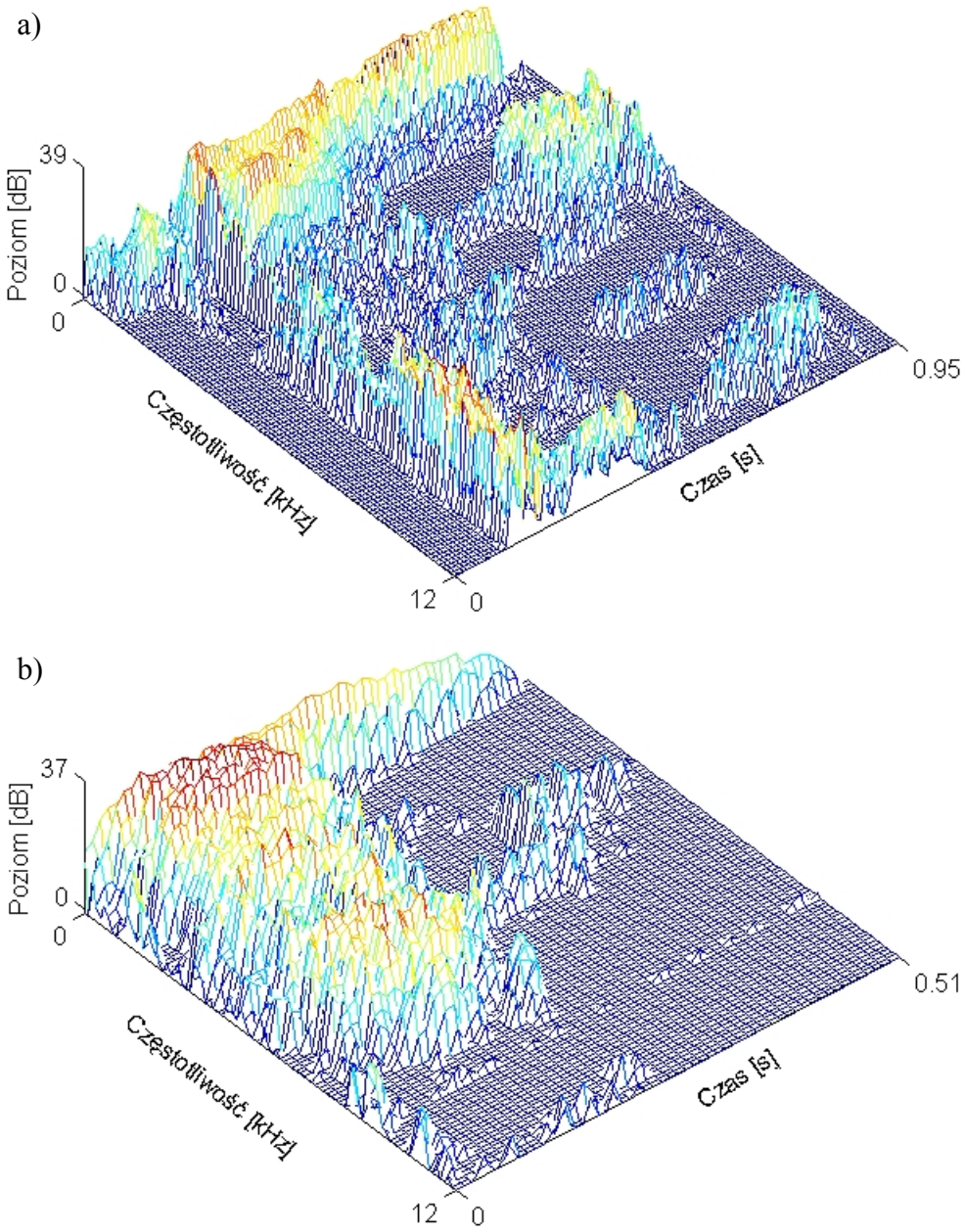


Rys. 1-7. Uśrednione widmo głoski „*s*" – wypowiedź patologiczna (rozszczep podniebienia wtórnego częściowy).

Najbardziej popularnym sposobem obrazowania sygnałów akustycznych są wspomniane już wyżej wideogramy, prezentujące zmianę widma sygnału w czasie. Wprawdzie wideogramy pozwalają na dokładną analizę sygnału, jednak nie we wszystkich przypadkach możliwe jest stwierdzenie, czy wideogram przedstawia wypowiedź wzorcową, czy patologiczną. Jako przykład niech posłużą wideogramy wypowiedzi wzorcowych i patologicznych słowa „*goni*". Na rysunku 1-8 przedstawiono dwa wideogramy wypowiedzi uznanych za wzorcowe, natomiast na rysunku 1-9 zamieszczono dwa wideogramy wypowiedzi patologicznych (pacjentów z częściowym rozszczepem podniebienia wtórnego).



Rys. 1-8. Wideogramy wzorcowych wypowiedzi słowa „*goni*".



Rys. 1-9. Wideogramy patologicznych (rozszczep podniebienia wtórnego częściowy) wypowiedzi słowa „*goni*”.

Analizując zamieszczonewyżej wykresy można dojść do wniosku, iż z samej postaci graficznej wideogramów ciężko jest wywnioskować, które wideogramy reprezentują wypowiedzi wzorcowe, a które patologiczne. Wydaje się, że wideogram przedstawiony na rysunku 1-9a, pomimo tego, że jest to obraz wypowiedzi patologicznej, ma więcej cech wspólnych z wzorcowym wideogramem przedstawionym na rys. 1-8a, niż drugi wzorcowy wideogram przedstawiony na rys. 1-8b. Kolejnym faktem znacząco komplikującym proces wizualizacji jest zmienna długość wypowiedzi tego samego

słowa. Wyraźnie widać to na rysunkach 1-8b i 1-9a, pomimo iż „*goni*" jest wyrazem krótkim to rozpiętość czasów jego wypowiedzi wynosi od 0,48 sek. do 0,95 sek. (najdłuższa wypowiedź jest prawie dwa razy dłuższa od wypowiedzi najkrótszej). Zatem konstruowany system obrazowania sygnału mowy musi wszystkie takie różnice uwzględniać, maksymalnie „upodobniając się" do preferencji właściwych psychofizycznym własnościom słuchu ludzkiego.

1.5. Teza pracy

Jak wspomniano powyżej, metoda zobrazowania mowy powinna nawiązywać koncepcyjnie do psychologii słyszenia, ponieważ chodzi o pokazanie na obrazie zmian w mowie dostrzegalnych słuchem, ale źle się przekładających na tradycyjną analizę parametryczną. W paragrafie poświęconym opisowi systemu słuchowego człowieka pokazano, iż niebagatelną rolę w procesie „obrazowania" sygnału mowy odgrywa błona podstawna (dokonująca analizy częstotliwościowej sygnału) oraz neurony nerwu słuchowego. Ponadto stwierdzono, że aby analizować sygnał mowy zgodnie z poznanymi dotychczas zasadami funkcjonowania układu słuchowego, należy jednocześnie przedstawić zmiany sygnału w dziedzinie częstotliwości, natężenia i czasu. Biorąc pod uwagę powyższe stwierdzenia wydaje się być uzasadnionym **oparcie konstrukcji systemu wizualizacji mowy o sieć neuronową Kohonena,** której wejście stanowić będzie opis sygnału w dziedzinie czasowo-częstotliwościowej. Jako tezę rozprawy można więc zaproponować następujące stwierdzenie:

> **Istnieje efektywna możliwość zastosowania sieci neuronowych samoorganizujących się typu Kohonena do prezentacji mowy patologicznej w postaci dwuwymiarowych obrazów. Obrazy takie mogą być bezpośrednio przedmiotem wizualizacji, wspomagając lekarzy w obiektywnej ocenie stopnia deformacji mowy, lub mogą służyć jako dane wejściowe do dalszych etapów przetwarzania sygnałów akustycznych – między innymi w systemach automatycznej diagnostyki medycznej.**

Komentując zaproponowaną tezę pracy należy podkreślić, że w pracy zaproponowano całkowicie oryginalną metodę wizualizacji sygnałów akustycznych z użyciem sieci neuronowej Kohonena, której struktura, metoda działania i sposób wykorzystania stanowią oryginalny własny dorobek autora rozprawy. W celu wykazania tezy pracy stworzono także specjalistyczne oprogramowanie (wykorzystujące zaproponowaną

metodę) umożliwiające zupełnie nową co do formy wizualizację sygnału mowy – zarówno prawidłowej, jak i patologicznej. Przy pomocy tego oprogramowania przeprowadzono oryginalne badania, których rezultaty zostały zamieszczone w pracy. Celem tych badań było ustalenie optymalnej formy wizualizacji, a także określenie wpływu następujących czynników na cechy generowanych obrazów:

- rodzaj obrazowanych fragmentów mowy (słowa, głoski),
- zastosowanie różnych definicji sąsiedztwa w sieci Kohonena,
- dobór rozmiaru sieci i rozdzielczości prezentacji,
- długość procesu uczenia, niezbędna do uzyskania dobrego wyniku bez efektów „przeuczenia".

W tekście pracy wykazane zostanie, że zaproponowana metoda daje dobre wyniki w wielu zadaniach związanych z wizualizacją mowy patologicznej, w szczególności jest przydatna do jakościowej oceny wypowiedzi dzieci z rozszczepem podniebienia.

1.6. Struktura pracy

Praca składa się z sześciu rozdziałów oraz wykazu wykorzystanej literatury. W rozdziale 1 (niniejszym) wskazano cel pracy, przedstawiono aktualny stan badań dotyczących przetwarzania sygnału mowy, omówiono podstawy funkcjonowania układu słuchowego człowieka oraz wskazano na trudności występujące podczas obrazowania sygnałów akustycznych. Na końcu rozdziału zamieszczono tezę rozprawy. Rozdział 2 poświęcono krótkiemu opisowi sieci neuronowych, ze szczególnym uwzględnieniem sieci Kohonena. W rozdziale 3 omówiono materiał badawczy, sposób jego zarejestrowania oraz strukturę danych stanowiącą dane wejściowe do dalszego etapu wizualizacji. Z kolei rozdział 4 zawiera opis proponowanej metody wizualizacji, przedstawia zatem istotę najistotniejszej koncepcji, mającej kluczowe znaczenie dla całej pracy. Rozdział 5 stanowi główną eksperymentalną cześć pracy i przedstawia wyniki przeprowadzonych przez autora badań. W pierwszej części zaprezentowano obrazy pojedynczych słów, natomiast w drugiej części obrazy izolowanych głosek, uzyskane z pomocą zaproponowanej w pracy techniki wizualizacji. W ostatniej części rozdziału przedstawiono ogólne wnioski dotyczące obrazowania sygnału mowy przy pomocy sieci Kohonena. Rozdział 6 stanowi podsumowanie uzyskanych wyników oraz wskazuje dalsze kierunki rozwoju metody wizualizacji zaproponowanej w rozprawie.

2. Sieci neuronowe

2.1. Ogólne własności sieci neuronowych

Historia sztucznych sieci neuronowych sięga lat 40-tych, kiedy to McCulloch i Pitts opracowali pierwszy model sztucznej komórki nerwowej [McC43], jednakże dopiero w połowie lat 80. zaobserwowano gwałtowny wzrost zainteresowania tą dziedziną. Ogromna popularność sieci neuronowych i rosnąca fala ich zastosowań mają swoje konkretne przyczyny. Są one związane z pewnymi właściwościami tych sieci, które warto tutaj skrótowo zestawić i podsumować.

- Możliwość nauki – sieci nie trzeba specjalnie programować, samoczynnie dopasowuje ona swoje parametry do charakterystyki rozwiązywanego zadania w trakcie procesu uczenia. Zamiast projektować algorytm wymaganego przetwarzania informacji stawia się więc sieci przykładowe zadania (najczęściej takie, dla których poprawne rozwiązanie jest znane) i automatycznie, zgodnie z założoną strategią uczenia, modyfikuje się połączenia sieci oraz współczynniki wagowe w taki sposób, by rozwiązanie dostarczane przez sieć było zgodne ze znanym rozwiązaniem wzorcowym.
- Umiejętność uogólniania nabytej wiedzy– jest to jedna z podstawowych zalet sieci neuronowych, pozwala ona na generowanie właściwego rozwiązania również dla takich danych, które nie zostały przedstawione podczas procesu nauki.
- Równoległe przetwarzanie danych przez neurony w warstwach sieci – każdy neuron w sieci działa niezależnie od pozostałych (wykorzystując jedynie docierające do niego sygnały), przeto stopień współbieżności obliczeń w sieciach neuronowych realizowanych sprzętowo (jako specjalizowane układy elektroniczne) może być setki razy większy niż w najnowocześniejszych systemach wieloprocesorowych.
- Mała wrażliwości na lokalne uszkodzenia– sieci zachowują zdolność poprawnego działania nawet po uszkodzeniu części ich elementów.

- Łatwa realizacja sprzętowa – najczęściej sieci realizowane są jako programy symulacyjne wykorzystywane do naśladowania działania sieci przez typowy komputer klasy PC lub Workstation, jednak istnieje możliwość zrealizowania sieci w postaci specjalizowanego układu w cyfrowej i analogowej technice VLSI [Gra89, LeC89, Mea89, Suz89] lub w technice optycznej czy optoelektronicznej [Abu89, Far89, Psa88, Psa89, Wag87].

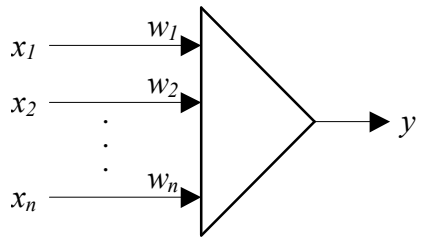
Do zalet sieci istotnych z punktu widzenia analizy i rozpoznawania mowy należy także, wcześniej już wspominana, możliwość zintegrowania sieci z takimi „klasycznymi" technikami rozpoznawania mowy, jak programowanie dynamiczne i ukryte modele Markowa [Ben92]. Wadą sieci neuronowych jest bez wątpienia czas potrzebny do nauki sieci. Ponadto na podstawie doniesień literaturowych można stwierdzić, że nie da się wykazać przewagi sieci neuronowych w tych problemach, w których istnieją dobrze zbadane relacje pomiędzy wejściem a wyjściem systemu, oparte na określonych prawach przyczynowych – na przykład na fizycznych równaniach bilansu masy i energii czy też prawach dynamiki. Pomimo tego faktu sieci neuronowe znajdują coraz szersze zastosowanie w wielu dziedzinach, aby wymienić tylko niektóre z nich:

- diagnostyka medyczna [Bax95, Bur97, Mic95, Świ97],
- klasyfikacja i rozpoznawanie wzorców [Bis95, Gaj99b, Keh97, Wid88],
- predykcja sygnałów [Hay93, Vem94, Wei90, Wu94],
- aproksymacja funkcji wielu zmiennych [Pog89, Bro88, Hor91],
- sterowanie procesami dynamicznymi [Kur98],
- zapamiętywanie wzorców, pamięci asocjacyjne [Zub94].

W kolejnych paragrafach zaprezentowane zostaną podstawowe pojęcia związane z sieciami neuronowymi. Istnieje jednak tak wiele typów i odmian sieci, iż nie można w jednym rozdziale, nawet bardzo pobieżnie, omówić ich wszystkich. Ponieważ w niniejszej pracy wykorzystane zostały sieci Kohonena, dlatego tylko one zostaną w miarę dokładnie przedstawione. Więcej informacji na temat innych sieci neuronowych można znaleźć w następujących pozycjach literaturowych [Hay94, Her93, Kac95, Koh84, Kor94, Mas92, Mor92, Orr96, Oso96, Ron95, Rum91, Spe91, Tad93, Tad98b].

2.2. Zasada funkcjonowania pojedynczego neuronu

Pojedynczy sztuczny neuron, podobnie jak jego biologiczny odpowiednik, posiada n wejść $(x_1, x_2, \mathrm{K}\ , x_n)$ oraz jedno wyjście y. Elementem mającym decydujący wpływ na funkcjonowanie neuronu jest zbiór współczynników wagowych $(w_1, w_2, \mathrm{K}\ , w_n)$, zwanych wagami synaptycznymi. W celu obliczenia wartości sygnału wyjściowego z neuronu, odpowiednie składowe sygnału wejściowego mnożone są przez te współczynniki wagowe. Na rysunku 2-1 przedstawiono schemat pojedynczego sztucznego neuronu.



Rys. 2-1. Schemat neuronu.

Sygnał wyjściowy y neuronu opisuje równanie:

$$y = \varphi(e) \tag{2.1}$$

gdzie: e – sygnał pobudzenia neuronu,

$\varphi(e)$ – funkcja aktywacji neuronu.

Sygnał pobudzenia neuronu obliczany jest zgodnie ze wzorem:

$$e = \sum_{i=1}^{n} w_i x_i + b \tag{2.2}$$

gdzie: x_i – i-ta składowa sygnału wejściowego,

w_i – i-ta waga wektora wag,

n – liczba wejść neuronu,

b – składnik stały (*bias*).

W przypadku neuronów z liniową funkcją aktywacji wartość b jest zerowa. Gdy wartość b jest różna od zera, to często poszerza się wektor sygnałów wejściowych oraz wektor wag o dodatkowy element:

$$x_0 = 1 \tag{2.3}$$

$$w_0 = b \tag{2.4}$$

Przy takim poszerzeniu wektorów wzór na sygnał pobudzenia neuronu można zapisać w następującej postaci:

$$e = \sum_{i=1}^{n} w_i x_i = \mathbf{w}\,\mathbf{x} \tag{2.5}$$

gdzie: $\mathbf{w} = [b, w_1, w_2, \mathrm{K}\ , w_n]$,

$\mathbf{x} = [1, x_1, x_2, \mathrm{K}\ , x_n]^T$.

Funkcja aktywacji neuronu $\varphi(e)$ może być funkcją liniową lub nieliniową. W przypadku funkcji nieliniowych najczęściej stosuje się następujące funkcje [Tad93]:

1. Funkcja signum:

$$\varphi_1(e) = \begin{cases} 1 & e > 0 \\ 0 & e = 0 \\ -1 & e < 0 \end{cases} \tag{2.6}$$

2. Funkcja logistyczna:

$$\varphi_2(e) = \frac{1}{1 + \exp(-\beta\, e)}, \quad \varphi_2(e) \in (0,1) \tag{2.7}$$

3. Tangens hiperboliczny:

$$\varphi_3(e) = \tanh(\beta\, e), \quad \varphi_3(e) \in (-1,1) \tag{2.8}$$

Współczynnik β w powyższych wzorach decyduje o kształcie funkcji, im jest on większy, tym bardziej stroma jest postać funkcji.

Wybór funkcji aktywacji ma istotne znaczenie nie tylko dla działania neuronu, ale także dla procesu jego treningu. Najbardziej istotne jest aby funkcja była różniczkowalna, chociaż stopień złożoności obliczeniowej samej funkcji aktywacji jak i jej po-

chodnej ma także pewne znaczenie. Funkcje $\varphi_2(e)$ i $\varphi_3(e)$ są różniczkowalne, a ich zaletą jest prosta metoda obliczania wartości ich pochodnych:

$$\frac{d\varphi_2(e)}{de} = \beta\,\varphi_2(e)\cdot(1-\varphi_2(e)) \tag{2.9}$$

$$\frac{d\varphi_3(e)}{de} = \beta\left(1-\varphi_3^2(e)\right) \tag{2.10}$$

które dają się wyrazić poprzez wartości samych funkcji, co jest szczególnie wygodne w sytuacji, gdy podczas procesu uczenia dostępna jest wartość tej funkcji (jako sygnał na wyjściu neuronu) a trzeba znaleźć wartość pochodnej w celu odpowiedniego skalowania wartości błędu podczas jego wstecznej propagacji.

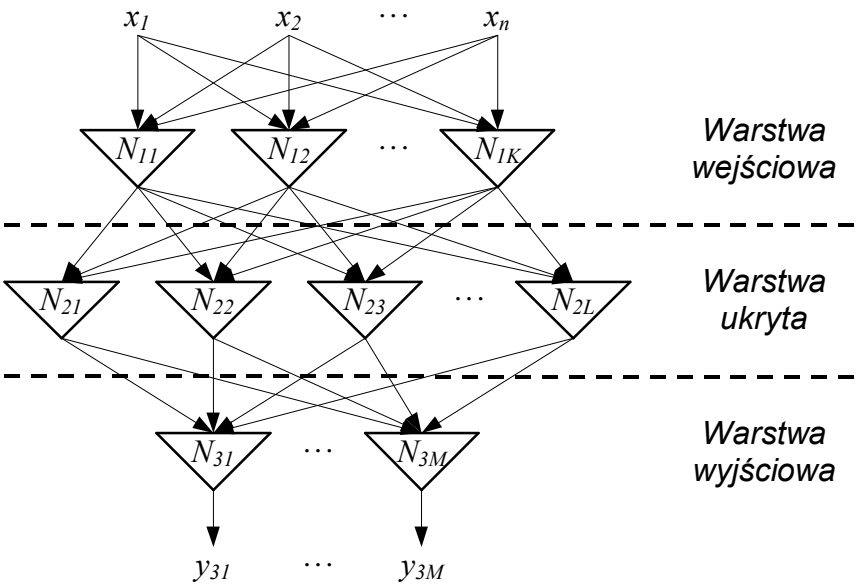
2.3. Rodzaje sieci

Istnieje wiele rodzajów i odmian sieci neuronowych. Ogólnie rzecz biorąc klasyfikację sieci neuronowych można dokonywać na podstawie jednej z poniższych cech:

- kierunek przepływu danych w sieci:
  - sieci jednokierunkowe (*feedforward*) – sygnały wyjściowe neuronów przekazywane są tylko do neuronów kolejnej warstwy,
  - sieci rekurencyjne – neurony mają sprzężenia zwrotne,
- rodzaj funkcji aktywacji neuronów:
  - sieci liniowe,
  - sieci nieliniowe,
- liczba warstw neuronów w sieci:
  - sieci jednowarstwowe,
  - sieci wielowarstwowe,
- sposób połączenia między kolejnymi warstwami neuronów:
  - połączenie zupełne,
  - połączenie niezupełne.

W przypadku sieci wielowarstwowych wyróżnia się 3 typy warstw: warstwę wejściową, warstwę wyjściową oraz warstwy ukryte (są to wszystkie warstwy, które nie są wejściowe ani wyjściowe). Każda warstwa sieci zazwyczaj zbudowana jest z neuronów tego samego typu, jednak poszczególne warstwy (w obrębie jednej sieci) mogą być zło-

żone z neuronów różnego typu. Na rysunku 2-2 przedstawiono przykładową sieć wielowarstwową złożoną z trzech warstw.



Rys. 2-2. Wielowarstwowa sieć neuronowa typu *feedforward*.

Sygnał wejściowy $(x_1, x_2, \mathrm{K}\ , x_n)$ podawany jest do neuronów warstwy wejściowej, zbudowanej z *K* neuronów. Następnie sygnał wyjściowy z tych neuronów przekazywany jest do neuronów warstwy ukrytej, z kolei wyjścia neuronów warstwy ukrytej połączone są z wejściami neuronów kolejnej warstwy ukrytej lub warstwy wyjściowej. Omawiana sieć jest siecią typu *feedforward* ponieważ przepływ danych w sieci następuje tylko w jednym kierunku (od warstwy wejściowej do warstwy wyjściowej). Dodatkowo w sieci tej połączenia pomiędzy neuronami poszczególnych warstw są zupełne, tzn. każdy neuron otrzymuje sygnały od każdego neuronu warstwy poprzedniej.

2.4. Inicjalizacja wag neuronów

Przed rozpoczęciem treningu sieci należy odpowiednio zainicjalizować wagi neuronów. Zły dobór początkowych wartości wag może być przyczyną znacznych trudności, a wręcz niepowodzeń, w procesie poprawnego treningu sieci. Należy tak dobrać początkowe wartości wag, aby zmniejszyć prawdopodobieństwo zatrzymania procesu nauki w minimum lokalnym i maksymalnie przyspieszyć proces treningu sieci. Jedną z metod rozwiązania tego problemu jest ustalenie odpowiednio małych wartości począt-

kowych wag. Najczęściej korzysta się z inicjacji losowej, gdzie rozkład generowanych wartości jest równomierny w określonym przedziale. Poniżej przedstawiono przykładowe rozmiary tych przedziałów [Thi95]:

1. Przedział Bottona:

$$\left[-\frac{a}{\sqrt{n}};\frac{a}{\sqrt{n}}\right] \tag{2.11}$$

gdzie: n – liczba wejść neuronu,
a – parametr dobierany dla każdego neuronu.

2. Przedział Śmieji:

$$\left[-\frac{\sqrt{n}}{2};\frac{\sqrt{n}}{2}\right] \tag{2.12}$$

gdzie: n – liczba wejść neuronu.

3. Przedział Nguyena-Widrowa:
   - dla neuronów warstwy wejściowej i ukrytej:

$$\left[-0.7\cdot\sqrt[n]{N_k}\;;\;0.7\cdot\sqrt[n]{N_k}\right] \tag{2.13}$$

gdzie: n – liczba wejść neuronu,
N_k – liczba neuronów w warstwie k.

   - dla neuronów warstwy wyjściowej:

$$\left[-0.5\;;\;0.5\right] \tag{2.14}$$

Podane wyżej reguły należy traktować wyłącznie jako wskazówkę orientacyjną, twórca sieci nie musi się do nich sztywno stosować.

2.5. Trening sieci

Trening sieci neuronowych jest procesem optymalizacji parametrów sieci, prowadzonym tak aby osiągnąć żądany sygnał na wyjściu sieci po pobudzeniu sieci zadanym sygnałem. Ogólnie metody treningu sieci można podzielić na dwie grupy: uczenie z nauczycielem (gdy znane są oczekiwane odpowiedzi neuronów) oraz uczenie bez nauczyciela zwane też samouczeniem.

2.5.1. Uczenie z nauczycielem

2.5.1.1. Reguła Delta

Najbardziej popularnym algorytmem uczenia jednowarstwowych sieci liniowych jest, wprowadzony przez Widrowa i Hoffa [Wid60], algorytm zwany regułą Delta. Algorytm ten zakłada, że z każdym wektorem wejściowym $\mathbf{x}$ do neuronu podawany jest sygnał z, będący oczekiwaną (wymaganą) odpowiedzią neuronu na sygnał $\mathbf{x}$. Na podstawie odpowiedzi neuronu i oczekiwanej odpowiedzi neuronu obliczany jest błąd jaki popełnił neuron:

$$\delta = z - y \tag{2.15}$$

Celem uczenia jest takie dopasowanie wektora wag $\mathbf{w}$, aby uzyskać zgodność odpowiedzi neuronu y z wymaganymi wartościami z. Zadanie to można sprowadzić do minimalizacji funkcjonału błędu średniokwadratowego:

$$Q = \frac{1}{2}\sum_{i=1}^{L}(z_i - y_i)^2 \tag{2.16}$$

gdzie: L – rozmiar ciągu uczącego.

Stosując metodę gradientową można wykazać [Tad93], że formuła uczenia pojedynczego neuronu ma postać:

$$\mathbf{w}' = \mathbf{w} + \eta\delta\,\mathbf{x} \tag{2.17}$$

gdzie: η – współczynnik uczenia.

Przedstawiona reguła Delta wykorzystywana jest także w trakcie uczenia sieci nieliniowych. Jedynym warunkiem jaki musi być spełniony, aby móc stosować powyższą regułę do treningu neuronów z nieliniowymi funkcjami aktywacji $\varphi(e)$, jest to aby funkcje aktywacji były różniczkowalne. W takim przypadku formuła uczenia pojedynczego neuronu przyjmuje postać:

$$\mathbf{w}' = \mathbf{w} + \eta\delta\,\frac{d\varphi(e)}{de}\,\mathbf{x} \tag{2.18}$$

gdzie: $\varphi(e)$ – funkcja aktywacji neuronu.

2.5.1.2. Algorytm wstecznej propagacji błędów

Opisany wyżej algorytm uczenia jest możliwy do bezpośredniego zastosowania jedynie w przypadku sieci jednowarstwowych. Dla sieci wielowarstwowych (które mają

istotnie szersze możliwości przetwarzania informacji) wzór (2.18) nie daje się zastosować, ponieważ dla wewnętrznych warstw sieci nie ma możliwości bezpośredniego określenia błędu δ. Rozwiązaniem powyższego problemu jest zaproponowany w połowie lat 80. algorytm wstecznej propagacji błędów (*backpropagation*) [Rum86, Wei90]. Idea tego algorytmu polega na tym, iż mając wyznaczony błąd popełniany przez wyjściową warstwę sieci (obliczony na podstawie ciągu uczącego) można ten błąd „rzutować" wstecz na kolejne warstwy. Regułą uczenia neuronów warstwy wyjściowej jest taka sama jak w przypadku algorytmu Delta, natomiast nauka pozostałych neuronów odbywa się zgodnie z następującą regułą:

$$\mathbf{w}' = \mathbf{w} + \eta\delta_m \frac{d\varphi(e)}{de}\mathbf{y}_k \tag{2.19}$$

gdzie: δ_m – błąd popełniany przez neuron,

$\mathbf{y}_k$ – sygnały wejściowe warstwy k.

Wartość błędu popełnianego przez neuron można obliczyć poprzez wsteczne rzutowanie błędów wykrytych w warstwie odbierającej sygnały od rozważanego neuronu:

$$\delta_m = \sum_{k\in M} w_{mk}\delta_k \tag{2.20}$$

gdzie: w_{mk} – wagi łączące rozważany neuron z neuronami, dla których już wyznaczono błąd δ_k,

M – zbiór numerów neuronów, do których aktualnie rozważany neuron wysyła swój sygnał wyjściowy.

Stosując opisaną wyżej technikę można stopniowo nauczyć całą sieć, ponieważ każdy neuron albo znajduje się w warstwie wyjściowej (wtedy jego błąd może być wyznaczony bezpośrednio na podstawie ciągu uczącego) albo znajduje się w jednej z głębiej ukrytych warstw – wtedy jego błąd można oszacować z chwilą obliczenia błędów w neuronach będących odbiorcami jego sygnałów.

2.5.2. Uczenie bez nauczyciela

Opisane w poprzednim paragrafie metody treningu sieci opierały się na założeniu, że dla każdego wektora z ciągu uczącego istnieje pewna oczekiwana wartość odpowiedzi sieci. Jednak w niektórych sytuacjach informacja taka jest niedostępna i dlatego w literaturze przedmiotu wiele uwagi poświęca się metodzie samouczenia zaobserwowanej przez Hebba w funkcjonowaniu żywej tkanki nerwowej i dopracowanej potemprzez

Suttona dla potrzeb uczenia systemów technicznych. Mowa o tak zwanej technice uczenia bez nauczyciela, zwanej *unsupervised learning* lub *hebbian learning*. Zasada tego uczenia (stosowanego do treningu neuronów z liniową funkcją aktywacji) polega na tym, że waga w_i i-tego wejścia neuronu wzrasta podczas prezentacji wektora wejściowego $\mathbf{x}$ proporcjonalnie do iloczynu i-tej składowej sygnału wejściowego x_i docierającego do rozważanej synapsy i sygnału wyjściowego y rozważanego neuronu:

$$w_i' = w_i + \eta\, x_i y_i \tag{2.21}$$

gdzie: η – współczynnik uczenia.

Proces samouczenia można uczynić bardziej efektywnym poprzez zastosowanie tek zwanego przyrostowego samouczenia (*differential hebbian learning*), polegającego na uzależnieniu zmiany wag od przyrostów sygnałów wejściowych na danej synapsie i wyjściowych na danym neuronie (górne indeksy oznaczają numery iteracji w procesie treningu):

$$w_i' = w_i + \eta \left[\left(x_i^{(n)} - x_i^{(n-1)}\right)\left(y_i^{(n)} - y_i^{(n-1)}\right)\right] \tag{2.22}$$

Inna modyfikacja algorytmu samouczenia pochodzi z wczesnych prac Grossberga [Gro76] i znana jest w literaturze jako koncepcja „gwiazdy wejść" (*instar training*). W koncepcji tej wybiera się pewien neuron (oznaczony za pomocą gwiazdki) i narzuca mu się taką strategię uczenia, by zapamiętał i potrafił potem rozpoznać aktualnie wprowadzany sygnał $\mathbf{x}$:

$$w_i^{*\,'} = w_i^* + \eta \left(x_i - w_i^*\right) \tag{2.23}$$

Na zasadzie analogii do *instar training* można wprowadzić także drugą propozycję Grossberga – koncepcję *outstar training*, w której rozważa się wagi wszystkich neuronów całej warstwy, jednak wybiera się wyłącznie wagi łączące te neurony z pewnym ustalonym wejściem. W sieciach wielowarstwowych wejście to pochodzi od pewnego ustalonego neuronu wcześniejszej warstwy i to właśnie ten neuron staje się centralnym punktem „gwiazdy wyjść". Powyższą zasadę można zapisać w postaci wzoru:

$$w_i' = w_i + \eta \left(y - w_i\right) \tag{2.24}$$

Warto jeszcze podkreślić odmienność powyższego wzoru w stosunku do poprzednio rozważanych. We wzorze (2.24) indeks i jest ustalony, natomiast formuła ta obliczana jest dla wszystkich możliwych neuronów w danej warstwie sieci.

2.6. Dobór wartości współczynnika uczenia

Istotnym czynnikiem z punktu widzenia szybkości i zbieżności procesu treningu jest odpowiedni dobór współczynnika uczenia η. W trakcie treningu sieci następuje zmiana współczynników wagowych poszczególnych neuronów – współczynnik η decyduje o wielkości tych zmian. Zbyt mała zmiana wag (mała wartość współczynnika η) wydłuża proces treningu (wymagana jest większa ilość iteracji aby odpowiednio dopasować wagi). Z drugiej strony zbyt duża zmiana wag może spowodować zbyt gwałtowne zmiany w sieci i w ich następstwie niemożność dopasowania wag do optymalnych wartości. Ogólnie istnieją dwie metody doboru współczynnika uczenia:

- współczynnik η jest stały,
- współczynnik η jest zmieniany w trakcie uczenia sieci.

W pierwszym przypadku wartość η wybierana jest arbitralnie, jedną z metod oszacowania tej wielkości może być zależność [Oso96]:

$$\eta \cong \left(\frac{1}{N}\right) \tag{2.25}$$

gdzie: N – całkowita liczba wag w danej warstwie sieci.

W przypadku metod ze zmiennym współczynnikiem uczenia stosuje się różne strategie w zależności od rodzaju sieci. W sieciach wykorzystujących koncepcję „gwiazdy wejść" (porównaj rozdział 2.5.2), można zastosować empiryczną regułę [Tad93]:

$$\eta^{(n)} = 0.1 - \lambda \cdot n \tag{2.26}$$

gdzie: λ – stały współczynnik (na tyle mały aby $\eta > 0$ w trakcie całego procesu treningu) .

Podobnie w sieciach stosujących koncepcję „gwiazdy wyjść", reguła zmiany wartości η przybiera postać:

$$\eta^{(n)} = 1 - \lambda \cdot n \tag{2.27}$$

W sieciach trenowanych przy pomocy reguły Delta proponuje się niekiedy jedną z dwóch poniższych formuł:

$$\eta^{(n)} = \left(\sum_{i=1}^{n} \|\mathbf{x}_i\|^2\right)^{-1} \tag{2.28}$$

$$\eta^{(n)} = \frac{\lambda}{\left\|\mathbf{x}_n\right\|^2} \tag{2.29}$$

gdzie: $\mathbf{x}_i$ – i-ty wektor z ciągu uczącego,

λ – stały współczynnik (zwykle $0.1 < \lambda < 1$).

Alternatywnym podejściem może być zmiana współczynnika uczenia na podstawie zmiany wartości błędu w kolejnych krokach treningu [Oso96]. Błąd popełniany przez sieć może być obliczony jako:

$$\varepsilon = \sum_{i=1}^{M} (y_i - z_i)^2 \tag{2.30}$$

gdzie: M – liczba neuronów wyjściowych w sieci,

y_i – sygnał wyjściowy i-tego neuronu,

z_i – oczekiwany sygnał wyjściowy i-tego neuronu.

Gdy w kolejnych iteracjach błąd wzrasta zbyt szybko [Oso96]:

$$\varepsilon^{(n)} > k_w \cdot \varepsilon^{(n-1)} \tag{2.31}$$

gdzie: k_w – współczynnik decydujący o dopuszczalnym tempie wzrostu błędu ε,

to współczynnik η jest zmniejszany w sposób liniowy zależnością:

$$\eta^{(n)} = \rho_d \cdot \eta^{(n-1)} \tag{2.32}$$

gdzie: ρ_d – współczynnik zmniejszający wartość η.

W przeciwnym wypadku, gdy nierówność (2.31) nie jest spełniona, wartość współczynnika η jest zwiększana liniową zależnością:

$$\eta^{(n)} = \rho_i \cdot \eta^{(n-1)} \tag{2.33}$$

gdzie: ρ_i – współczynnik zwiększający wartość η.

Przykładowe wartości współczynników wynoszą: $k_w = 1.04$, $\rho_d = 0.7$, $\rho_i = 1.05$.

2.7. Liczba warstw i neuronów w sieci

Liczba warstw w sieci ma istotny wpływ na możliwości przetwarzania informacji przez sieć. Możliwości klasyfikacyjne jednej warstwy neuronów ograniczone są do wzorców liniowo separowalnych. Jeśli nie istnieje płaszczyzna dzieląca wzorce to po-

blem nie jest liniowo separowalny i zadanie klasyfikacji nie może być spełnione przez pojedynczą warstwę neuronów. Dodanie kolejnych warstw pozwala na zwiększenie zakresu odwzorowań realizowanych przez sieć. Dowolne odwzorowanie może być zrealizowane przez sieć o trzech warstwach, pod warunkiem dostatecznie dużej liczby neuronów w warstwach [Cyb89, Her93, Tad93].

W trakcie rozważań dotyczących projektowania i uczenia sieci przyjmuje się zwykle założenie, że warstwa wejściowa i wyjściowa sieci jest ustalona; to znaczy, że z góry wiadomo, jakie zmienne stanowić będą wejścia do sieci i jakie są w niej oczekiwane wartości wyjściowe. Istotnie można się zgodzić, że te drugie są zawsze znane, gdyż samo sformułowanie zadania dla sieci jednoznacznie wskazuje, czego od niej oczekujemy, a z tego w oczywisty sposób wynika, ile i jakich zmiennych wyjściowych należy użyć. Nieco ostrożniej można tę tezę sformułować w taki sposób, że przynajmniej w tych zagadnieniach, w których stosuje się uczenie z nauczycielem, nie powinno być wątpliwości, jakie sygnały pojawiają się na wyjściu sieci – ponieważ nauczyciel musi podawać dla tych sygnałów wyjściowych wzorce ich poprawnych wartości.

Natomiast wybór zmiennych wejściowych dla sieci jest o wiele trudniejszy, a sformułowane zadanie często pozostawia w tym zakresie sporą dowolność. Istota problemu polega na tym, że budując sieć często nie wiadomo, jaki zestaw (spośród wielu kandydujących zmiennych wejściowych) jest w rzeczywistości użyteczny dla skutecznego wyznaczania przez sieć potrzebnej wartości wyjściowej. Wybór dobrego zbioru wejść jest dodatkowo utrudniony z uwagi na konieczność rozważenia licznych istotnych przesłanek i problemów, które teraz kolejno zostaną wymienione.

Problem wymiarowości. Każdy dodatkowy neuron wejściowy w sieci powoduje zwiększenie wymiaru przestrzeni, w której znajdują się przypadki danych. Komplikuje to strukturę sieci, a także stwarza większe wymagania odnośnie zbiorów uczących. Należy bowiem pamiętać o konieczności zapewnienia dostatecznej liczby punktów uczących, które w wystarczającym stopniu wypełnią N-wymiarową przestrzeń sygnałów wejściowych. Musi być tych punktów uczących na tyle dużo, żeby wystarczająco gęsto wypełniły przestrzeń wejść, tak aby możliwe było „zobaczenie” istniejącej struktury danych i wierne odwzorowanie w sieci jej właściwości. Liczba punktów niezbędnych do tego rośnie bardzo szybko wraz z wymiarem przestrzeni sygnałów wejściowych, co stanowi istotny problem. Szacunkowo można podać, że liczba punktów (elementów zbioru danych) potrzebnych do w miarę wiernego odwzorowania zasadniczych cech tworzonego modelu jest rzędu 2^N. Oznacza to, że dla sieci neuronowych (podobnie jak dla większości innych technik modelowania) nie można bezkarnie powiększać liczeb-

ności wejściowego wektora danych, gdyż trzeba wtedy mieć bardzo wiele obserwacji, aby uzyskać w miarę reprezentatywne dane.

Podane wyżej uwagi można czasem nieco złagodzić. Liczne eksperymenty dowodzą, że większość typów sieci neuronowych (w szczególności sieci MLP) w rzeczywistości jest w mniejszym stopniu podatna na problemy związane z wymiarem przestrzeni, niżby to wynikało z przytoczonych wyżej teoretycznych rozważań. Wynika to z zaobserwowanego wielokrotnie faktu, że sieci neuronowe w trakcie uczenia potrafią od razu na początku „skoncentrować się" na pewnym podzbiorze przestrzeni wejściowej, ignorując resztę wejściowych danych. Oznacza to, że sieć umie w trakcie uczenia skupić się na spontanicznie utworzonym „zadaniu zastępczym", charakteryzującym się mniejszym wymiarem przestrzeni sygnałów wejściowych. W takiej „wykrojonej" podprzestrzeni (będącej często bardzo małym fragmentem oryginalnej wysokowymiarowej przestrzeni wejść) sieć neuronowa potrafi zaskakująco dobrze odwzorować specyfikę rozwiązywanego zadania. Jest to unikatowa, ale bardzo korzystna cecha sieci neuronowych – przynajmniej w porównaniu z innymi metodami modelowania matematycznego (na przykład w porównaniu z metodami statystycznymi). Wspomniane „przycinanie" wejściowej przestrzeni odbywa się w sieci neuronowej bardzo prosto: poprzez bardzo szybkie wyzerowanie wag wychodzących od wybranego wejścia sieć MLP może już na początku uczenia całkowicie zignorować daną zmienną wejściową. Oznacza to, że w trakcie uczenia sieci może dochodzić równocześnie do dwóch procesów: optymalizacji funkcji przenoszenia całej sieci, będącej modelem poszukiwanej zależności, oraz spontanicznej selekcji zmiennych wejściowych z podziałem ich na takie, które będą argumentami rozważanej funkcji, oraz na takie, które będą odrzucone (zignorowane).

Niemniej wspomnianego wyżej mechanizmu nie należy nadużywać i dlatego trzeba przyznać, że problem wymiaru wejściowych danych dla sieci neuronowych nadal istnieje, a działanie wielu sieci może być z pewnością znacznie polepszone poprzez eliminację (od samego początku) zbędnych zmiennych wejściowych.

W związku z tym, że precyzyjna ocena stopnia informatywności poszczególnych danych wejściowych bywa bardzo trudna, nie zawsze wiadomo, które zmienne są zbędne, a które nie. W związku z tym, w celu ograniczenia problemów związanych z wymiarem przestrzeni wejść, czasami wskazane jest usunięcie nawet takich zmiennych, które zapewne wnoszą pewną ilość informacji, jednak ich udział w budowanym modelu jest niewielki.

Współzależności pomiędzy zmiennymi. Gdyby istniała możliwość niezależnej oceny użyteczności każdej potencjalnej zmiennej wejściowej, to wówczas stosunkowo łatwo byłoby wybrać na wejścia sieci tylko te spośród nich, które są najbardziej przy-

datne. Niestety, taka możliwość istnieje rzadko, gdyż zwykle istotne dodatkowe informacje mieszczą się we współwystępowaniu zmiennych.

Chodzi mianowicie o to, że dwie (lub większa liczba) współzależnych zmiennych mogą razem dostarczać istotnych informacji, podczas gdy każdy podzbiór takiego zestawu zmiennych informacji takich już nie zawiera. Klasycznym przykładem zadania, w którym współwystępowanie zmiennych wnosi informacje niedostępne w inny sposób może być szeroko znany problem dwóch spiral. Zadanie to polega na zaliczaniu wejściowych danych do jednej z dwóch klas. Te dwie klasy rozpoznawanych danych tworzą w dwuwymiarowej przestrzeni wejściowej dwie splecione ze sobą spirale. Łatwo zauważyć, że w tym zadaniu żadna ze zmiennych wejściowych samodzielnie nie dostarcza użytecznych informacji (wzdłuż każdej z osi rozważane dwie klasy wyglądają na całkowicie przemieszane), ale korzystając łącznie z obu zmiennych można dokładnie rozróżnić obie klasy. Przytoczony przykład wskazuje na to, że zmienne wejściowe nie mogą być zwykle oceniane ani wybierane całkiem niezależnie, lecz powinny być niekiedy oceniane w pewnych grupach, wynikających z natury rozwiązywanego problemu. Na marginesie można dodać, że wspomniany klasyczny problem dwóch spiral ilustruje jeszcze jedną, dość uniwersalną, ale i dość kłopotliwą cechę sieci neuronowych. Otóż problem ten jest bardzo trudny do rozwiązania jeśli sieć otrzymuje dane o stawianym jej zadaniu w postaci współrzędnych kartezjańskich poszczególnych punktów oraz wzorca informacji wyjściowej, podającej, do której z dwóch klas dany punkt należy. Natomiast zadanie staje się banalnie łatwe jeśli wejściowe dane (współrzędne punktu) podawane są w postaci współrzędnych biegunowych. Świadczy to o celowości wytrwałego poszukiwania najlepszego zestawu danych wejściowych, gdyż oczywiście w większości praktycznych przypadków wybór lepszych danych wejściowych nie jest tak łatwy ani naturalny, jak w przypadku pomysłu z użyciem współrzędnych biegunowych w problemie dwóch spiral, ale zwykle po wybraniu dobrego zestawu danych można naprawdę wiele skorzystać.

Redundancja zmiennych. Często w zadaniach rozwiązywanych z pomocą sieci neuronowych zdarza się, że liczne zmienne mogą dostarczać na różne sposoby tych samych informacji. Na przykład wzrost i waga ludzi mogą, w wielu okolicznościach, dostarczać podobnych informacji, ponieważ te dwie zmienne są zwykle ze sobą silnie skorelowane. W przypadku, kiedy zmienne wejściowe są silnie skorelowane wystarczy użyć w charakterze wejść pewnego ich podzbioru (czyli wybrać spośród zmiennych skorelowanych ich odpowiednią reprezentację, tak zbudowaną, by zmienne zachowane w tej reprezentacji już dalej nie były wzajemnie skorelowane). W podanym wyżej przykładzie oznacza to, że do danych wejściowych włączona powinna być albo informacja o

wzroście, albo informacja o wadze rozważanego człowieka, nie celowe jest natomiast podawanie obydwu tych skorelowanych danych (oczywiście nie dotyczy to tych zadań, kiedy subtelne zmiany proporcji wspomnianych skorelowanych danych są najbardziej istotną informacją – na przykład w zagadnieniach leczenia otyłości).

W opisanym przypadku ważne jest usunięcie nadmiarowych zmiennych, ale całkowicie dowolny jest wybór zachowywanego podzbioru (czyli decyzja o tym, które zmienne usunąć, a które pozostawić). Wybór ten może być dyktowany jakimiś dodatkowymi czynnikami – na przykład faktem, że jedne zmienne wejściowe są łatwiejsze do uzyskania, niż inne. Przewaga stosowania podzbioruskorelowanych zmiennych zamiast korzystania z pełnego ich zbioru jest oczywistą konsekwencją dyskutowanego wyżej problemu wymiaru wejściowego wektora danych.

Dokonując opisanej w poprzednim punkcie redukcji danych musimy być pewni, że w rozważanym problemie nie pojawia się merytoryczna potrzeba uwzględniania (jako dodatkowej ważnej informacji) wzajemnych relacji lub proporcji skorelowanych zmiennych. Jak już nadmieniono, w rozważanym przykładzie z wagą i wzrostem może się okazać, że uwzględnienie zarówno wagi, jak i wzrostu, jest celowe mimo ich skorelowania. Będzie się tak działo w przypadku, gdy w procesie rozpoznawania ważną informacją może być zarówno nadmierne wychudzenie lub (przeciwnie) nadmierna otyłość. Gdyby przeznaczeniem budowanej sieci miało być jedno z zadań diagnostyki medycznej, związane z badaniami metabolizmu, wówczas z pewnością informacja o proporcji wagi do wzrostu będzie ważną przesłanką dla wymaganych decyzji. Jednak nawet w takim przypadku nie jest na ogół celowe zwyczajne podawanie obok siebie obu skorelowanych danych (np. wzrostu i wagi), bo wtedy sieć może się koncentrować na tym, że jedna z tych wartości daje się przewidywać na podstawie drugiej z pomocą prostej regresji, a nie o to w tym przypadku chodzi. Dlatego mając dane wejściowe silnie skorelowane i mając równocześnie świadomość konieczności uwzględniania obydwu skorelowanych informacji – lepiej jest podać jako wejście sieci jedną ze skorelowanych wartości oraz – na przykład – stosunek drugiej wartości do tej pierwszej. Taka wartość względna (na przykład iloraz wagi i wzrostu) nie jest skorelowana, więc unika się paradoksów w działaniu sieci, a jednocześnie odciąża się sieć od konieczności „odkrywania" w trakcie uczenia, że taka właśnie proporcja stanowi ważną przesłankę przy podejmowaniu wymaganej decyzji.

Powyższe uwagi wskazują, że wybór zmiennych wejściowych jest krytycznym elementem projektowania sieci neuronowej. Aby dokonać optymalnego wyboru zmiennych wejściowych przed użyciem sieci można skorzystać z własnej wiedzy lub zasięgnąć opinii ekspertów z zakresu rozwiązywanego problemu. Można także użyć standar-

dowych testów statystycznych dla uzyskania dokładniejszej informacji o naturze posiadanych danych. Warto wiedzieć, że istnieje sporo narzędzi, przy pomocy których można przetestować różne kombinacje zmiennych wejściowych. Na przykład można wykorzystać możliwość „nieuwzględniania" pewnychzmiennych i podejmować wysiłek budowania próbnych sieci, które nie wykorzystują tych zmiennych jako wejść. Można wielokrotnie eksperymentalnie dodawać i usuwać różne kombinacje danych, budując za każdym razem nową sieć i sprawdzając jej działanie.

W trakcie takich eksperymentów szczególnie przydatne są sieci probabilistyczne i sieci realizujące regresję uogólnioną. Chociaż w porównaniu z bardziej zwartymi sieciami MLP oraz RBF są one powolne w działaniu, ale za to ich uczenie jest prawie natychmiastowe – co jest bardzo istotne, gdyż w trakcie testowania dużej liczby kombinacji zmiennych wejściowych, proces budowy i uczenia sieci musi być wielokrotnie powtarzany. Co więcej, sieci PNN oraz GRNN są (podobnie jak sieci RBF) przykładami sieci radialnych. Oznacza to, że posiadają one neurony radialne w pierwszej warstwie ukrytej i budują funkcje będące kombinacjami funkcji Gaussa. Jest to dużą zaletą w trakcie doboru zmiennych wejściowych, ponieważ sieci radialne są w rzeczywistości bardziej podatne na problem wymiaru niż sieci liniowe, więc kontrola danych odbywa się w tym przypadku przy użyciu wyjątkowo wymagającego testu.

Aby wyjaśnić powyższe stwierdzenie należy rozpatrzyć efekt, jaki wywoła dodanie do sieci dodatkowej, wybranej w sposób niepoprawny zmiennej wejściowej. Sieć liniowa, taka jak MLP, może „ukryć" ten błędny wybór zmiennej wejściowej w taki sposób, że w trakcie uczenia może nadać wszystkim wagom wychodzącym od błędnego wejścia wartość zero. W praktyce takie „izolowanie" niektórych zmiennych wejściowych dosyć często zachodzi (w sposób całkowicie nie zauważalny dla użytkownika), ponieważ jest bardzo łatwe do realizacji. Po prostu pewne wagi, o niewielkich wartościach początkowych (jak wszystkie wagi po procesie losowej inicjalizacji sieci) pozostają po procesie uczenia nadal małe, zaś inne wartości wag (te wychodzących od wejść niosących istotną informację) będą się w trakcie uczenia systematycznie odchylać od niewielkich wartości początkowych w kierunku określonych wartości końcowych – dodatnich albo ujemnych.

Tak bywa w sieciach typu MLP. W odróżnieniu od tego sieci radialne, takie jak PNN lub GRNN, nie dostarczają takiego luksusu: generowane w neuronach radialnych skupienia, które miałyby określony kształt w przestrzeni zmiennych istotnych, posiadającej niższy wymiar, zostają istotnie zaburzone i zniekształcone przez pojawiające się w danych wejściowych nieistotne składowe, nadające przestrzeni, w której budowane są skupienia, niepotrzebnie i niekorzystnie większy wymiar. Pokrycie przestrzeni o więk-

szej wymiarowości wymaga wtedy większej liczby neuronów, aby uwzględnić (nieistotną, ale pojawiającą się w danych) zmienność generowaną przez „nadmiarowe" wymiary. Ujawnia się to natychmiast w postaci złych wyników uzyskiwanych przez sieć, po usunięciu nadmiarowych zmiennych jakość działania sieci radykalnie się poprawia.

Możliwość dokonania optymalizacji przestrzeni wejść poznaje się więc na tej podstawie, że „próbna" sieć, która gorzej się zachowuje w przypadku korzystania z całej zbiorowości sygnałów wejściowych (w tym także z tych niepoprawnie dobranych, ale trudnych do wykrycia zmiennych wejściowych) radykalnie polepsza swoje działanie przy próbie eliminacji pewnych wejść (które na tej podstawie mogą być zidentyfikowane jako błędnie dobrane czy wręcz niepotrzebne).

Ponieważ takie eksperymenty są bardzo czasochłonne (trzeba wielokrotnie ponawiać próby uczenia i oceny sieci, eliminując po kolei różne zmienne i ich kombinacje), opracowano również mechanizm umożliwiający ich przeprowadzenie w sposób automatyczny, bez angażowania uwagi i wysiłku użytkownika. Mechanizm ten wykorzystuje do wyboru właściwej kombinacji wejść tak zwany algorytm genetyczny.

Algorytmy genetyczne są techniką globalnej optymalizacji, bardzo dobrze się sprawdzającą przy takich problemach jak tu opisywane, gdyż posiadają zdolność przeszukania dużej liczby kombinacji (w tym przypadku – zestawów zmiennych wejściowych) w celu znalezienia najlepszego rozwiązania. Stosowanie algorytmów genetycznych jest szczególnie korzystne w sytuacji, gdy mogą istnieć współzależności pomiędzy optymalizowanymi zmiennymi i gdy przy ich eliminacji trzeba brać pod uwagę wpływ ich wzajemnych interakcji.

Innym podejściem do rozwiązania problemu redukcji wymiaru przestrzeni sygnałów wejściowych, które stanowić może alternatywę lub uzupełnienie w stosunku do opisanych wyżej metod wyboru zmiennych, jest redukcja wymiaru przestrzeni wejść metodą zastosowania odpowiedniej transformacji. W trakcie transformacji pierwotny zbiór zmiennych jest przetwarzany w celu utworzenia nowego, (mniejszego) zbioru zmiennych, który zawiera maksymalnie dużo informacji zawartej w zbiorze pierwotnym.

Jako przykład danych dobrze nadających się do użycia takiej transformacji rozważyć można taki zbiór danych, w którym wszystkie punkty leżą na pewnej płaszczyźnie, umieszczonej skośnie w przestrzeni trójwymiarowej. W takim przypadku oryginalne kodowanie danych wymaga, żeby każda z nich miała trzy składowe, podczas gdy faktyczny wymiar przestrzeni danych wynosi dwa. Jeśli płaszczyzna ta zostanie wy-

kryta i gdy zostanie zdefiniowana transformacja rzutująca całą przestrzeń na tę jedynie ważną płaszczyznę, to sieć neuronowa może posiadać mniejszy wymiar danych wejściowych, dzięki czemu posiada większą szansę na poprawną pracę.

Najpopularniejszym sposobem redukcji wymiaru przestrzeni sygnałów wejściowych jest analiza głównych składowych PCA(*Principal Components Analysis*). Jest to transformacja liniowa, która wyznacza kierunki maksymalnej zmienności pierwotnych danych wejściowych i dokonuje rotacji układu współrzędnych w taki sposób, żeby maksymalna zmienność danych zachodziła po transformacji wzdłuż tych nowych, obróconych osi. Zwykle dokonuje się przy tym takiego uporządkowania numeracji wykrytych nowych osi (składowych głównych) żeby początkowe (mające najniższe numery) główne składowe odpowiadały kierunkom największej zmienności danych, czyli (jak się przypuszcza) zawierały najwięcej informacji. Można wtedy oprzeć działanie sieci jedynie na kilku początkowych składowych głównych, co radykalnie zmniejsza wysiłek związany z operowaniem wejściowym zbiorem sygnałów.

Ponieważ analiza głównych składowych jest matematycznie opisywana transformacją liniową, więc może być wykonywana przez liniową sieć neuronową. Możliwość ta jest chętnie stosowana, ponieważ często PCA jest w stanie wyodrębnić z ogromnej liczby danych wejściowych pewną bardzo małą liczbę składowych głównych o tak dużej zawartości informacyjnej, że wystarczą one do skutecznego opisania badanej zależności. W takim przypadku wystarcza podanie tylko tych kilku („wyodrębnionych" przez osobną liniową sieć neuronową) najważniejszych składowych głównych, na wejście nieliniowej sieci neuronowej modelującej badaną zależność. Taka kombinacja sieci liniowej, dokonującej PCA i małej (ze względu na małą liczbę wejść) sieci modelującej, z powodzeniem może zastąpić podawanie wszystkich oryginalnych danych wejściowych do dużej sieci modelującej badaną zależność regresyjną.

W przypadku szerokiego stosowania PCA pewnym ograniczeniem może być liniowy charakter tej techniki, który może w niektórych przypadkach powodować utratę ważnych informacji na temat struktury pierwotnych danych. Żeby uniknąć tej niedogodności można również realizować pewien rodzaj „nieliniowej PCA" stosując w tym celu sieć autoasocjacyjną.

Koncepcję takiej sieci zgłosili Bouland i Kamp w 1988 roku. Jest to sieć neuronowa, która jest uczona w taki sposób, aby dokonywała maksymalnie wiernej reprodukcji wartości wejściowych na swoich wyjściach. Rozważana sieć posiada w warstwie środkowej (ukrytej) zdecydowanie mniejszą liczbą neuronów niż w warstwie wejściowej czy wyjściowej. Przez to „ucho igielne" muszą przecisnąć się wszystkie dane przesyłane z wejścia na wyjście. Ponieważ w sieci autoasocjacyjnej nie ma żadnych bezpo-

średnich połączeń między warstwą wejściową a wyjściową – jedynym sposobem wiernego odtworzenia sygnałów wejściowych na wyjściu jest właśnie znalezienie sposobu kompresji sygnału między warstwą wejściową a ukrytą, a także sposobu rekonstrukcji sygnału między warstwą ukrytą a warstwą wyjściową. Osiągnięcie stanu prawidłowej pracy sieci po okresie uczenia oznacza, że sieć autoasocjacyjna w trakcie uczenia zdobyła umiejętność redukcji wymiaru wejściowych danych.

Po zakończeniu uczenia część sieci od strony wyjść może zostać usunięta, tak aby pozostała część realizowała potrzebną (nieliniową) redukcję wymiaru danych wejściowych. Tak spreparowane dane mogą trafiać na wejścia sieci modelującej interesującą zależność, gwarantując wszystkie jej wyżej dyskutowane zalety, związane z możliwością poruszania się w zbiorze danych o mniejszej wymiarowości.

W wielu przypadkach zadanie nieliniowej kompresji danych wejściowych (i nieliniowej ich dekompresji na wyjściu) jest zadaniem zbyt złożonym, by mogło być ono wykonane z użyciem sieci o jednej warstwie ukrytej. Z tego powodu sieć autoasocjacyjna jest najczęściej siecią MLP posiadającą trzy warstwy ukryte, z których środkowa wyznacza reprezentację o zredukowanym wymiarze. Dwie pozostałe warstwy ukryte są potrzebne do przeprowadzenia przez sieć dwóch skomplikowanych transformacji nieliniowych: transformacji wejść do postaci skompresowanej dostępnej w środkowej warstwie ukrytej oraz odpowiednio transformacji danych pojawiających się w środkowej warstwie ukrytej do postaci wyjściowej. Sieć autoasocjacyjna posiadająca tylko jedną warstwę ukrytą może wyłącznie realizować liniową redukcję wymiaru i w rzeczywistości uczy się aproksymacji standardowej PCA.

Należy również zaznaczyć, że liczba neuronów w sieci ma zasadnicze znaczenie w przypadku zadań związanych z rozpoznawaniem i klasyfikacją wzorców. Jednym z ograniczeń zastosowań sieci może być „pojemność pamięci", bezpośrednio związana z liczbą neuronów w sieci. Przyjmuje się, że dla sieci liczącej k neuronów maksymalna liczba możliwych do zapamiętania wzorów wyraża się przybliżonym empirycznym wzorem [Tad93]:

$$P_{\max} \approx \frac{k}{2 \log k} \tag{2.34}$$

2.8. Sieci Kohonena

Sieci Kohonena realizują uczenie w trybie bez nauczyciela, w którym stosuje się tak zwane uczenie z rywalizacją (*competitive learning*) [Koh84, Koh90a, Koh90b].

Wprowadził je Kohonen przy tworzeniu sieci neuronowych realizujących dowolne odwzorowanie $\mathbf{X} \Rightarrow \mathbf{Y}$. Sieci Kohonena uczą się **rozpoznawania skupień** występujących w zbiorze nieskategoryzowanych danych uczących. Z góry nie wiadomo, ani ile takich skupień będzie, ani czym się od siebie różnią, ani jakie są kryteria przynależności tego czy innego obiektu do określonego skupienia. Sieć Kohonena uczy się całkiem sama, wyłącznie obserwując przesyłane do niej dane, których wewnętrzna struktura i ukryta w tej strukturze nieznana logika – decydować będzie o końcowym obrazie klasyfikacji i o ostatecznym działaniu sieci. W związku z brakiem zastosowania w takiej sieci wzorcowych danych wyjściowych proces samouczenia sieci Kohonena prowadzony jest na podstawie danych uczących, które zawierają same tylko wartości zmiennych wejściowych.

Można założyć, że w sieci Kohonena poszczególne neurony identyfikują i rozpoznają poszczególne skupienia danych. Przez skupienie danych rozumieć należy grupę danych cechujących się tym, że dane należące do skupienia są (pomiędzy sobą) w jakimś stopniu podobne, natomiast dane należące do różnych skupień w jakimś stopniu różnią się pomiędzy sobą. W sieci Kohonena przebiega dość specyficzny proces samouczenia nazywany także często samoorganizacją. W wyniku tej samoorganizacji zbliżone do siebie skupienia danych reprezentowane są (po zakończeniu procesu nauki) przez neurony warstwy wyjściowej położone blisko siebie, co powoduje, że neurony te tworzą *mapę topologiczną* wejściowych danych. Ta właściwość sieci Kohonena bywa bardzo użyteczna przy analizie i ocenie tych danych. Wykazali to, między innymi, Haykin w 1994 roku i Fausett oraz Patterson w 1996.

Sieci Kohonena są wykorzystywane w całkowicie inny sposób niż inne rodzaje sieci, gdyż głównym wynikiem ich pracy jest utworzona przez nie mapa topologiczna wejściowych danych. Neurony w sieci Kohonena nie są ze sobą połączone, natomiast często dla wygody interpretacji wyników działania sieci zbiorowość neuronów wyjściowych przedstawiana jest w postaci dwuwymiarowej struktury (neurony uszeregowane są w kolumny, a te kolumny ułożone są jedna obok drugiej i tworzą łącznie regularną siatkę – neurony są jak gdyby węzłami tej siatki).

Zaproponowany przez Kohonena algorytm uczenia modyfikuje współczynniki wagowe w neuronach znajdujących się w warstwie stanowiącej mapę topologiczną w taki sposób, aby zbliżyć je do centrów skupień występujących w danych uczących. Odbywa się to w taki sposób, że w trakcie uczenia algorytm Kohonena najpierw wyznacza neuron mający współczynniki wagowe najbardziej zbliżone do aktualnie przetwarzanego przypadku uczącego. Neuron ten nazywany jest „zwycięzcą” i proces uczenia koncentruje się na takiej zmianie jego wartości wagowych by wzmocnić i utrwalić

skłonność, dzięki której neuron ten stał się „zwycięzcą". Zasada uczenia sieci Kohonena, jest bardzo podobna do reguły *instar* (porównaj rozdział 2.5.2):

$$w_i^{*'} = w_i^* + \eta\left(x_i - w_i^*\right) \tag{2.35}$$

z dwoma dość istotnymi uzupełnieniami. Po pierwsze wektor wejściowy $\mathbf{x}$ jest przed procesem uczenia normalizowany (tak aby $\left\|\mathbf{x}'\right\| = 1$):

$$x_i' = \frac{x_i}{\sqrt{\sum_{j=1}^{n}\left(x_j\right)^2}} \tag{2.36}$$

Po drugie, treningowi poddawany jest tylko „zwycięzca", tzn. ten neuron, którego sygnał wyjściowy y^* jest największy. Oznacza to, że przy każdorazowym podaniu sygnału wejściowego $\mathbf{x}$ neurony rywalizują ze sobą i tylko ten neuron, który wygra w danym momencie jest uczony. Efektem tego jest jeszcze lepsze dopasowanie wag tego „zwycięskiego" neuronu do rozpoznawania sygnałów podobnych do wektora $\mathbf{x}$.

2.8.1. Sąsiedztwo w sieci Kohonena

Reguła uczenia Kohonena bywa często wzbogacana o dodatkowy element związany z topologią sieci. Po uporządkowaniu neuronów w sieci (poprzez nadanie im numerów) można wprowadzić pojęcie sąsiedztwa – na przykład uznając za sąsiednie te neurony, które mają numery m różniące się o ustaloną wartość. Uogólniona reguła samoorganizacji sieci Kohonena polega na tym, że uczeniu podlega nie tylko neuron „zwycięzca", ale także neurony, które z nim sąsiadują. Formalnie regułę tą można zapisać w postaci wzoru (górny indeks oznacza numer neuronu):

$$w_i^{'m} = w_i^m + \eta\, h\left(m, m^*\right)\left(x_i - w_i^m\right) \tag{2.37}$$

gdzie: m – indeks rozważanego neuronu,

m^* – indeks „zwycięzcy",

$h\left(m, m^*\right)$ – funkcja określająca sąsiedztwo.

Funkcja określająca sąsiedztwo, zwana też współczynnikiem uczenia sąsiadów ponieważ determinuje stopień nauki sąsiadów, jest najbardziej popularna w postaci:

$$h\left(m, m^*\right) = \begin{cases} 1 & m = m^* \\ 0.5 & \left|m - m^*\right| = 1 \\ 0 & \left|m - m^*\right| > 1 \end{cases} \tag{2.38}$$

W bardziej skomplikowanych zadaniach funkcja ta może być wyrażona za pomocą odległości, na przykład:

$$h(m,m^*)=\frac{1}{\rho(m,m^*)} \tag{2.39}$$

$$h(m,m^*)=\exp\left(-\left[\rho(m,m^*)\right]^2\right) \tag{2.40}$$

gdzie: $\rho(m,m^*)$ – odległość między neuronami o indeksach m i m^*.

Każdy neuron wchodzący w skład sąsiedztwa zwycięskiego neuronu jest modyfikowany w taki sposób, aby jego zestaw wag upodobnić do przypadku uczącego, który wyłonił „zwycięzcę". Warto podkreślić, że w wielu przypadkach może to oznaczać diametralną zmianę wag neuronu należącego do sąsiedztwa, gdyż jego początkowy zestaw wag mógł być diametralnie odmienny od rozważanego wejściowego sygnału.

Sąsiedztwo odgrywa zasadniczą rolę w uczeniu sieci Kohonena. Poprzez dokonywanie zmian nie tylko w zwycięskim neuronie ale również w neuronach otaczających zwycięzcę, metoda uczenia Kohonena przypisuje powiązane ze sobą dane do sąsiadujących obszarów mapy topologicznej. W trakcie uczenia wielkość sąsiedztwa jest progresywnie zmniejszana, co powoduje, że zasięg wpływu zwycięskich neuronów na proces uczenia ich sąsiadów powoli zanika i pod koniec uczenia każdy neuron jest w istocie uczony niezależnie. Podobnie dzieje się ze współczynnikiem uczenia – na początku jest duży, co wymusza duże zmiany w położeniu wektora wag „zwycięzcy" i jego sąsiadów, potem jednak maleje i zmiany stają się subtelne, wręcz „kosmetyczne". W wyniku tych procesów we wczesnych etapach procesu uczenia tworzona jest mapa o charakterze przybliżonym (posiada ona zwykle duże skupiska neuronów, które odpowiadają na podobne przypadki), natomiast precyzyjne detale zostają uwzględniane i odwzorowane w sieci później, gdy już proces uczenia jest na ukończeniu. W tych późniejszych etapach procesu uczenia pojedyncze neurony wchodzące w skład skupiska zostają precyzyjnie „rozprowadzone", uwzględniając różne formy i odmiany powiązanych ze sobą przypadków. Dzięki temu nawet mały stopień zróżnicowania poszczególnych podgrup wejściowych danych wchodzących w skład rozpoznawanej grupy może zostać uwzględniony w formie oddzielnych (chociaż sąsiadujących ze sobą) neuronów sieci, będących ich detektorami.

2.8.2. Pojęcie błędu w sieciach Kohonena

W sieciach Kohonena wykorzystuje się całkowicie odmienny błąd niż ten, który prezentowany był w trakcie działania metody wstecznej propagacji błędów. W metodzie

wstecznej propagacji błędów standardową funkcją błędu jest suma kwadratów różnic pomiędzy wartościami obliczanymi przez sieć a wartościami podawanymi przez nauczyciela. Można powiedzieć, że w tym przypadku RMS jest to miara, która służy do pomiaru odległości od wektora wyjściowego sieci do wektora zadanych wartości wyjściowych. W efekcie takiej interpretacji przebieg RMS prezentowany przez wiele programów modelujących sieci na *wykresie błędu* jest rzeczywiście miarą błędów produkowanych przez sieć sygnałów wyjściowych. Miara ta liczona jest z reguły dla całego zbioru uczącego, jest więc pewną średnią miarą błędu dla sieci.

Przy uczeniu sieci Kohonena zwyczajowo przyjmowaną funkcją błędu jest odległość wektora wag zwycięskiego neuronu od wektora wejściowego. W rezultacie takiego założenia wartość przedstawiana dla sieci Kohonena na *wykresie błędu* jest miarą rozbieżności pomiędzy wartością sygnałów wejściowych do sieci, a wartościami najlepszych wzorców (prototypów) tych sygnałów, przechowywanych w postaci wektorów wag w poszczególnych neuronach sieci. Ponieważ pokazywana na wykresie wartość jest wartością średnią tak zdefiniowanego błędu, liczoną dla całego zbioru uczącego, przeto nadal w dość sensowny sposób można ją interpretować. Po prostu jest to miara niedoskonałości działania sieci i jej malenie wskazywać będzie na postęp procesu uczenia. Takie rozumowanie jest w pełni uprawnione, gdyż malejąca średnia rozbieżność pomiędzy wektorem wejściowym a najlepszym wzorcem w sieci dowodzi polepszającego się rozmieszczenia wzorców.

Nie można jednak zapominać, że mimo podobnego zastosowania (do oceny i śledzenia postępu procesu uczenia) – tylko w przypadku sieci MLP lub RBF podawany błąd może być traktowany jako miara stopnia spełniania przez sieć postawionegojej zadania. W przypadku sieci Kohonena ten sam wykres błędu pokazuje polepszające dopasowywanie się sieci do danych wejściowych– co jednak wcale nie musi być tożsame z coraz lepszym wykonywaniem przez sieć powierzonego jej zadania. Jest to swoista „cena", jaką trzeba zapłacić za fakt, że decydując się na wykorzystanie techniki uczenia bez nauczyciela nie podaje się sieci wyraźnych wskazówek, czego naprawdę się od niej oczekuje. W wyniku tego samoistny proces doskonalenia działania sieci może niekiedy prowadzić na manowce.

2.8.3. Etapy procesu uczenia sieci Kohonena

W sieci Kohonena możliwe jest określenie wartości początkowych i końcowych dla współczynnika uczenia oraz dla rozmiaru sąsiedztwa. Ustalając te wartości można decydować o tym, czy zmiany zachodzące w trakcie procesu samouczenia sieci będą

szybkie i radykalne, czy też polegać będą na delikatnym i subtelnym dostrajaniu jej parametrów do drobnych szczegółów rozwiązywanego zadania.

Uczenie Kohonena jest często wyraźnie dzielone na fazę wstępną (porządkowanie sieci) oraz fazę, w której odbywa się precyzyjne modyfikowanie parametrów neuronów (dostrajanie sieci). W pierwszej z nich współczynnik uczenia i zasięg sąsiedztwa powinny być duże, natomiast w drugiej fazie parametry te podlegają radykalnemu zmniejszeniu – czasem nawet w przypadku zasięgu sąsiedztwa – do zera. Kiedy sieć Kohonena będzie już nauczona, można eksperymentalnie sprawdzić, gdzie tworzą się skupienia i co one reprezentują, podając na wejście sieci obiekty o znanej (ustalonej) interpretacji.

Do jakościowej oceny wyniku uczenia sieci Kohonena szczególnie przydatna jest statystyka pokazująca częstość zwycięstw poszczególnych neuronów sieci. Statystyka ta pozwala na obserwowanie, gdzie na mapie topologicznej tworzą się skupienia odpowiadające szczególnie dużej częstości występowania wejściowych obiektów. W trakcie uruchomienia sieci przeprowadza się zliczanie, ile razy każdy z neuronów zwycięża (tzn. ile razy jest najbliższy prezentowanym przypadkom uczącym, testowym lub walidacyjnym). Wysokie częstości zwycięstw wskazują na centra skupień na mapie topologicznej. Z kolei neurony z zerową częstością zwycięstw nie są wcale wykorzystywane, co powszechnie uważa się za wskaźnik świadczący o nie całkowitym powodzeniu uczenia. Można tak sądzić, bowiem przy zerowej częstości zwycięstw pewnych neuronów nie są one wcale wykorzystywane, a to oznacza, że sieć nie wykorzystuje wszystkich dostępnych w niej zasobów. Jednak niekiedy liczba przypadków uczących jest zbyt mała, żeby „wysycić" możliwości sieci, co może powodować, że wystąpienie pewnych niewykorzystanych neuronów będzie nieuchronne i nie należy się tym sugerować.

Korzystne jest, jeśli wyświetlane są statystyki częstości zwycięstw oddzielnie dla zbioru uczącego i oddzielnie dla zbioru walidacyjnego. Jeśli wykryty w czasie testowania działania sieci Kohonena podział na skupienia jest istotnie różny w obu zbiorach to wskazuje to, że sieć nie nauczyła się istotnych cech pozwalających na grupowanie informacji, w wyniku czego nie potrafi dobrze generalizować uzyskanej umiejętności klasyfikacji danych.

2.8.4. Mapa topologiczna Kohonena

Przy ostatecznej ocenie działania sieci Kohonena bardzo cennym sposobem prezentacji wyników jest tak zwana *mapa topologiczna*. Przedstawia ona w sposób graficzny warstwę wyjściową sieci, najczęściej opartą na regularnej siatce neuronów rozło-

żonych równomiernie w dwóch wymiarach (chociaż bywają także przypadki wykorzystania sieci w układzie jednowymiarowym lub trójwymiarowym przy wykorzystaniu aksonometrii).

Działanie mapy topologicznej jest następujące. W trakcie prezentacji konkretnego przypadku wejściowego każdy neuron informuje o stopniu swojego podobieństwa do wprowadzonego wektora danych. Informacja ta wyrażana jest graficznie, na przykład przy pomocy wypełnionego czarnego kwadratu. Na tego typu mapie (analogicznej do pewnego stopnia do tzw. diagramów Hintona, używanych często w sieciach neuronowych do wizualizacji wartości wag) większy kwadrat reprezentuje większe podobieństwo, zaś neuron zwycięski jest wyróżniony poprzez otoczenie go prostokątną ramką.

Jeśli przetestuje się w opisany wyżej sposób różne przypadki wejściowe to na podstawie sposobu aktywacji neuronów można zaobserwować grupy powiązanych ze sobą neuronów, które mocniej odpowiadają na podobne przypadki wejściowe. Po stwierdzeniu tego można rozpocząć nadawanie etykiet neuronom w celu identyfikacji ich znaczenia. Można tego dokonać na podstawie interpretacji danych występujących w poszczególnych skupieniach rozpoznawanych przez poszczególne neurony. Na przykład jeśli sygnały wejściowe wprowadzane do sieci odpowiadają różnym zestawom objawów (symptomów) rejestrowanych u pacjentów, to wówczas po wytrenowaniu sieci Kohonena może się ujawnić grupa neuronów sygnalizujących na przykład pacjentów cierpiących na zapalenie wątroby, osobno będzie grupa takich, u których choroba ta jest dodatkowo komplikowana przez cukrzycę, zaś jeszcze osobno może ujawnić się grupa, u której przyczyną choroby nie jest zapalenie, tylko rak. Może się zdarzyć, że ten sam neuron zwycięży zarówno dla przypadku *zapalenie* jak również dla przypadku *rak* (w oparciu o wiele symptomów te dwa zbiory nie są dobrze separowalne); nadaje się wówczas takiemu neuronowi etykietę „?" (nieokreślony). Po zakończeniu pracy ze wszystkimi znanymi przypadkami, można także wszystkim pozostałym neuronom (które nigdy nie były zwycięzcami) nadać etykietę „?" (którą można tu interpretować jako niewykorzystany). Inny sposób postępowania z „niezidentyfikowanymi" neuronami może polegać na ponownym uruchomieniu sieci, sprawdzeniu, na jaki typ danych te niewykorzystane neurony reagują najmocniej i nadaniu im takiej właśnie etykiety.

Kiedy już wszystkim neuronom zostanie nadana etykieta, można skorzystać z *mapy topologicznej* aby sprawdzić, jak dobrze sieć Kohonena klasyfikuje zbiór walidacyjny. Jeśli sieć Kohonena jest stosowana do rozwiązania jakiegoś rzeczywistego problemu, to zwykle z góry nie wiadomo, jakie skupienia występują w danych. W takim przypadku należy po prostu nadać skupieniom (zidentyfikowanym przy pomocy wyżej

opisanej procedury) jakieś umowne nazwy symboliczne (na przykład „Klasa 1”, „Klasa 2” itd.), a następnie analizując dane próbować określić, jakie znaczenie może zostać przypisane poszczególnym skupieniom.

3. Materiał badawczy

3.1. Opis materiału badawczego

Badania, których wyniki zamieszczono w tekście pracy, przeprowadzono na próbkach mowy patologicznej pochodzących z jednego tylko źródła, chociaż generalnie badania nad postaciami deformacji mowy prowadzone są w Laboratorium Biocybernetyki AGH na wielu różnych typach problemów prowadzących do efektu patologii mowy: chirurgii szczękowej, protetyce stomatologicznej, chirurgii laryngologicznej itp. Dla zachowania jednorodności wyników w niniejszej pracy ograniczono się wyłącznie do próbek mowy zawartych w zarejestrowanych wypowiedziach dzieci z rozszczepem podniebienia – operowanych z powodu rozszczepu podniebienia pierwotnego i wtórnego w Klinice Chirurgii Plastycznej Akademii Medycznej w Gdańsku. Jako porównawcze próbki mowy wzorcowej posłużyły wypowiedzi dzieci z Przedszkola PSK nr 1 w Gdańsku. Zgromadzony materiał badawczy podzielono na 5 kategorii [Łab97]:

- wypowiedź wzorcowa,
- rozszczep podniebienia wtórnego częściowy,
- rozszczep podniebienia wtórnego całkowity,
- rozszczep podniebienia pierwotnego i wtórnego całkowity obustronny,
- rozszczep podniebienia pierwotnego i wtórnego całkowity lewostronny lub prawostronny.

Wprawdzie z punktu widzenia medycznego rozszczep podniebienia pierwotnego i wtórnego całkowity lewostronny oraz rozszczep podniebienia pierwotnego i wtórnego całkowity prawostronny stanowią dwie oddzielne grupy patologiczne, jednakże z punktu widzenia akustycznego jest to ten sam rodzaj patologii i dlatego zdecydowano się na stworzenie jednej wspólnej kategorii obejmującej te dwie grupy.

Ocenę medyczną wszystkich dzieci przeprowadzono w Zespole Wielospecjalistycznym Ortopedów Szczękowo-Twarzowych Instytuty Stomatologii AMG, którzy na podstawie badania rysów twarzy, wykonanych modeli diagnostycznych oraz badania czynnościowego narządu żucia dokonali oceny wady zgryzu, a także wpływu wykrytych zaburzeń na poprawność mowy dziecka (ocenianej subiektywnie „na słuch"). Następnie dzieci były oceniane przez laryngologa– foniatrę, gdzie poza oceną zmian estetyczno-morfologicznych przeprowadzono także badanie subiektywne zaburzeń czynności mowy wywoływanych przez rozważaną wadę morfologiczną.

Rys. 3-1. Przykłady rozszczepów podniebienia u dzieci.

Ze względu na obszerność próby i fakt szybkiego męczenia się dzieci ograniczono rozmiar rejestrowanych wypowiedzi do minimum. Każde dziecko wypowiadało tylko jedno zdanie:

Rudy pies goni szybko kota.

Zdanie to wybrano kierując się zarówno względami akustycznymi (występują w nim wszystkie fonemy, przy których spodziewano się wystąpienia deformacji wywołanych rozszczepem podniebienia) jak i względami psychologicznymi (zdanie to odwołuje się do sytuacji łatwej do zrozumienia i inspirującej wyobraźnię dziecka, a więc jego wypowiadanie – czasem wielokrotne – mniej nuży i obciąża badanego pacjenta). Nagrań wypowiedzi dokonano w Rozgłośni Polskiego Radia w Gdańsku, w warunkach studyjnych na magnetofonie pomiarowym ***Nagra IV SJ***. Następnie zarejestrowane wypowiedzi poddano analizie akustycznej z użyciem aparatury firmy ***Bruel&Kjaer*** [Wsz94].

3.2. Przekształcenie zarejestrowanych wypowiedzi do postaci cyfrowej

Rejestrację wypowiedzi stanowiących materiał badawczy wykorzystywany w niniejszej pracy dokonano w warunkach studyjnych za pomocą mikrofonu, którego dynamika wynosiła od 22 do 160 dB, dla zakresu częstotliwości 4 Hz do 40 kHz i przy nierównomierności jego charakterystyki mniejszej od 2 dB. Sygnał z wyjścia mikrofonu był rejestrowany poprzez przedwzmacniacz na rejestratorze magnetycznym w postaci analogowej (czasowych przebiegów napięcia). Rejestrator zapewniał pasmo przenoszenia od 25 Hz do 20kHz, nierównomierność charakterystyki na poziomie nie większym od 1 dB i dynamikę nie mniejszą niż 50 dB. Zarejestrowany na taśmie magnetycznej sygnał odtworzono następnie przy pięciokrotnie mniejszej prędkości przesuwu taśmy niż podczas procesu rejestracji sygnału. Celem tej operacji było uzyskanie pięciokrotnej transformacji czasowej i częstotliwościowej, niezbędnej dla pełnego wykorzystania charakterystyk zastosowanego w badaniach analizatora pasmowego. Następnie sygnał poddano preemfazie przy pomocy filtru górnoprzepustowego RC o tłumieniu 5 dB na oktawę, w zakresie częstotliwości 25 Hz do 1400 Hz. W sygnale pierwotnym odpowiada to częstotliwościom z przedziału od 125 Hz do 7000 Hz. Celem tej operacji było uwydatnienie w sygnale składowych o wyższych częstotliwościach, mających (jak wiadomo) duże znaczenie przy analizie sygnału mowy. Po filtracji sygnał poddano atenuacji (wzmocnieniu) do poziomu niezbędnego do pełnego wysterowania wejścia używanego analizatora. Tak przekształcony sygnał podano na wejście wąskopasmowego analizatora w celu wyznaczenia (na drodze pomiaru) jego dynamicznego widma amplitudowo-częstotliwościowego. W wyniku powyższej analizy otrzymano widmo sygnału w paśmie od 125 Hz do 12 kHz, przy szerokości pasma filtrów analizujących równej 125 Hz (jednakowej w całym paśmie przenoszenia), dyskretyzowanego czasowo z krokiem próbkowania około 9 ms.

Jak wykazały liczne wcześniejsze badania [Jas73, Sap66, Tad88a] informacja akustyczna zawarta w widmie fazowo-częstotliwościowym nie ma praktycznego wpływu na proces rozpoznawania sygnału akustycznego mowy. Dlatego postanowiono opracować poszukiwaną w tej pracy metodę wizualizacji patologii sygnału mowy wykorzystującą jedynie informacje zawarte w widmie amplitudowo-częstotliwościowym, właśnie takim, jakie udostępniał wspomniany analizator.

Nagrane i przetworzone w powyżej opisany sposób wypowiedzi wzorcowe oraz patologiczne zapisano w zdyskretyzowanej postaci macierzy widm chwilowych. Ma

cierz taka posiada 96 kolumn (odpowiadają one poszczególnym pasmom filtrów analizujących widmo) oraz liczbę wierszy, będącą długością wypowiedzi podzieloną przez szerokość okna czasowego (9 ms):

$$S_{kl} = \begin{bmatrix} s_{11} & s_{12} & \Lambda & s_{1l} \\ s_{21} & s_{22} & \Lambda & s_{2l} \\ \mathrm{M} & \mathrm{M} & & \mathrm{M} \\ s_{k1} & s_{k2} & \Lambda & s_{kl} \end{bmatrix} \tag{3.1}$$

gdzie: k – ilość wierszy w macierzy (proporcjonalna do długości wypowiedzi),

l – ilość filtrów pasmowych,

s_{ij} – poziom sygnału (wyrażony w dB).

Każdy pojedynczy wiersz takiej macierzy można interpretować jako widmo chwilowe (w danym oknie czasowym) rozważanego sygnału akustycznego, zaś każdą kolumnę – jako czasową zmienność sygnału w wydzielonym paśmie częstotliwości.

Zarejestrowany materiał badawczy, w opisanej wyżej postaci dwuwymiarowych macierzy widm chwilowych, zobrazowano w celach poglądowych w postaci cyfrowych wideogramów, a następnie (posługując się pomocniczo wspomnianą wizualizacją) wyizolowano z ciągłego zapisu te fragmenty sygnału, które odpowiadały pojedynczym wyrazom oraz wybranym głoskom (ze względu na spodziewane formy zniekształceń rozważanego sygnału mowy skupiono uwagę głównie na głoskach szumowych „*s*” i „*sz*”). Każdy taki wydzielony fragment został zapisany w postaci oddzielnej macierzy widm chwilowych, będącej podstawowym kwantem rozważnej dalej informacji. Macierze sporządzone w opisany sposób stanowiły dane wejściowe do zaproponowanej w pracy metody wizualizacji sygnału mowy.

4. Konwersja sygnału do postaci graficznej

4.1. Uwagi wstępne

Sygnały dźwiękowe, przetworzone i zapisane w postaci macierzy widm chwilowych są zazwyczaj przedstawiane graficznie przy pomocy wideogramów. Przykłady takich form wizualizacji zaprezentowano w rozdziale 1. Ponieważ wideogramy są wykresami trójwymiarowymi, to przedstawienie takiego wykresu na płaszczyźnie dwuwymiarowej nie zawsze jest czytelne oraz sprzyjające łatwej analizie sygnału. Co więcej, wideogramy koncentrują uwagę badacza na elementach fizycznych sygnału (czasowych zmianach spektrum) podczas gdy istotne z punktu widzenia problemów rozważanych w tej pracy cechy artykulacyjne sygnału mowy i jego właściwości percepcyjne prezentowane są na typowym wideogramie w sposób mało czytelny i uwikłany. Na koniec warto zauważyć, że wideogram będący prostą graficzną reprezentacją dynamicznego widma sygnału zawiera ogromną liczbę szczegółowych danych o cechach sygnału, przy czym w zdecydowanej większości są to dane całkowicie nieistotne dla osoby poszukującej w tym obrazie odpowiedzi na pytanie, czy sygnał reprezentuje wypowiedź poprawnie artykułowaną, czy patologicznie zniekształconą. Fakty te skłoniły autora do poszukiwania nowej, bardziej efektywnej metody obrazowania sygnałów akustycznych ze szczególnym uwzględnieniem cech i właściwości sygnału mowy patologicznej. Metoda ta w swoim założeniu miała być podobna do sposobu realizacji tego zadania przez człowieka. Założono przy tym, że słuchanie polega na tworzeniu pewnych „obrazów dźwiękowych", które następnie są przetwarzane przez wyższe piętra układu nerwowego.

Szczegóły funkcjonowania układu słuchowego człowieka zostały przedstawione w paragrafie 1.3, w tym miejscu należy jedynie przypomnieć, iż niebagatelną rolę w procesie słyszenia odgrywa błona podstawna (dokonująca pewnego rodzaju analizy czę-

stotliwościowej sygnału) oraz neurony nerwu słuchowego (przekazujące informacje akustyczne do systemu nerwowego). Biorąc pod uwagę powyższe przesłanki zdecydowano się na zastosowanie sieci neuronowej, której dane wejściowe stanowić będzie sygnał akustyczny zapisany w postaci macierzy widm chwilowych (opisanych w rozdziale 3.2). Sygnały wyjściowe z sieci będą podstawą do generacji dwuwymiarowych obrazów wizualizowanego sygnału akustycznego.

4.2. Wybór rodzaju sieci

Przy zastosowaniu sieci neuronowych do rozpoznawania i klasyfikowania różnego rodzaju danych niezwykle ważnym zagadnieniem jest odpowiedni wybór rodzaju i struktury sieci. Dokonanie na tym etapie błędnego wyboru może być przyczyną znacznych trudności, a wręcz niepowodzeń w procesie treningu i późniejszego funkcjonowania sieci. Możliwości w tym względzie jest wiele (porównaj paragraf 2.3), ostatecznie jednak zdecydowano się na wybór sieci Kohonena. Za trafnością tak dokonanego wyboru przemawiało kilka przesłanek:

- Po pierwsze sieć Kohonena jest siecią samouczącą się, co w znacznym stopniu ułatwia przygotowanie zbioru treningowego. Jest to wielka zaleta we wszelkiego rodzaju przypadkach, w których prawidłowa odpowiedź sieci nie jest z góry znana.
- Po drugie sieć taka może zostać zorganizowana w dowolnym układzie topologicznym, w tym w szczególności w postaci macierzy dwuwymiarowej. Postać taka bardzo ułatwia generowanie obrazów na podstawie odpowiedzi sieci.
- Kolejnym argumentem przemawiającym za zastosowaniem sieci Kohonena jest fakt, że w sieci takiej tylko jeden neuron jest zwycięzcą (posiada największą wartość sygnału wyjściowego), co pozwala na koncentrowanie uwagi wyłącznie na lokalizacji tego neuronu z pominięciem sygnałów wyjściowych wszystkich innych elementów sieci. To uproszczenie ma zasadnicze znaczenie podczas procesu generowania czytelnego obrazu dostarczanego przez sieć sygnału.
- Ponadto sieć neuronowa Kohonena pozwala na wprowadzenie tzw. pojęcia sąsiedztwa (własność tą omówiono w 2.8.1), dzięki czemu sygnały o podobnych parametrach rozpoznawane są (zazwyczaj) przez sąsiednie neurony (znajdujące się w pewnym podobszarze sieci).

Oczekiwano, że właśnie dzięki wprowadzeniu w sieci Kohonena sąsiedztwa, wzorcowe sygnały mowy poprawnej (z założenia podobne) będą rozpoznawane przez

pewną grupę sąsiadujących ze sobą neuronów, w przeciwieństwie do sygnałów odpowiadających mowie patologicznej, na które (w większości przypadków) reagowały będą także neurony pozostałe. Również dalsze zalety sieci Kohonena, opisane i przedyskutowane w rozdziale 2.8, miały znaczący wpływ na dokonany wybór rodzaju użytej sieci.

4.3. Struktura sieci

Równie istotnym zagadnieniem jak wybór rodzaju sieci jest dobór odpowiedniej wielkości sieci. Przy za małej ilości neuronów sieć może dokonać zbytniego uogólnienia sygnałów, co może prowadzić do nie rozróżniania kategorii o relatywnie wysokim stopniu wzajemnego podobieństwa. Natomiast zastosowanie zbyt dużej liczby neuronów powoduje wystąpienie tendencji uczenia się przez sieć wzorców na pamięć. Sytuacja taka może spowodować zaklasyfikowanie dwóch w miarę podobnych sygnałów do różnych kategorii. Przytoczone uwagi stanowią istotną wskazówkę ułatwiającą uzyskanie sieci o pożądanych właściwościach, jednak ani te ogólne (jakościowe) rozważania, ani dokładne studia literaturowe niemogły dostarczyć konkretnej sugestii na temat tego, jak ostatecznie powinna wyglądać sieć wykorzystywana do wizualizacji sygnałów mowy patologicznej.

W związku z tym w pracy niniejszej przeprowadzono obszerne badania, mające na celu wyznaczenie optymalnej liczby neuronów potrzebnych do prawidłowego zobrazowania różnego rodzaju sygnałów akustycznych. Wyniki tych badań zamieszczono w kolejnym rozdziale, natomiast na obecnym etapie rozważań można przyjąć, iż sieć została skonstruowana z N^2 neuronów zorganizowanych w postaci kwadratowej siatki o boku złożonym z N neuronów. W zależności od wyboru wartości N uzyskuje się zobrazowania sygnału o większej lub mniejszej rozdzielczości, ponieważ topologia sieci ma oczywiście ścisły związek z formatem generowanych obrazów. Ważne jest aby zastosowana rozdzielczość była dostatecznie duża dla wykrycia wszystkich istotnych postaci deformacji rozważanego sygnału, jednak z kolei nie tak duża, by absorbowała uwagę badacza różnicami o marginalnym czy drugorzędnym znaczeniu. Stąd poszukiwania właściwego rozmiaru sieci miały tak duże znaczenie w tej pracy i z tego powodu poświęcono im tak wiele uwagi.

Inne elementy struktury używanej sieci były łatwiejsze do wyznaczenia. Ustalono, że każdy z neuronów sieci będzie posiadał 96 wejść (odpowiada to wymiarowi pojedynczego wiersza macierzy widm chwilowych) oraz będzie realizował następującą formułę:

$$y = \sum_{i=1}^{96} x_i w_i \tag{4.1}$$

gdzie: y – sygnał wyjściowy neuronu,

x_i – i-ta składowa sygnału wejściowego,

w_i – i-ta waga neuronu.

W przeprowadzonych badaniach zastosowano dwa typy sieci Kohonena: bez sąsiedztwa oraz z sąsiedztwem.

Pojęcie sąsiedztwa zostało zdefiniowane w kategoriach metryki odległości wprowadzonej w dwuwymiarowej strukturze sieci. W pracy zastosowano następującą postać funkcji określającej sąsiedztwo w sieci:

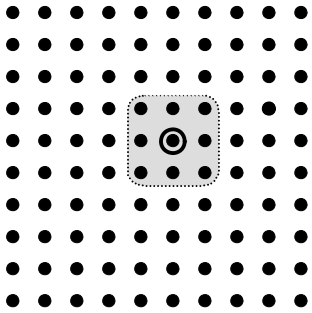
$$h\left(m, m^*\right) = \begin{cases} 1 & \rho\left(m, m^*\right) = 0 \\ 0.5 & 0 < \rho\left(m, m^*\right) < 2 \\ 0 & \rho\left(m, m^*\right) \geq 2 \end{cases} \tag{4.2}$$

gdzie: m – indeks badanego neuronu,

m^* – indeks zwycięzcy,

ρ – metryka odległości między neuronami.

Jako metrykę odległości pomiędzy neuronami przyjęto metrykę Euklidesa przy założeniu, że odległość między dwoma sąsiednimi neuronami leżącymi na jednej prostej wynosi 1. Powyższa reguła, mimo iż skomplikowana w swojej formie, sprowadza się do bardzo prostej zależności – neurony sąsiednie stanowi zbiór neuronów leżących w bezpośrednim sąsiedztwie neuronu zwycięskiego. Sytuację tą zilustrowano na rysunku 4-1. Na rysunku tym ciemne punkty reprezentują pojedyncze neurony, kółkiem zaznaczono neuron będący zwycięzcą, natomiast punkty znajdujące się w zaciemnionym obszarze są – w rozumieniu definicji przyjętej w tej pracy– „sąsiadami” zwycięzcy.



Rys. 4-1. Sąsiedztwo w sieci Kohonena.

W bardziej zaawansowanych pracach wykorzystujących sieci Kohonena zwykle wprowadza się w trakcie uczenia sieci mechanizm stopniowego zawężania zasięgu „sąsiedztwa". W opisywanych tu badaniach element ten nie odgrywał znaczącej roli ze względu na bardzo prostą, podaną wyżej, definicję pojęcia sąsiedztwa, której dynamiczna transformacja w trakcie uczenia sieci jest trudna do wyobrażenia. Jednak rozważając kierunki dalszych badań, które mogą być podjęte w przyszłości w tym samym temacie, można by było sprawdzić, w jakim stopniu sieć o zmienianym w trakcie uczenia zasięgu sąsiedztwa mogłaby lepiej spełniać pokładane w niej nadzieje.

4.4. Trening sieci

Spośród całego zarejestrowanego materiału badawczego wyodrębniono sygnały wzorcowe, które posłużyły do utworzenia zbioru treningowego sieci. Warto podkreślić, iż zbiór treningowy był tworzony z wszystkich wypowiedzi wzorcowych należących do tego samego rodzaju wypowiedzi (to samo słowo lub ta sama głoska). Po wyselekcjonowaniu odpowiednich macierzy widm chwilowych, wszystkie wektory tych macierzy ponumerowano, tworząc w ten sposób jeden duży zbiór treningowy. Dodatkowo każdy z tych wektorów znormalizowano, tak aby jego długość wynosiła 1 (porównaj rozdział 2.8).

Przed przystąpieniem do procesu treningu zainicjalizowano wektory wagowe neuronów, a następnie również je znormalizowano. Inicjalizacji wag dokonano wartościami losowymi z przedziału $[-1.0\,;1.0]$ (po normalizacji przedział wartości współczynników wagowych wynosił mniej więcej $[-0.1\,;0.1]$).

W procesie treningu sieci w sposób losowy wybierano spośród zbioru treningowego pojedynczy wektor (reprezentujący widmo chwilowe sygnału) i podawano go na wejście sieci neuronowej. Dla każdego neuronu obliczano (zgodnie ze wzorem 4.1)

wartość sygnału wyjściowego oraz wyznaczono zwycięzcę. Numer zwycięskiego neuronu został wyznaczony w następujący sposób:

$$m^* = \arg\max_i(y_i) \tag{4.3}$$

gdzie: i – przebiega zbiór 1K N^2.

Następnie dokonywano korekty wag zwycięzcy oraz (w przypadku sieci z sąsiedztwem) jego sąsiadów. Na uwagę zasługuje fakt, iż funkcja $h(m,m^*)$ nie tylko określa sąsiedztwo w sieci, ale także determinuję stopień uczenia sąsiadów (w stosunku do uczenia zwycięzcy). Korekta wag odbywała się zgodnie z następującą formułą:

$$w_i^{'m} = w_i^m + \eta\, h(m,m^*)(x_i - w_i^m) \tag{4.4}$$

gdzie: η – współczynnik uczenia sieci,

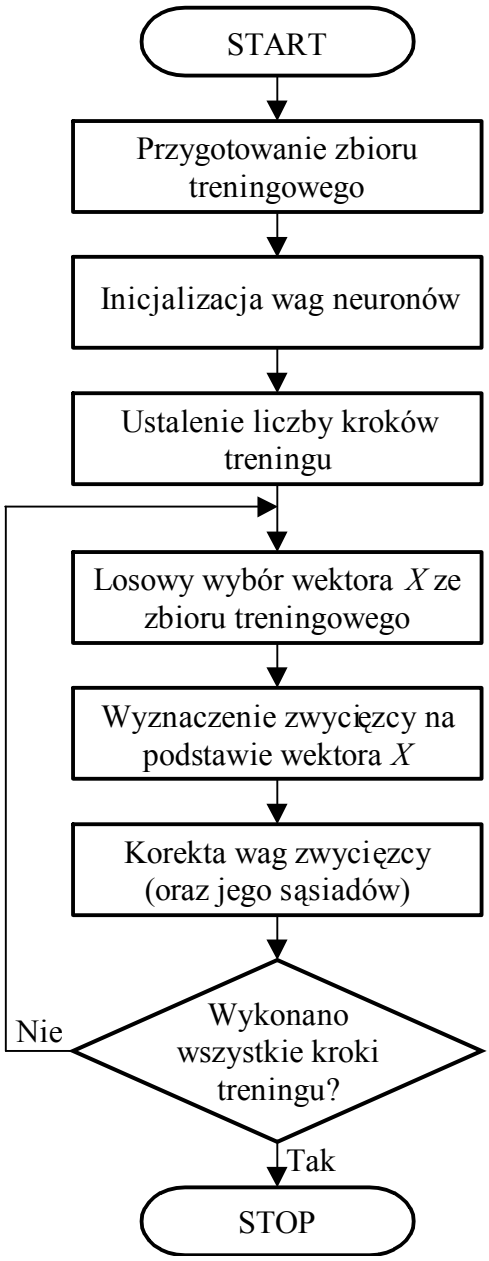
m – indeks rozważanego neuronu,

m^* – indeks „zwycięzcy”,

$h(m,m^*)$ – funkcja określająca sąsiedztwo.

W przeprowadzonych badaniach zdecydowano się na zastosowanie stałego współczynnika uczenia sieci o wartość 0.1. Dobierając tą wartość oparto się na rezultatach badań, które wykazały, iż stosowanie zmiennego współczynnika uczenia sieci nie powoduje poprawy rezultatów treningu sieci, ani też znacząco nie przyspiesza treningu.

Powyższy proces został powtórzony tyle razy ile wynosiła liczba kroków treningu. Cały proces treningu sieci można zilustrować za pomocą schematu blokowego przedstawionego na rysunku 4-2.



Rys. 4-2. Schemat blokowy treningu sieci.

4.5. Generacja obrazów topologicznych

Po zakończeniu procesu treningu, sieć jest już gotowa do generacji obrazów topologicznych, będących swoistym odwzorowaniem zapisanej w strukturze sygnałów mowy metody ich artykulacji (wraz z deformacjami wynikającymi z rozważanych w pracy form patologii). Podczas treningu pojedynczy element zbioru treningowego reprezentował wartości amplitud sygnału w poszczególnych pasmach, w danej chwili cza-

sowej. Wziąwszy pod uwagę zdolność sieci Kohonena do samoorganizacji i automatycznego grupowania obiektów o podobnych właściwościach spodziewano się, że podobne wektory, wchodzące w skład macierzy widm chwilowych, będą rozpoznawane przez te same neurony. Zatem poszczególne neurony sieci Kohonena zostaną uwrażliwione na pewną szczególna postać sygnału. Ponadto dzięki wprowadzeniu do sieci neuronowej Kohonena sąsiedztwa, w obrębie sieci utworzą się pewne niewielkie obszary, złożone z kilku neuronów odpowiedzialnych za rozpoznawanie danego typu sygnałów.

Przed przystąpieniem do generacji obrazu sygnału akustycznego zapisanego w postaci macierzy widm chwilowych należy znormalizować wszystkie wiersze tej macierzy. Następnie na wejście sieci podaje się kolejne wiersze tej macierzy. Na podkreślenie zasługuje fakt, że w przeciwieństwie do procesu treningu sieci, wiersze macierzy widm chwilowych podawane są w porządku czasowym (czemu odpowiada ich kolejność występowania w macierzy). Dla każdego z powyższych wektorów wyznaczany jest zwycięzca (neuron generujący na swoim wyjściu sygnał o największej wartości) oraz zapisywane są jego współrzędne. W kroku następnym węzły odpowiadające kolejnym zwycięskim neuronom (reprezentowane na obrazie przez punkty) są łączone odcinkami linii prostej. Powstała w ten sposób trajektoria, łącząca lokalizacje „zwycięskich" neuronów, stanowi obraz reprezentujący badany sygnał mowy.

5. Wyniki badań

W niniejszym rozdziale zaprezentowane zostaną obrazy uzyskane przez autora pracy jako wynik wizualizacji sygnału mowy prawidłowej i patologicznej. Prezentowane obrazy zostały wygenerowane zgodnie z algorytmem opisanym w poprzednim rozdziale. Obrazy te można podzielić na dwie grupy. W pierwszej części zaprezentowane zostaną obrazy całych słów, natomiast w części drugiej przedstawiono obrazy pojedynczych głosek. Należy podkreślić, iż ze względu na obszerność materiału badawczego, niemożliwe było zaprezentowanie wszystkich wygenerowanych w trakcie badań obrazów, chociaż takie wyczerpujące studium byłoby wysoce pouczające. Dlatego dla każdej omawianej grupy przedstawiono w tekście pracy jedynie kilka najbardziej reprezentatywnych obrazów, natomiast w celu wskazania ogólnych własności obrazów generowanych przy pomocy proponowanej metody wizualizacji zamieszczono tabele z wynikami przeprowadzonych testów, dających pewien pogląd w kwestii stopnia czułości i specyficzności tej metody.

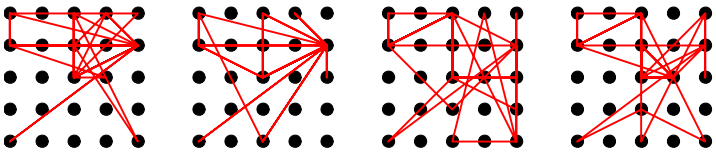
5.1. Obrazy wyników wizualizacji pojedynczych wyrazów

5.1.1. Wybrane wyniki badań sieci z wprowadzonym sąsiedztwem

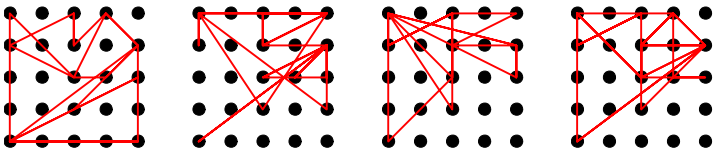
W punkcie tym przedstawiono obrazy pojedynczych wyrazów wygenerowane przy pomocy dwuwymiarowej sieci o topologii kwadratu. W sieci tej wprowadzono sąsiedztwo opisane reguła (4.2). W trakcie badań poszukiwano optymalnej metody zobrazowania, w związku z czym obrazy będące wizualizacjami rozważanych próbek mowy prawidłowej i patologicznej wygenerowano kilkakrotnie dla każdego z rozważanych wyrazów, zmieniając w każdym z eksperymentów rozmiar sieci oraz liczbę kro-

ków treningu. Tak jak już wspomniano, niemożliwe jest tu zaprezentowanie wszystkich wygenerowanych obrazów, jakkolwiek dopiero taka pełna prezentacja mogłaby dać kompletny pogląd na temat możliwości i ograniczeń opracowanej metody. Jednak starając się zaprezentować, nawet w ograniczonym zakresie, jak najbardziej zróżnicowaną część tych obrazów, zdecydowano się na prezentację 5 grup wyników procesu wizualizacji, podzielonych na wspomniane grupy ze względu na założoną (w każdej grupie inną) liczbę neuronów w sieci. Dla każdej grupy zaprezentowano obrazy będące wynikiem wizualizacji wypowiedzi wzorcowych oraz wypowiedzi należących do wszystkich 4 kategorii wypowiedzi patologicznych. W każdej grupie zaprezentowano obraz innego wyrazu, w celu zapewnienia możliwie najpełniejszego przeglądu możliwości opracowanej metody, jednak w uzupełnieniu należy podkreślić, że w obrębie jednej grupy i tej samej liczby kroków treningu, rezultaty uzyskane dla poszczególnych słów były do siebie bardzo podobne.

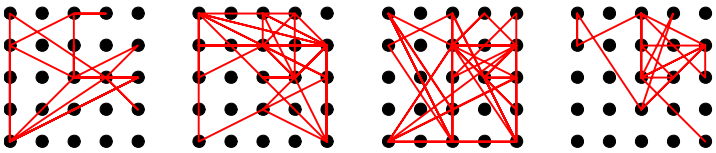
Na rys. 5-1 do 5-5 przedstawiono obrazy słowa „*rudy*". Obrazy te uzyskano dla sieci składającej się z 25 neuronów trenowanej przez 5000 kroków.



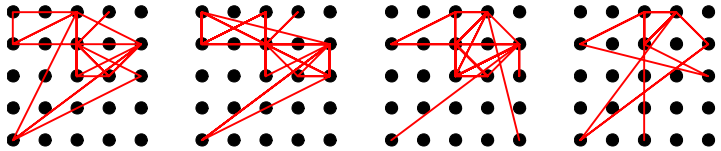
Rys. 5-1. Obrazy wzorcowych wypowiedzi słowa „*rudy*".



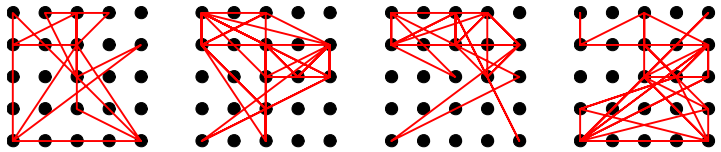
Rys. 5-2. Obrazy wypowiedzi słowa „*rudy*" – rozszczep podniebienia wtórnego częściowy.



Rys. 5-3. Obrazy wypowiedzi słowa „*rudy*" – rozszczep podniebienia wtórnego całkowity.



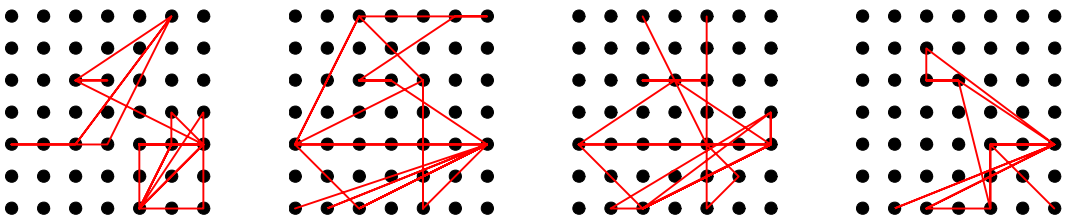
Rys. 5-4. Obrazy wypowiedzi słowa „*rudy*" – rozszczep podniebienia pierwotnego i wtórnego całkowity obustronny.



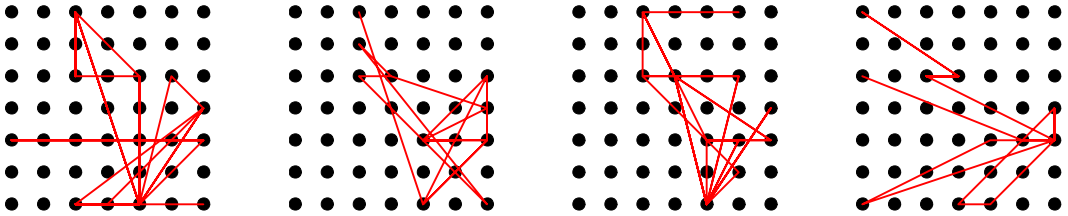
Rys. 5-5. Obrazy wypowiedzi słowa „*rudy*" – rozszczep podniebienia pierwotnego i wtórnego całkowity lewostronny lub prawostronny.

Na rysunkach przytoczonej grupy wizualizacji daje się zauważyć generalna tendencja, która potem będzie się powtarzała także i przy dalszych zobrazowaniach– otóż poszczególne wypowiedzi prawidłowe różnią się oczywiście wzajemnie od siebie, podobnie jak zróżnicowane są między sobą poszczególne trajektorie odpowiadające mowie różnych pacjentów cierpiących na tę samą wadę rozwojową (ten sam typ rozszczepu podniebienia). Jednak generalnie można zauważyć, że trajektorie wypowiedzi prawidłowych cechują się prostszym i bardziej regularnym obrazem od trajektorii dla mowy patologicznej, zaś w obrębie tego samego typu patologii wyniki wizualizacji cechują się pewnym wzajemnym podobieństwem produkowanych przez sieć neuronową trajektorii, z zachowaniem wyraźnej odmienności zobrazowań uzyskiwanych dla różnych typów i form patologii.

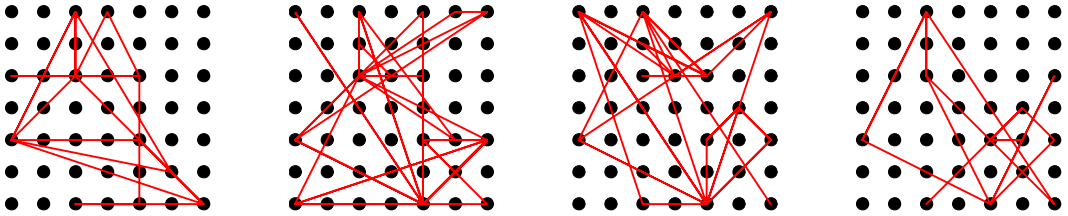
Przedstawione wyżej obrazy zostały uzyskane dla sieci neuronowej o małej rozdzielczości, relatywnie krótko uczonej. Kolejną prezentowaną grupę stanowią obrazy uzyskane dla nieco większej sieci, podlegającej także nieco dłuższemu procesowi uczenia. Jako wypowiedzi testowej użyto w tym przypadku słowa „*pies*". Sieć trenowana była 8000 kroków i składała się z 49 neuronów.



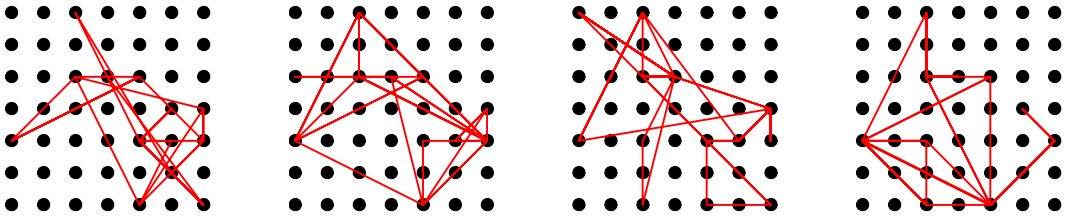
Rys. 5-6. Obrazy wzorcowych wypowiedzi słowa „*pies*".



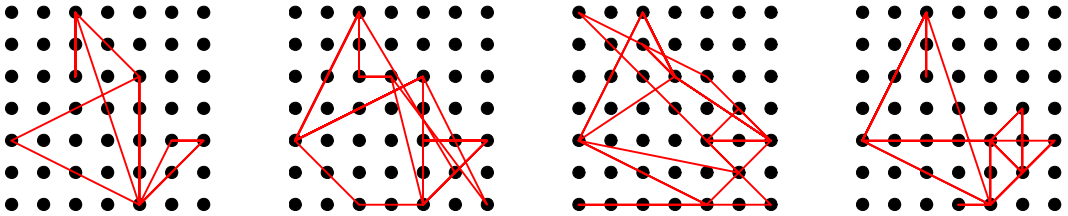
Rys. 5-7. Obrazy wypowiedzi słowa „*pies*" – rozszczep podniebienia wtórnego częściowy.



Rys. 5-8. Obrazy wypowiedzi słowa „*pies*" – rozszczep podniebienia wtórnego całkowity.



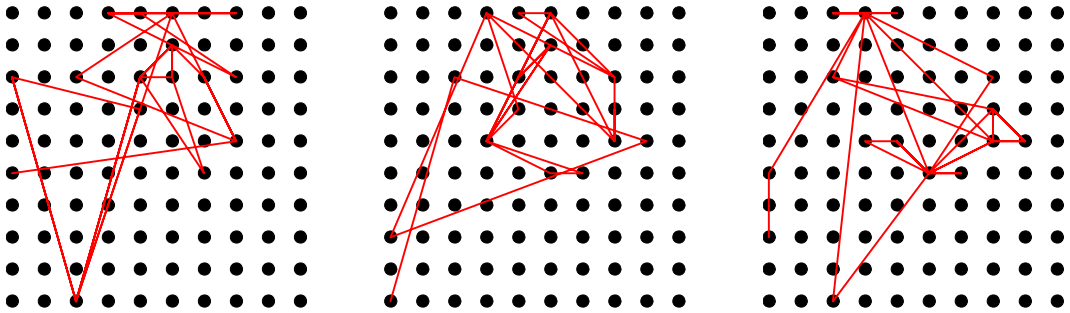
Rys. 5-9. Obrazy wypowiedzi słowa „*pies*" – rozszczep podniebienia pierwotnego i wtórnego całkowity obustronny.



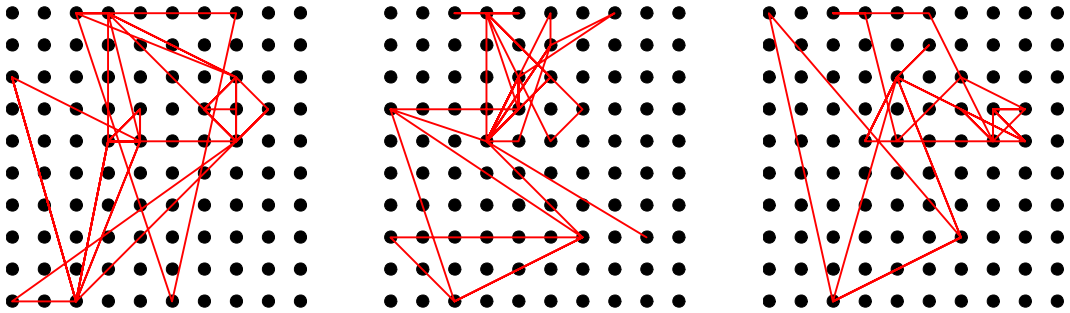
Rys. 5-10. Obrazy wypowiedzi słowa „*pies*" – rozszczep podniebienia pierwotnego i wtórnego całkowity lewostronny lub prawostronny.

Ogólne wnioski przedstawione dla poprzednio rozważanej sieci znajdują tutaj swoje potwierdzenie, jakkolwiek obraz trajektorii jest w tym przypadku generalnie mniej skomplikowany, co jest wynikiem zarówno większej rozdzielczości sieci jak i prostszej struktury fonetycznej rozważanego przykładowego wyrazu.

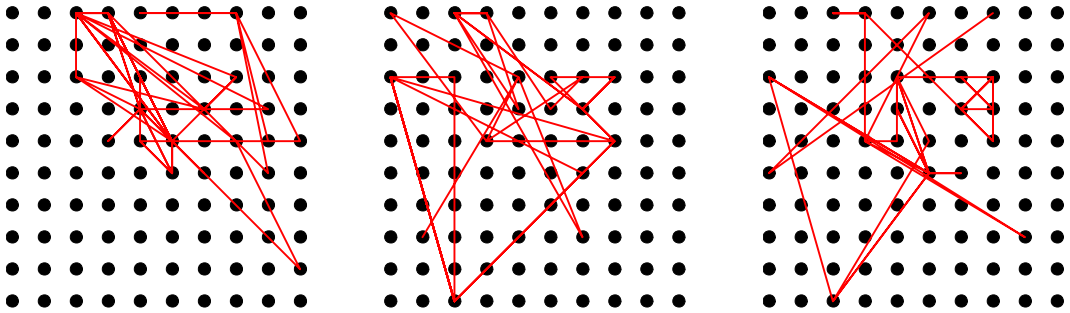
Jedną z hipotez roboczych, jakie testowano w trakcie badań, była hipoteza o celowości korzystania przy tworzeniu wizualizacji z sieci „niedouczonych". W związku z tym użyto większej sieci (o rozmiarach 10 na 10 neuronów) uczonej na tyle krótko, że trening przerwano w fazie nie całkiem ukształtowanych w sieci wzorców percepcyjnych. Wyniki były bardzo ciekawe. Rysunki 5-11 do 5-15 przedstawiają przykładowo wybrane dla tej serii badań obrazy słowa „*goni*" uzyskane przy pomocy sieci trenowanej przez 3000 kroków i składającej się ze 100 neuronów.



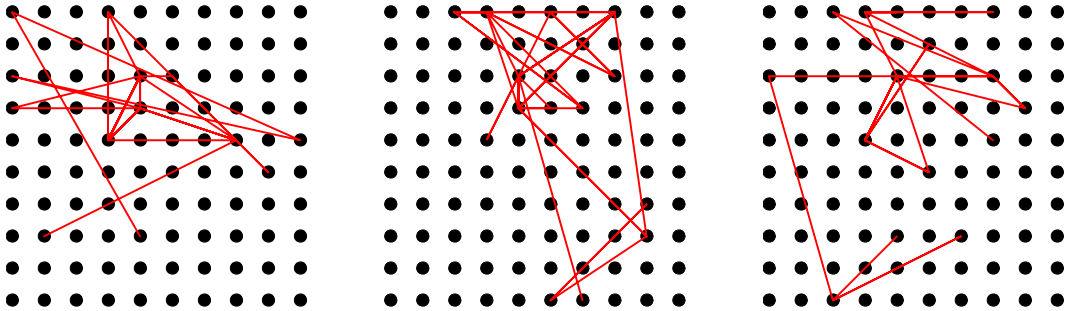
Rys. 5-11. Obrazy wzorcowych wypowiedzi słowa „*goni*".



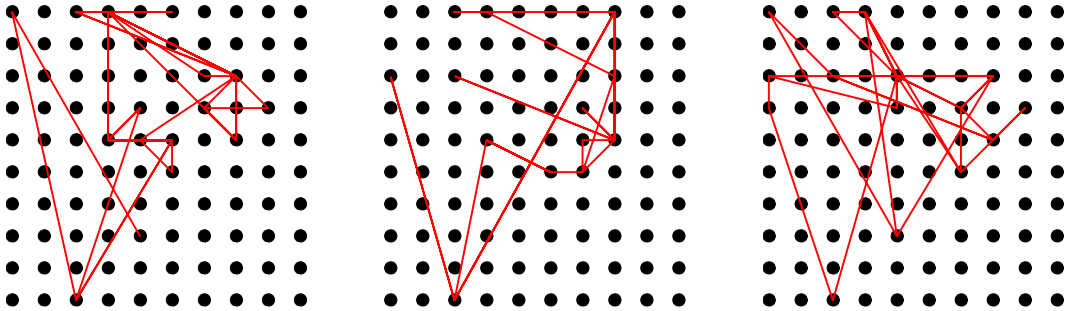
Rys. 5-12. Obrazy wypowiedzi słowa „*goni*" – rozszczep podniebienia wtórnego częściowy.



Rys. 5-13. Obrazy wypowiedzi słowa „*goni*" – rozszczep podniebienia wtórnego całkowity.



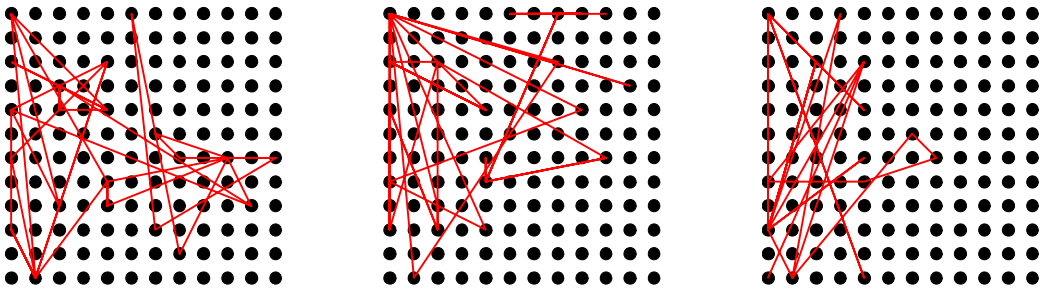
Rys. 5-14. Obrazy wypowiedzi słowa „*goni*" – rozszczep podniebienia pierwotnego i wtórnego całkowity obustronny.



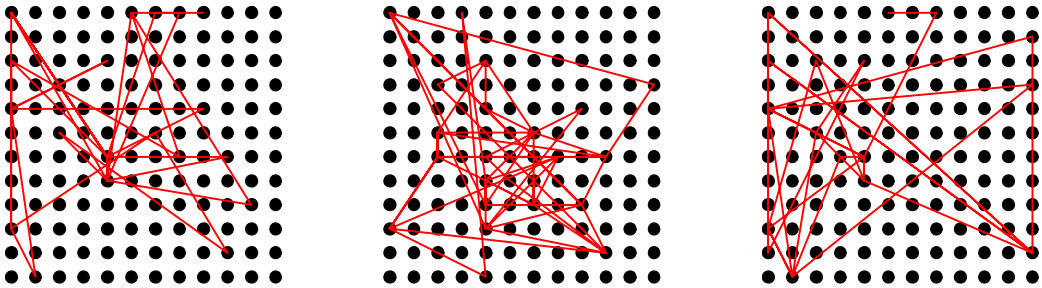
Rys. 5-15. Obrazy wypowiedzi słowa „*goni*" – rozszczep podniebienia pierwotnego i wtórnego całkowity lewostronny lub prawostronny.

Na uwagę zasługuje w tym przypadku dobre różnicowanie przez sieć różnych form patologii, co można uznać za częściowe potwierdzenie sformułowanej hipotezy.

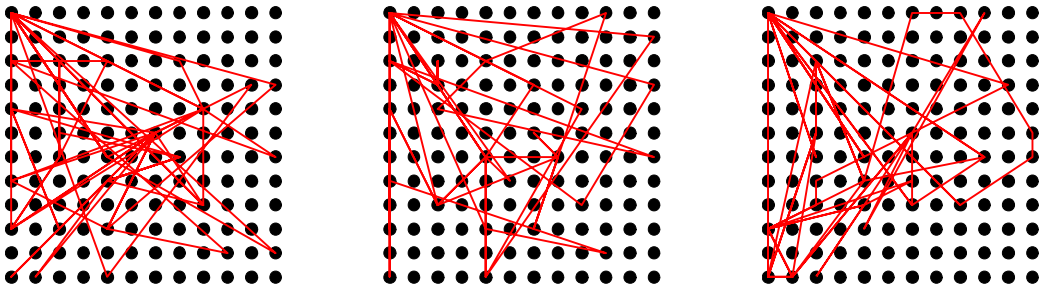
Podejmowano także próby użycia dużych, dobrze wytrenowanych sieci. Wyniki tej fazy badań reprezentują obrazy słowa „*szybko*" przedstawione na rys. 5-16 do 5-20. Otrzymano je przy pomocy sieci złożonej ze 144 neuronów, trenowanej w długim okresie aż 10000 kroków.



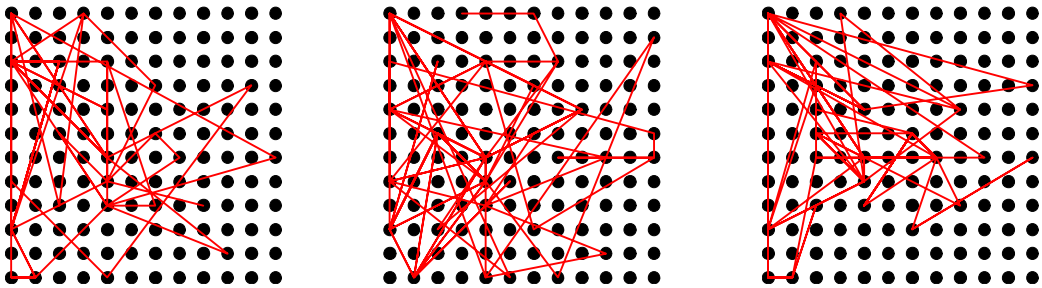
Rys. 5-16. Obrazy wzorcowych wypowiedzi słowa „*szybko*".



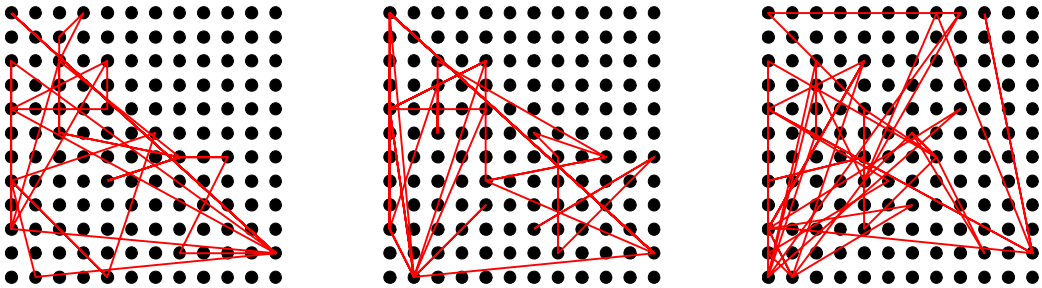
Rys. 5-17. Obrazy wypowiedzi słowa „*szybko*" – rozszczep podniebienia wtórnego częściowy.



Rys. 5-18. Obrazy wypowiedzi słowa „*szybko*" – rozszczep podniebienia wtórnego całkowity.

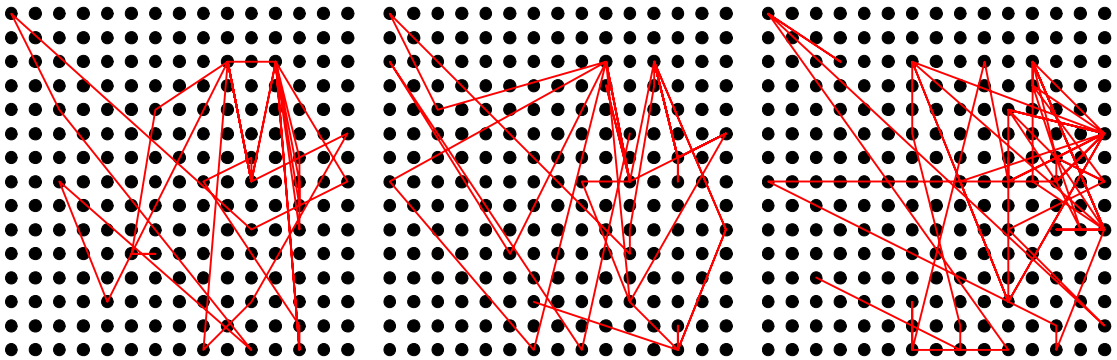


Rys. 5-19. Obrazy wypowiedzi słowa „*szybko*" – rozszczep podniebienia pierwotnego i wtórnego całkowity obustronny.

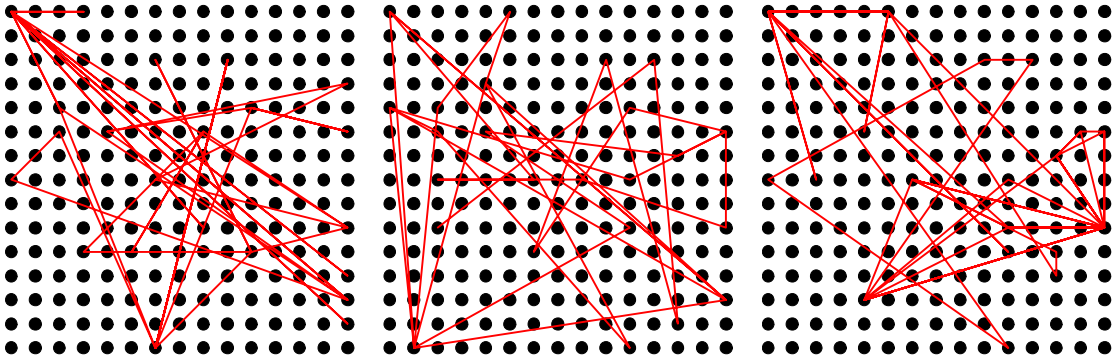


Rys. 5-20. Obrazy wypowiedzi słowa „*szybko*" – rozszczep podniebienia pierwotnego i wtórnego całkowity lewostronny lub prawostronny.

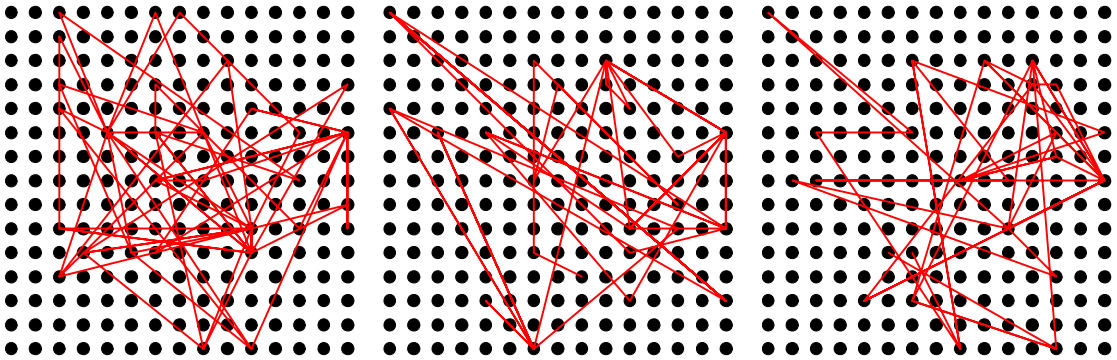
Wyniki były zachęcające, w związku z czym wykonano kolejny krok w tym samym kierunku (to znaczy zwiększono rozmiar sieci oraz czas uczenia). Wyniki tej próby ilustrować mogą obrazy słowa „*kota*", otrzymane przy pomocy sieci trenowanej przez 15000 kroków i składającej się z 225 neuronów.



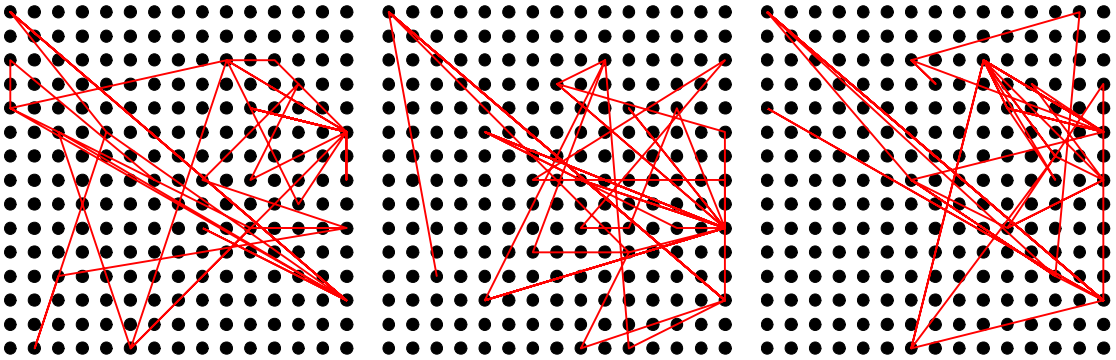
Rys. 5-21. Obrazy wzorcowych wypowiedzi słowa „*kota*".



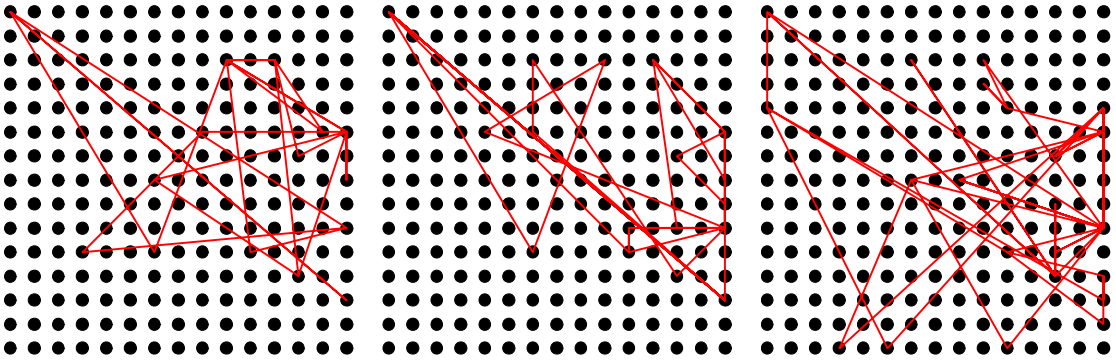
Rys. 5-22. Obrazy wypowiedzi słowa „*kota*" – rozszczep podniebienia wtórnego częściowy.



Rys. 5-23. Obrazy wypowiedzi słowa „*kota*" – rozszczep podniebienia wtórnego całkowity.



Rys. 5-24. Obrazy wypowiedzi słowa „*kota*" – rozszczep podniebienia pierwotnego i wtórnego całkowity obustronny.



Rys. 5-25. Obrazy wypowiedzi słowa „*kota*" – rozszczep podniebienia pierwotnego i wtórnego całkowity lewostronny lub prawostronny.

5.1.2. Dyskusja uzyskanych wyników dla całych wyrazów w sieci z wprowadzonym sąsiedztwem

Uzyskane wyniki należy uznać za interesujące. Należy podkreślić oryginalność uzyskanych form wizualizacji– w takiej postaci nikt nigdzie nie próbował analizować sygnałów mowy – ani prawidłowych, ani patologicznych. Niestety z przykrością należy także stwierdzić, że powyższe obrazy, mimo iż posiadają interesującą formę, są nadal zbyt złożone i nieregularne w swoim kształcie, w związku z czy z praktycznego (medycznego) punktu widzenia należy uznać je za mało przydatne. W obrazach tych trajektoria stanowiąca wynik wizualizacji sygnału rozprzestrzenia się po całym obszarze obrazu i ma skomplikowaną formę, tak że mimo dostrzegania w opisanych zapisach pewnych prawidłowości (komentowanych bezpośrednio podczas prezentacji poszczególnych obrazów) niemożliwe jest w oparciu o te formy wizualizacji łatwe rozróżnienie

wypowiedzi wzorcowych i patologicznych, a także znacznie utrudnione jestróżnicowanie różnych form i postaci patologii.

Niemniej mimo generalnie małej **bezpośredniej** użyteczności wykonanych prób, pozwoliły one na sformułowanie kilku użytecznych spostrzeżeń. Spostrzeżenia te zostaną teraz kolejno wymienione, ponieważ będą one istotnymi wskazówkami przy prowadzeniu wszystkich dalszych badań.

Po pierwsze badania przeprowadzone dla zakresu od 100 do 100000 kroków treningu wykazały, że liczba 2000-5000 kroków procesu uczenia (dokładna wartość zmienia się w zależności od rozmiaru sieci) jest wartością optymalną. Warto dodać, że wskazana wartość jest wartością graniczną pomiędzy siecią „niewytrenowaną", a siecią już „przetrenowaną". Analizując generowane obrazy trajektorii stwierdzono przy tym, że niewystarczające wytrenowanie sieci objawia się dużą ilością przypadkowych węzłów na generowanych obrazach. Spowodowane to jest faktem, że w sieci na zmiany w wejściowym sygnale reaguje wtedy jeszcze stosunkowo dużo neuronów posiadających losowe (wynikające z procesu inicjalizacji) wektory wagowe. Wprawdzie przeprowadzenie większej liczby kroków treningu eliminuje te „losowo reagujące" neurony oraz powoduje, iż w generowanych obrazach zaczyna się pojawiać coraz więcej wspólnych węzłów, dodatkowo jednak wzrasta całkowita liczba neuronów reagujących na wizualizowany sygnał, zmniejszając w ten sposób czytelność obrazu. W związku z tym najbardziej czytelny obraz (a zatem i najlepszy wynik wizualizacji) uzyskuje się przy sieci nieco niedouczonej. Wychwycenie właściwego momentu w toku procesu uczenia, kiedy sieć w optymalnym stopniu oddala się od obydwu wskazanych zagrożeń (całkowitego braku wiedzy i „przeuczenia") wymaga pewnej praktyki, ale jest możliwe. Tym niemniej występowanie takiej wrażliwości badanego narzędzia na czynnik długości procesu uczenia uznano za wadę rozważanego tu wariantu sieci i poszukiwano rozwiązania o lepszych właściwościach.

W pierwszej kolejności podjęto próbę ustalenia, czy polepszenia wyników nie przyniesie rezygnacja z efektu sąsiedztwa. Oparto się przy tym na następującej obserwacji. W trakcie treningu sieci z wprowadzonym sąsiedztwem poprawiane są również wagi sąsiadów „zwycięzcy", co powoduje, że w tworzonej podczas uczenia neuronowej mapie topologicznej, odwzorowującej potem (w trakcie eksploatacji sieci) poszczególne elementy sygnału mowy, tworzą się całe „domeny" neuronów o zbliżonych właściwościach. Każdy z neuronów domeny może odpowiedzieć na pewną ustaloną formę wejściowego sygnału i jest wysoce prawdopodobne, że przy sygnale zdeformowanym (a więc odmiennym od wzorca używanego w trakcie uczenia) któryś z sąsiadów wytrenowanego zwycięzcy okaże się być lepiej dopasowany do wizualizowanego wektora niż

sam zwycięzca. W takiej sytuacji podczas rozpoznawania i wizualizacji nieznanego wcześniej sygnału na skutek efektu sąsiedztwa pojawia się deformacja trajektorii, utrudniająca jej jednoznaczną interpretację, przy czym może się to zdarzyć zarówno dla próbek mowy wzorcowej jak i dla rozważanych form patologii. Zewnętrznie przejawia się to w taki sposób, że w trakcie uczenia sieć szybko traci swą zdolność uogólniania i zaczyna przejawiać efekty „nauki na pamięć". Skutkiem tego jest większe skomplikowanie kształtu generowanej krzywej oraz wzrost jej długości, a to zdecydowanie utrudnia jej interpretacjezgodnie z celami stawianymi w tej pracy.

Jeśli przytoczone wyżej rozumowanie jest słuszne, to sieć Kohonena pozbawiona efektu sąsiedztwa powinna lepiej się uczyć i lepiej spełniać stawiane przed nią zadania. Postanowiono sprawdzić tę hipotezę. Wyniki opisano w kolejnym podrozdziale.

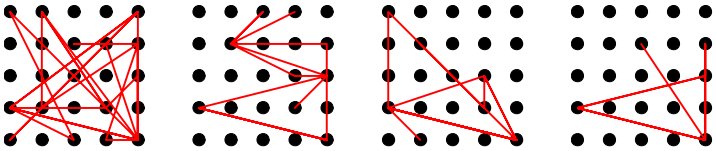
5.1.3. Sieć bez sąsiedztwa

Ponieważ – jak stwierdzono wyżej – sieć Kohonena z wprowadzonym sąsiedztwem niezbyt dobrze nadaje się do obrazowania różnic, jakie zachodzą pomiędzy mową prawidłową i patologiczną podczas wypowiadania całych (pojedynczych) wyrazów, postanowiono zbadać możliwość zastosowania do wizualizacji artykulacji pojedynczych wyrazów sieci Kohonena bez sąsiedztwa. W sieci tego typu, w fazie treningu, tylko jeden neuron jest uczony, a zatem można przypuszczać, że sieć taka posiada o wiele większe możliwości uogólniania danych niż sieć z sąsiedztwem. Oczekiwano, iż dzięki tej własności generowane krzywe będą miały bardziej regularne kształty i tym samym będzie możliwe łatwiejsze rozróżnienie obrazów wypowiedzi wzorcowych i patologicznych.

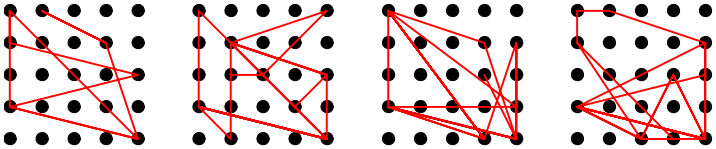
Dla zweryfikowania tej tezy wykonano próby wizualizacji pojedynczych wyrazów dla wielu konfiguracji sieci bez sąsiedztwa oraz dla wielu strategii uczenia. Podobnie jak w poprzednim podrozdziale zaprezentowana tutaj zostanie jedynie niewielka część wszystkich uzyskanych w trakcie badań obrazów, pozwalająca jednak na orientację w najważniejszych zaobserwowanych tendencjach. Tak jak poprzednio wyniki zostały podzielone, ze względu na liczbę neuronów w sieci, na 5 grup. W każdej grupie przedstawiono obrazy tego samego wyrazu co w analogicznej grupie w rozdziale poprzednim. Starano się w ten sposób pokazać różnicę pomiędzy obrazami generowanymi przez te dwie poddawane badaniom odmiany sieci Kohonena. Dodatkowo zaprezentowanie w każdej grupie sieci złożonej z innej liczby neuronów pozwala na zbadanie wpływu rozmiaru sieci na jakość generowanych obrazów. Należy podkreślić, iż wprawdzie rezultaty uzyskane dla poszczególnych wyrazów różniły się oczywiście między

sobą, jednak przytoczone dane można uznać za dość reprezentatywne. Na podstawie badań można było ustalić stopień przydatności rozważanego typu sieci do realizacji rozważanego tu zadania, a także można było na ich podstawie wyciągnąć ogólniejsze wnioski dotyczące między innymiminimalnej wymaganej liczby neuronów w sieci oraz najwłaściwszej strategii uczenia.

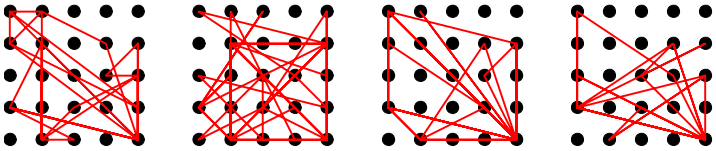
Pierwszą z prezentowanych grup obrazów stanowią obrazy słowa „*rudy*" (wypowiadanego przez dzieci z prawidłową budową narządów mowy oraz przez dzieci z różnymi formami rozszczepu podniebienia) uzyskane przy pomocy sieci złożonej z 25 neuronów i trenowanej przez 10000 kroków.



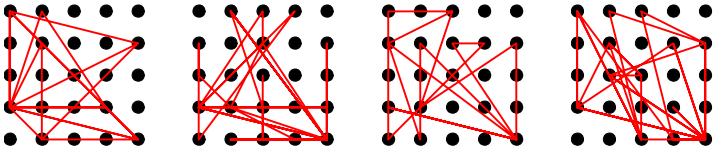
Rys. 5-26. Obrazy wzorcowych wypowiedzi słowa „*rudy*".



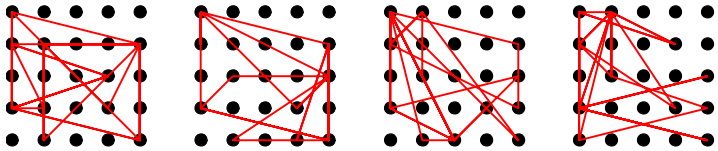
Rys. 5-27. Obrazy wypowiedzi słowa „*rudy*" – rozszczep podniebienia wtórnego częściowy.



Rys. 5-28. Obrazy wypowiedzi słowa „*rudy*" – rozszczep podniebienia wtórnego całkowity.

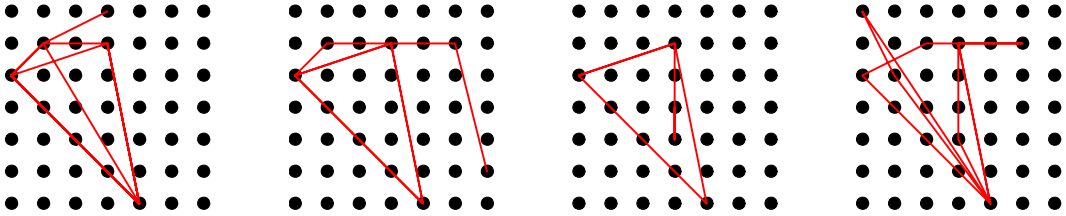


Rys. 5-29. Obrazy wypowiedzi słowa „*rudy*" – rozszczep podniebienia pierwotnego i wtórnego całkowity obustronny.

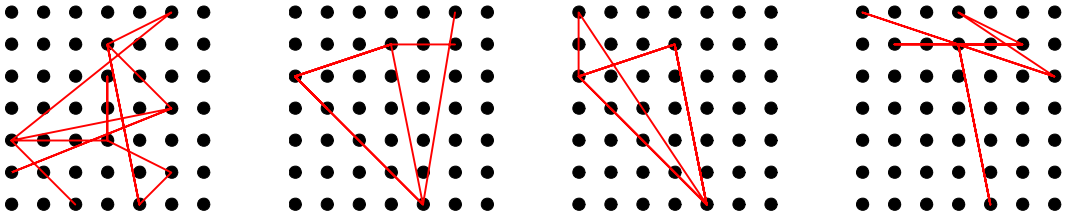


Rys. 5-30. Obrazy wypowiedzi słowa „*rudy*" – rozszczep podniebienia pierwotnego i wtórnego całkowity lewostronny lub prawostronny.

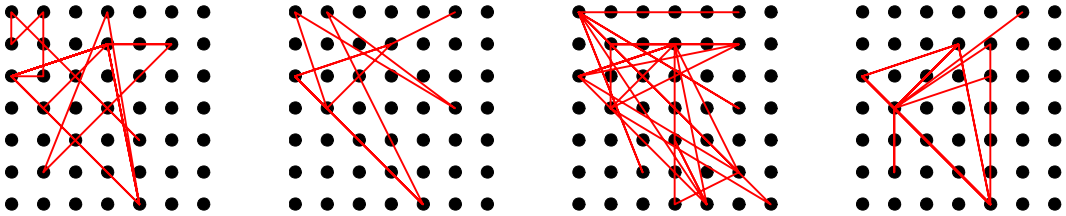
Na rys. 5-31 do 5-35 przedstawiono obrazy słowa „*pies*". Sieć wykorzystana do wygenerowania tych obrazów trenowana była przez 5000 kroków treningu i składała się z 49 neuronów.



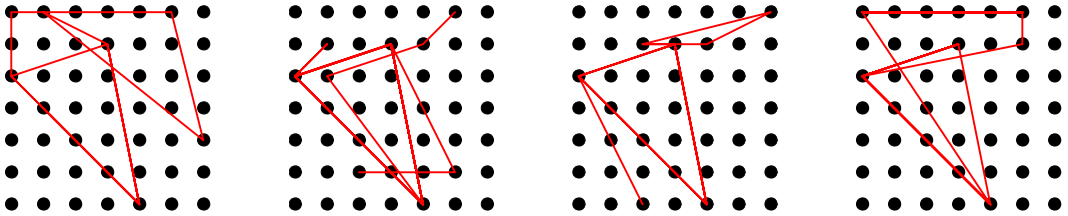
Rys. 5-31. Obrazy wzorcowych wypowiedzi słowa „*pies*".



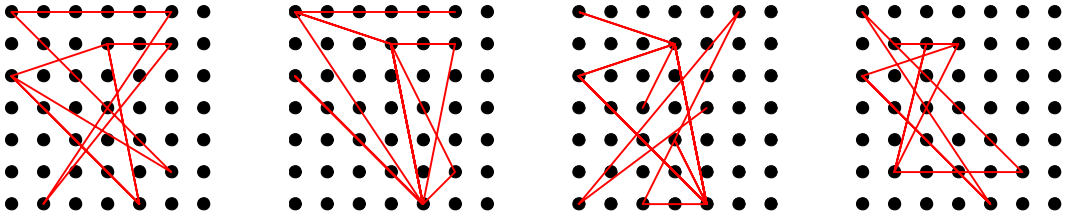
Rys. 5-32. Obrazy wypowiedzi słowa „*pies*" – rozszczep podniebienia wtórnego częściowy.



Rys. 5-33. Obrazy wypowiedzi słowa „*pies*" – rozszczep podniebienia wtórnego całkowity.

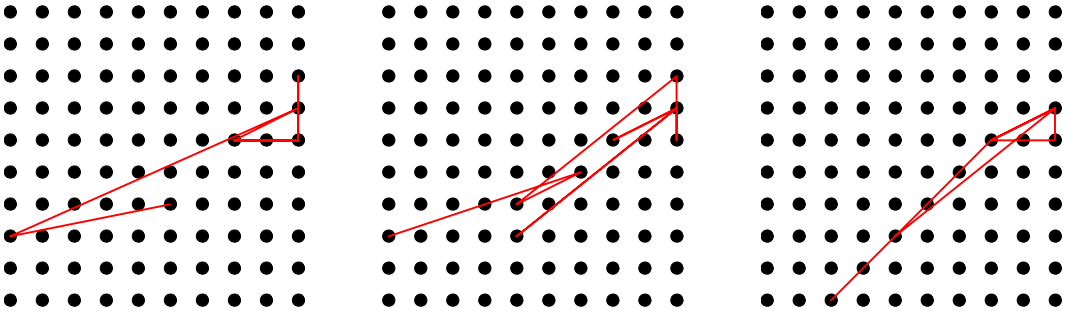


Rys. 5-34. Obrazy wypowiedzi słowa „*pies*" – rozszczep podniebienia pierwotnego i wtórnego całkowity obustronny.

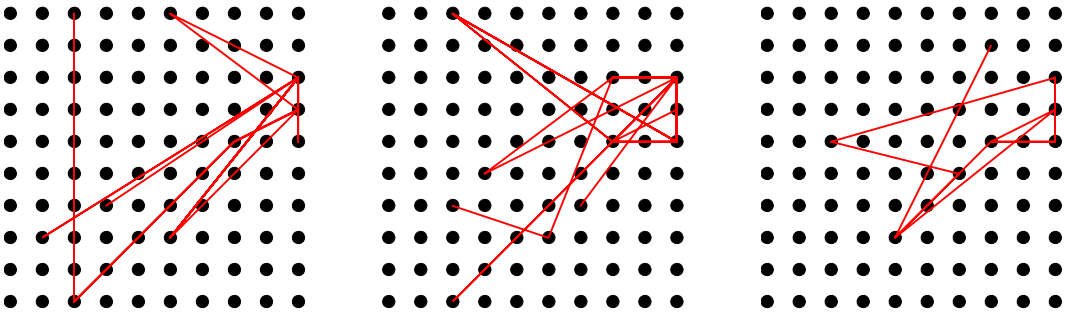


Rys. 5-35. Obrazy wypowiedzi słowa „*pies*" – rozszczep podniebienia pierwotnego i wtórnego całkowity lewostronny lub prawostronny.

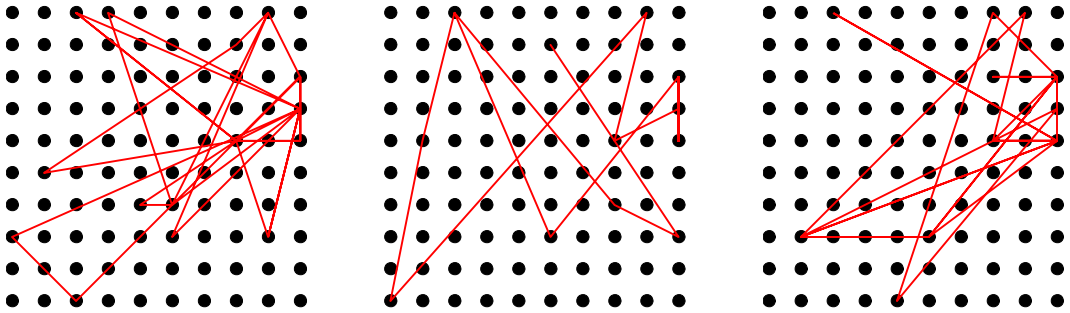
Kolejny zestaw rysunków prezentuje obrazy wyrazu „*goni*" uzyskane przy pomocy sieci złożonej ze 100 neuronów, trenowanej przez 10000 kroków.



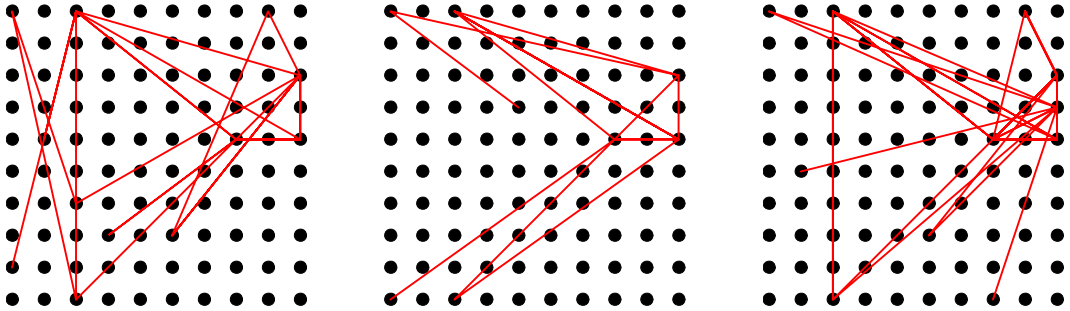
Rys. 5-36. Obrazy wzorcowych wypowiedzi słowa „*goni*".



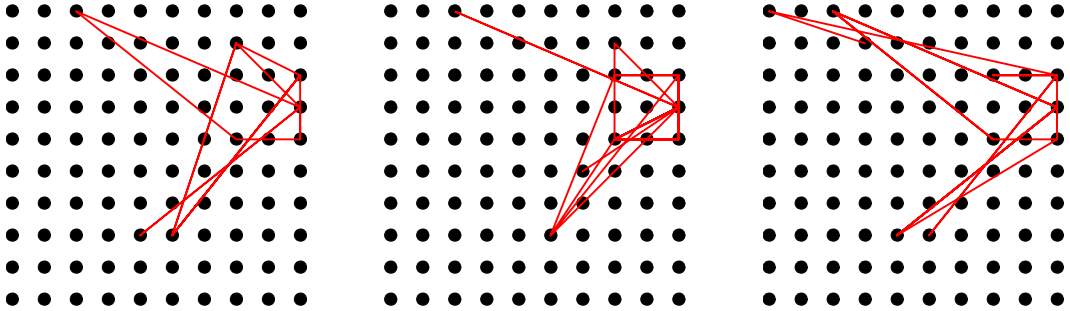
Rys. 5-37. Obrazy wypowiedzi słowa „*goni*" – rozszczep podniebienia wtórnego częściowy.



Rys. 5-38. Obrazy wypowiedzi słowa „*goni*" – rozszczep podniebienia wtórnego całkowity.

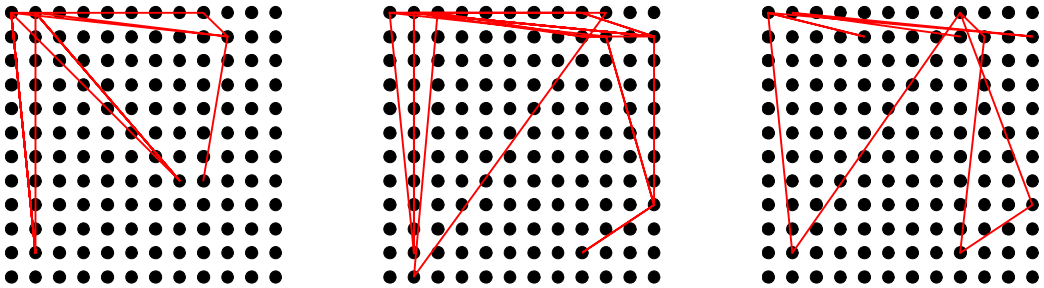


Rys. 5-39. Obrazy wypowiedzi słowa „*goni*" – rozszczep podniebienia pierwotnego i wtórnego całkowity obustronny.

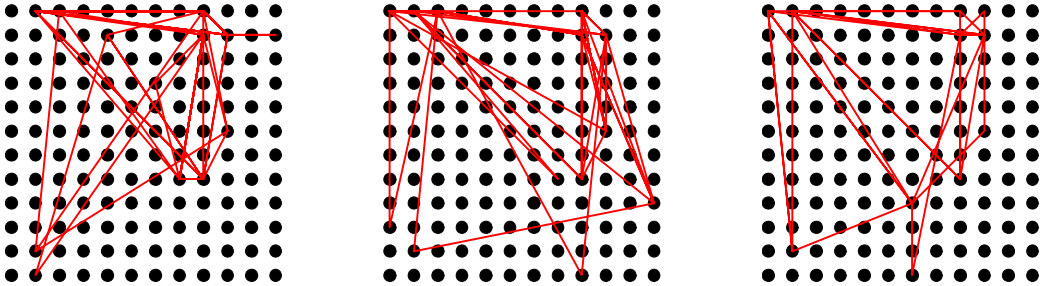


Rys. 5-40. Obrazy wypowiedzi słowa „*goni*" – rozszczep podniebienia pierwotnego i wtórnego całkowity lewostronny lub prawostronny.

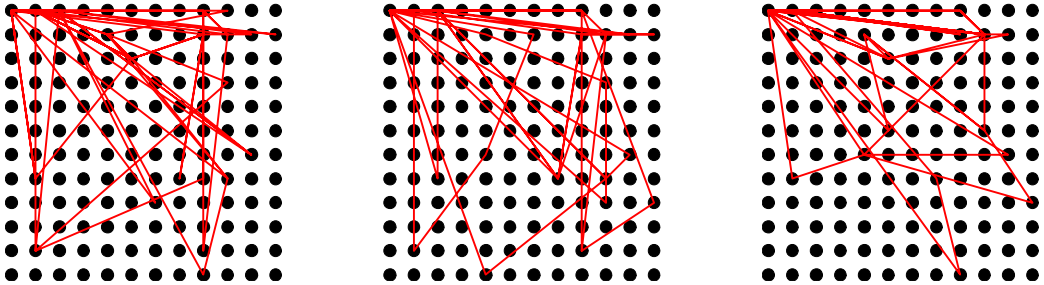
Czwartą z prezentowanych grup stanowią obrazy słowa „*szybko*" uzyskane przy pomocy sieci złożonej ze 144 neuronów i trenowanej przez 20000 kroków.



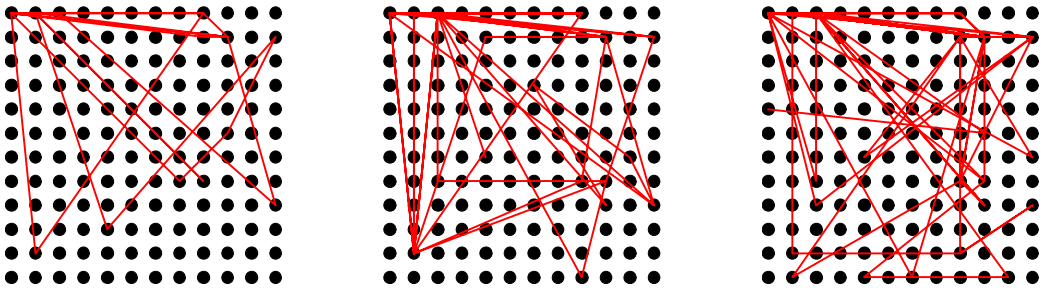
Rys. 5-41. Obrazy wzorcowych wypowiedzi słowa „*szybko*”.



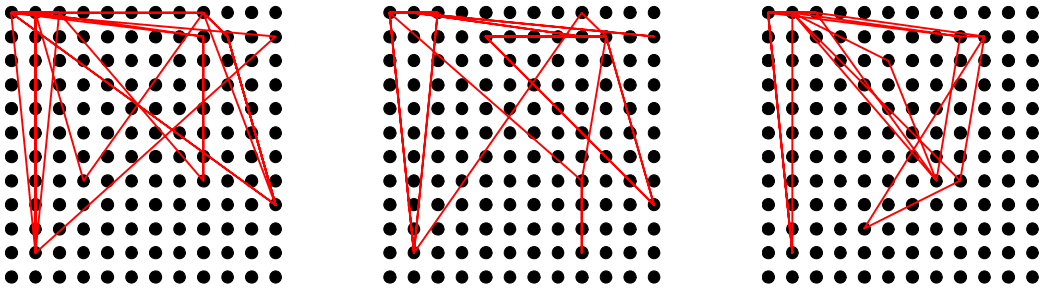
Rys. 5-42. Obrazy wypowiedzi słowa „*szybko*” – rozszczep podniebienia wtórnego częściowy.



Rys. 5-43. Obrazy wypowiedzi słowa „*szybko*” – rozszczep podniebienia wtórnego całkowity.

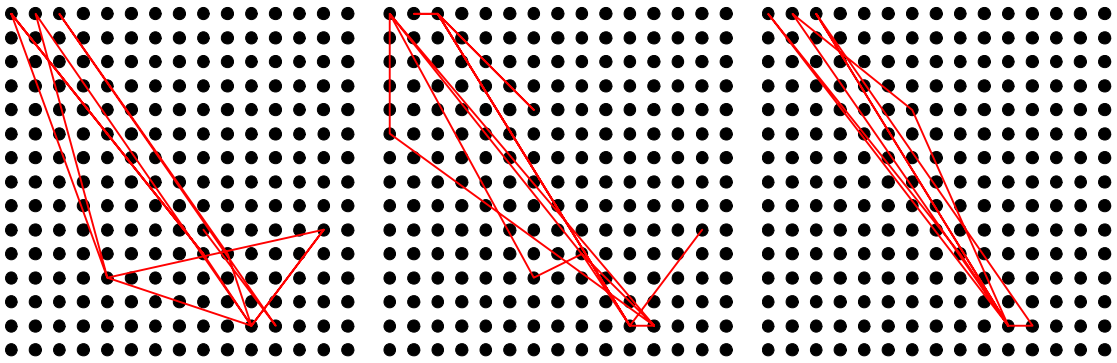


Rys. 5-44. Obrazy wypowiedzi słowa „*szybko*” – rozszczep podniebienia pierwotnego i wtórnego całkowity obustronny.

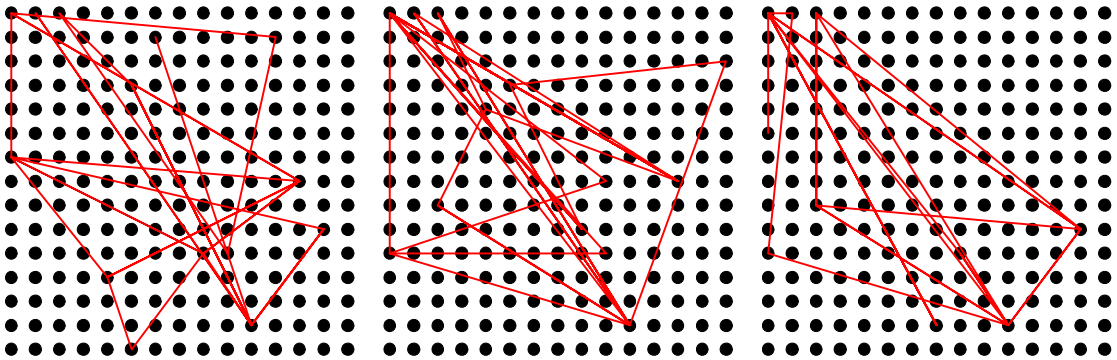


Rys. 5-45. Obrazy wypowiedzi słowa „*szybko*" – rozszczep podniebienia pierwotnego i wtórnego całkowity lewostronny lub prawostronny.

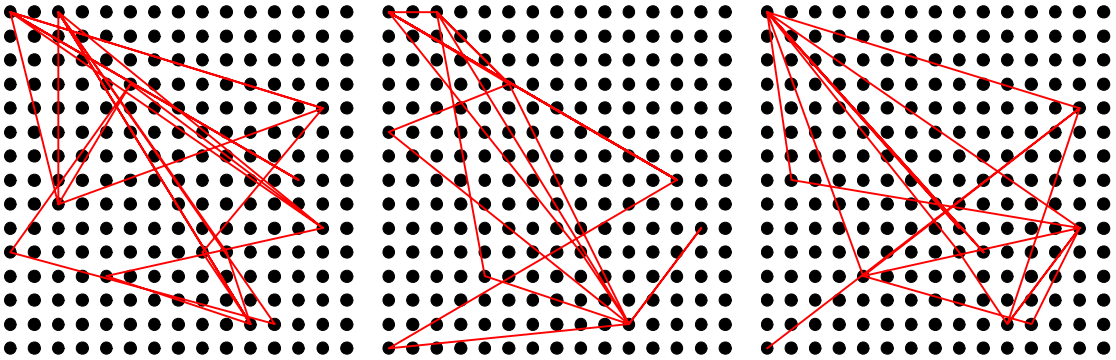
Ostatnią prezentowaną w tym paragrafie grupą obrazów są zamieszczone na rys. 5-46 do 5-50 obrazy słowa „*kota*" uzyskane przy pomocy sieci trenowanej przez 10000 kroków i złożonej z 255 neuronów.



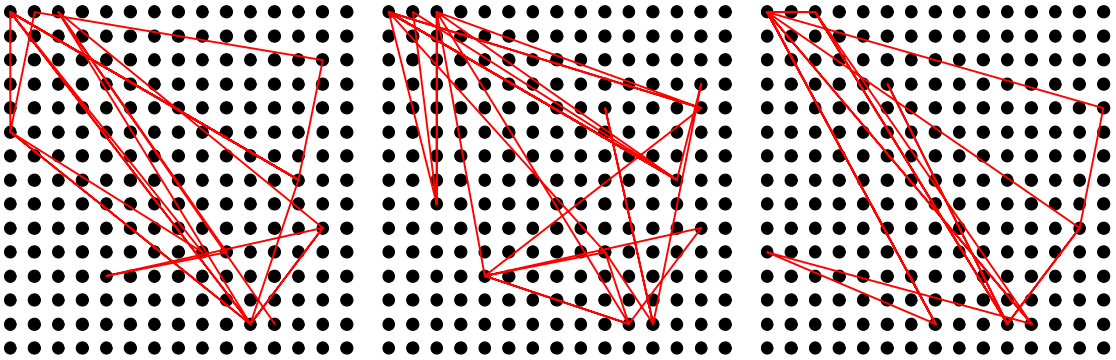
Rys. 5-46. Obrazy wzorcowych wypowiedzi słowa „*kota*".



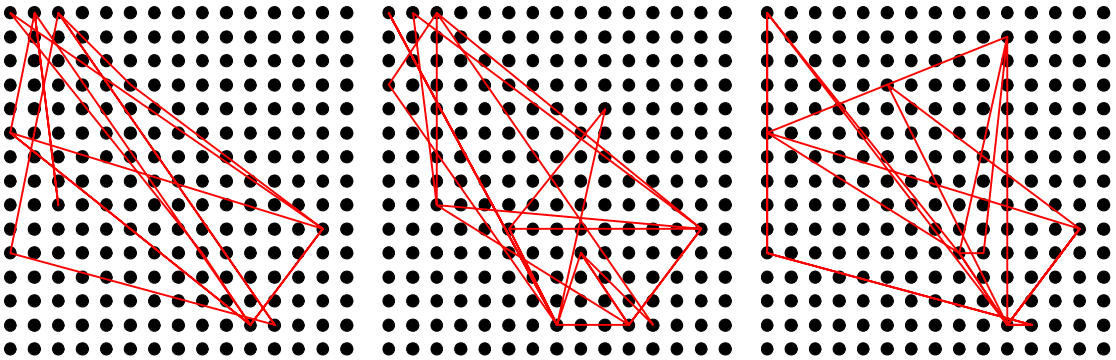
Rys. 5-47. Obrazy wypowiedzi słowa „*kota*" – rozszczep podniebienia wtórnego częściowy.



Rys. 5-48. Obrazy wypowiedzi słowa „*kota*" – rozszczep podniebienia wtórnego całkowity.



Rys. 5-49. Obrazy wypowiedzi słowa „*kota*" – rozszczep podniebienia pierwotnego i wtórnego całkowity obustronny.



Rys. 5-50. Obrazy wypowiedzi słowa „*kota*" – rozszczep podniebienia pierwotnego i wtórnego całkowity lewostronny lub prawostronny.

5.1.4. Dyskusja uzyskanych wyników dla całych wyrazów w sieci bez sąsiedztwa

Z postaci zaprezentowanych wyżej obrazów można wyciągnąć kilka wniosków dotyczących własności sieci Kohonena bez sąsiedztwa wykorzystywanej jako narzędzie do wizualizacji form patologii mowy. Po pierwsze widać, iż na każdym z tych obrazów istnieje pewien wspólny dla wypowiedzi wzorcowych oraz patologicznych obszar sieci, przez który przebiega krzywa obrazująca wypowiedź. Zjawisko to można wytłumaczyć w ten sposób, iż w każdej z tych wypowiedzi istnieją pewne, w miarę podobne do siebie fragmenty mowy (fonemy, mikrofonemy), które reprezentowane są przez te same neurony znajdujące się (po wytrenowaniu) w tym właśnie wspólnym obszarze sieci. Dodatkowo, na większości obrazów wypowiedzi wzorcowych (za wyjątkiem obrazów wygenerowanych przez sieć o zbyt małym rozmiarze, którą między innymi z tego powodu należy uznać za nieprzydatną) można wyróżnić pewien podobny kształt krzywych reprezentujący te wypowiedzi. Kształt ten jest również widoczny na obrazach wypowiedzi patologicznych, ale jest on zdeformowany i nie tak wyraźny jak w przypadku wypowiedzi wzorcowych. Jest to dokładnie zgodne z intuicyjnym wyobrażeniem procesu artykulacji mowy patologicznej jako procesu polegającego na zaburzonym (a więc także bardziej skomplikowanym) naśladownictwie procesu artykulacji poprawnej. Co więcej, można przyjąć, że im większy jest stopień patologii rozważanej wypowiedzi, tym bardziej znaczące jest także zdeformowanie i skomplikowanie trajektorii odwzorowującej tę wypowiedź na płaszczyźnie obrazu produkowanego przez sieć.

Własność ta powoduje, iż możliwe jest wprowadzenie prostego kryterium odróżniającego obrazy wypowiedzi wzorcowych od obrazów wypowiedzi patologicznych. Można mianowicie wykazać, że obrazy wypowiedzi prawidłowych charakteryzują się mniejszą długością krzywej oraz jej prostszą formą niż obrazy wypowiedzi patologicznych. Miara ta (tzn. zamiana długości trajektorii przy przechodzeniu od mowy poprawnej do patologicznej) będzie dalej wykorzystana do porównawczego charakteryzowania jakości wizualizacji uzyskiwanej przez różne sieci.

Opisany wyżej rezultat jest zgodny z teoretycznymi oczekiwaniami. Wynika to z następującego rozumowania. W fazie treningu sieci, neurony uczone są tego, że mają reagować (uzyskiwać status „zwycięzcy”) w odpowiedzi na wektory podobne do wektorów znajdujących się w zbiorze treningowym. Ponieważ zbiór ten utworzony jest wyłącznie z wypowiedzi wzorcowych, zatem w trakcie obrazowania wypowiedzi wzorcowych (także odmiennych od tych, które wchodziły w skład zbioru uczącego) na poszczególne fragmenty sygnału mowy reagują ściśle określone neurony (najlepiej dopa-

sowane do wektorów wejściowych). Sekwencja zjawisk fonetyczno-akustycznych składająca się na prawidłową wypowiedź znajduje zatem w wytrenowanej sieci sekwencję dobrze zdefiniowanych wzorców, które kolejno identyfikują nadchodzące elementy sygnału, tworząc tym samym w miarę regularny kształt budowanej krzywej.

W przypadku obrazowania wypowiedzi patologicznych jest wysoce prawdopodobne, że na wejście sieci podany zostanie wektor, który nie jest podobny do żadnego z wektorów ze zbioru treningowego. W takiej sytuacji zwycięzcą może zostać jakiś przypadkowy neuron, najczęściej spoza tej ściśle określonej grupy neuronów, dla których w poprawnej wypowiedzi istnieją konkretne odpowiedniki fonetyczne, co powoduje zniekształcenie, wydłużenie i skomplikowanie generowanej krzywej.

Pożądanym efektem, jaki jest następstwem opisanych wyżej mechanizmów, jest odmienny i względnie łatwy do interpretacji wygląd obrazu produkowanego przez sieć odpowiednio dla mowy prawidłowej i dla mowy patologicznej, przy czym – co warto podkreślić – większym (subiektywnie) patologicznym zniekształceniom sygnału mowy towarzyszą większe deformacje przebiegu wizualizowanej trajektorii. Efekt ten (większej złożoności kształtu krzywej) jest łatwiejszy do wzrokowej oceny i to właśnie on będzie zapewne podstawowym kryterium oceny wizualizowanego sygnału przez lekarzy, jednak dla obiektywnych porównań komputerowych wygodniej jest odwołać się do parametru pośrednio związanego z większą złożonością kształtu trajektorii, mianowicie do jej sumarycznej długości. Wprawdzie długość krzywej trudno jest dokładnie ocenić wizualnie, jednak cecha ta może być wyjątkowo łatwo wyznaczana obliczeniowo, a jednocześnie może posłużyć jako podstawa do stworzenia obiektywnej metody oceny generowanych w opisany powyżej sposób obrazów. Niektóre dodatkowe aspekty stosowania miary długości trajektorii przy automatycznej ocenie stopnia deformacji mowy patologicznej opisano we wcześniejszych pracach autora [Gaj99c, Kap00a, Kap00b], nie będą więc one tutaj dodatkowo dyskutowane.

Sugerowana metoda oceny złożoności trajektorii polega na pomiarze całkowitej długości krzywych reprezentujących poszczególne słowa, a następnie porównaniu wartości uzyskanych dla wypowiedzi wzorcowych i patologicznych. W tab. 5-1 do 5-5 zamieszczono średnie wartości długości krzywych otrzymane dla poszczególnych wyrazów i kategorii chorobowych w zależności od rozmiaru sieci. Dla każdego badanego przypadku wybierano indywidualnie liczbę kroków treningu, należy jednak podkreślić, że dla raz wytrenowanej sieci długości krzywych były obliczane dla wszystkich 5 kategorii wypowiedzi.

Przeprowadzone badania ponownie wykazały, że najlepsze rezultaty uzyskuje się gdy sieć jest trenowana nie za krótko (bo nie zdąży jeszcze zgromadzić wymaganej ilo-

ści wiadomości o naturze badanego sygnału) i nie za długo (bo grozi to „przeuczeniem"). Z przeprowadzonych badań wynika, że optymalna wartość długości procesu ucznia wynosi od 5000 do 20000 kroków. Niestety niemożliwe jest bardziej dokładne oszacowanie tego przedziału, ponieważ trening sieci jest (nawet przy ustalonym zbiorze uczącym) w pewnym sensie procesem niedeterministycznym (inicjalizacja wag sieci odbywa się wartościami losowymi oraz wektory treningowe wybierane są w losowej kolejności), co powoduje, że rezultaty uzyskane w kolejnych testach mogą się znacząco różnić między sobą. Wartości optymalnej długości procesu uczenia były także silnie zróżnicowane głównie ze względu na rodzaj wizualizowanego słowa, natomiast stwierdzono, że rozmiar sieci miał mniejszy wpływ na optymalną liczbę kroków treningu co było nieco zaskakujące w kontekście danych przytaczanych w literaturze przedmiotu.

	Liczba neuronów tworzących sieć Kohonena				
	25	**49**	**100**	**144**	**225**
Wypowiedź wzorcowa	48	85	92	187	234
Rozszczep podniebienia wtórnego częściowy	41	83	93	194	255
Rozszczep podniebienia wtórnego całkowity	54	110	116	207	265
Rozszczep podniebienia pierwotnego i wtórnego całkowity obustronny	53	102	97	223	275
Rozszczep podniebienia pierwotnego i wtórnego całkowity lewo- lub prawostronny	51	94	104	221	287

Tab. 5-1. Średnie długości krzywej w obrazach słowa „*rudy*".

	Liczba neuronów tworzących sieć Kohonena				
	25	**49**	**100**	**144**	**225**
Wypowiedź wzorcowa	27	42	66	91	108
Rozszczep podniebienia wtórnego częściowy	43	59	98	124	162
Rozszczep podniebienia wtórnego całkowity	49	69	110	140	174
Rozszczep podniebienia pierwotnego i wtórnego całkowity obustronny	45	60	112	143	177
Rozszczep podniebienia pierwotnego i wtórnego całkowity lewo- lub prawostronny	47	64	107	149	179

Tab. 5-2. Średnie długości krzywej w obrazach słowa „*pies*".

	Liczba neuronów tworzących sieć Kohonena				
	25	**49**	**100**	**144**	**225**
Wypowiedź wzorcowa	39	63	74	131	118
Rozszczep podniebienia wtórnego częściowy	44	84	108	164	138
Rozszczep podniebienia wtórnego całkowity	49	77	94	149	147
Rozszczep podniebienia pierwotnego i wtórnego całkowity obustronny	62	94	121	155	164
Rozszczep podniebienia pierwotnego i wtórnego całkowity lewo- lub prawostronny	49	74	107	156	158

Tab. 5-3. Średnie długości krzywej w obrazach słowa „*goni*".

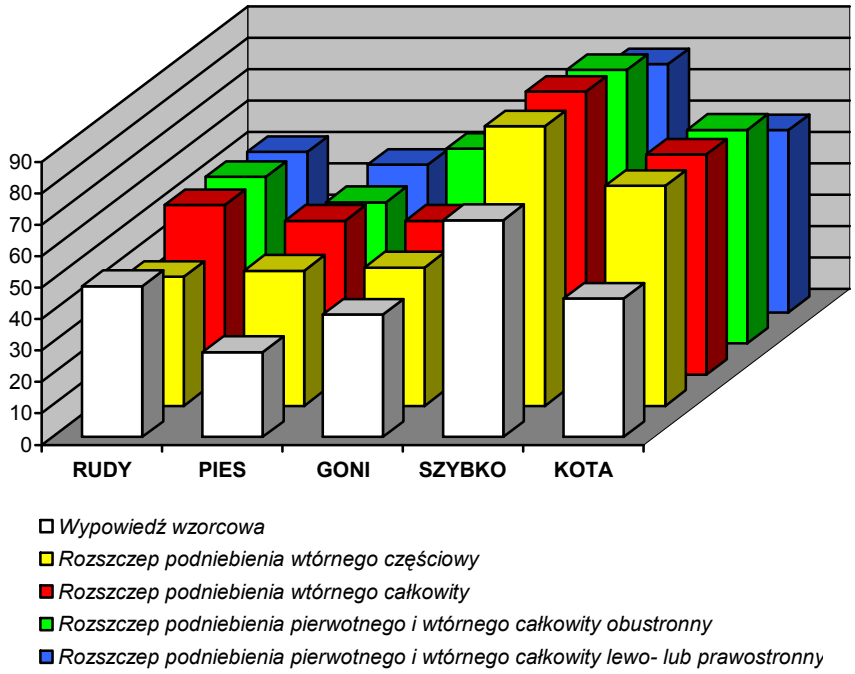
	Liczba neuronów tworzących sieć Kohonena				
	25	**49**	**100**	**144**	**225**
Wypowiedź wzorcowa	69	81	136	142	219
Rozszczep podniebienia wtórnego częściowy	89	126	155	228	311
Rozszczep podniebienia wtórnego całkowity	90	118	178	213	304
Rozszczep podniebienia pierwotnego i wtórnego całkowity obustronny	87	124	173	222	295
Rozszczep podniebienia pierwotnego i wtórnego całkowity lewo- lub prawostronny	79	107	164	200	251

Tab. 5-4. Średnie długości krzywej w obrazach słowa „*szybko*".

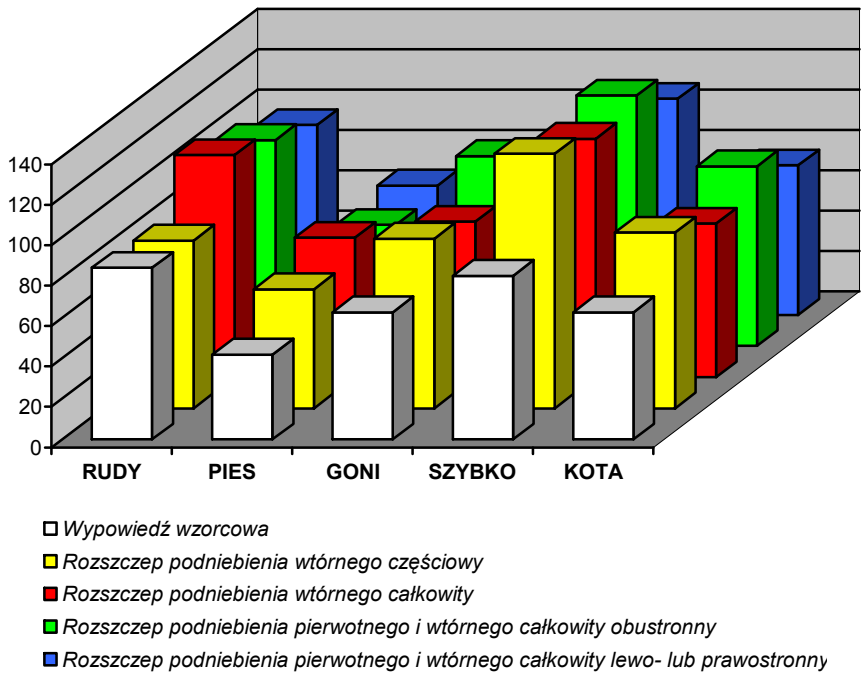
	Liczba neuronów tworzących sieć Kohonena				
	25	**49**	**100**	**144**	**225**
Wypowiedź wzorcowa	44	63	88	131	133
Rozszczep podniebienia wtórnego częściowy	70	87	149	200	227
Rozszczep podniebienia wtórnego całkowity	70	76	140	189	203
Rozszczep podniebienia pierwotnego i wtórnego całkowity obustronny	68	89	140	205	195
Rozszczep podniebienia pierwotnego i wtórnego całkowity lewo- lub prawostronny	58	74	118	169	166

Tab. 5-5. Średnie długości krzywej w obrazach słowa „*kota*".

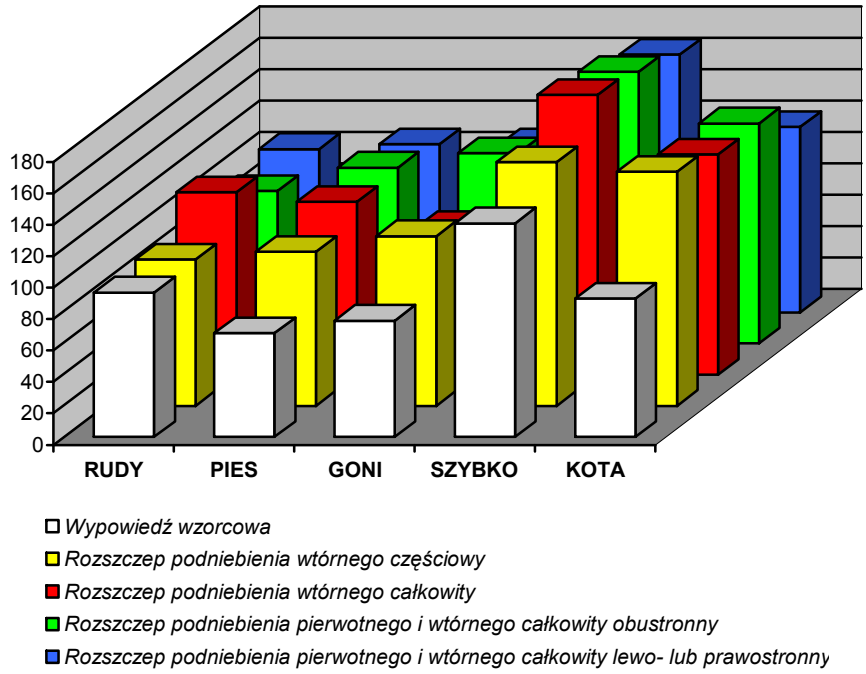
W celu ułatwienia Czytelnikowi interpretacji powyższych danych przedstawiono je w także postaci wykresów słupkowych – wykresy 5-1 do 5-5 prezentują dane z tabel 5-1 do 5-5 pogrupowane w zależności od liczby neuronów tworzących sieć Kohonena.



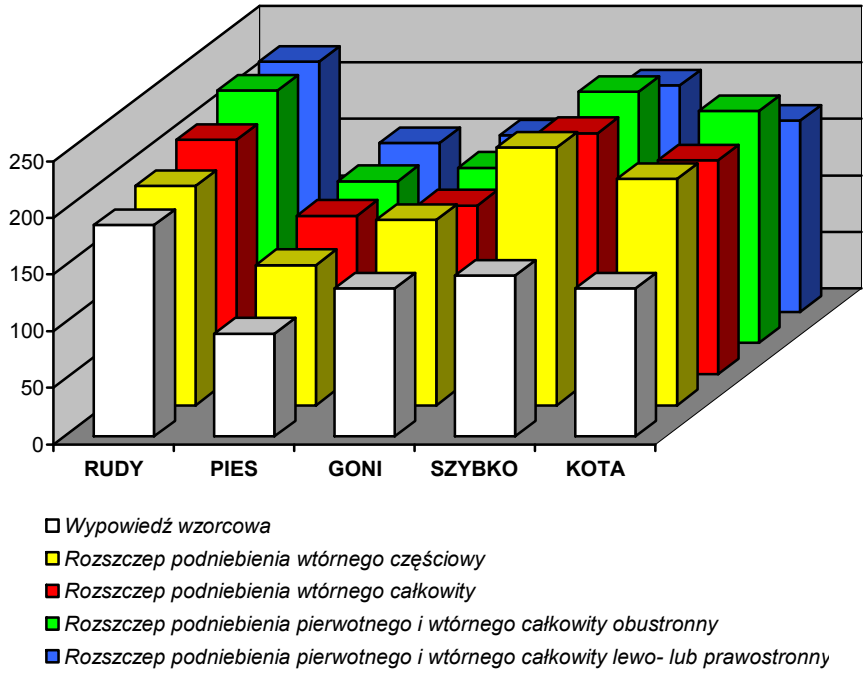
Wyk. 5-1. Średnie długości krzywej w obrazach badanych słów, dane dla sieci o rozmiarze 5 na 5 neuronów.



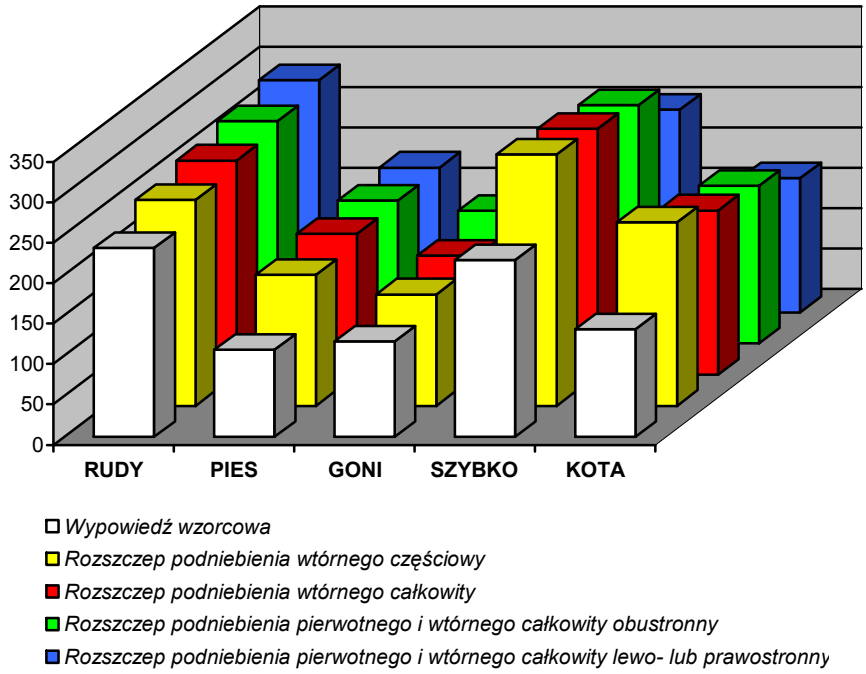
Wyk. 5-2. Średnie długości krzywej w obrazach badanych słów, dane dla sieci o rozmiarze 7 na 7 neuronów.



Wyk. 5-3. Średnie długości krzywej w obrazach badanych słów, dane dla sieci o rozmiarze 10 na 10 neuronów.



Wyk. 5-4. Średnie długości krzywej w obrazach badanych słów, dane dla sieci o rozmiarze 12 na 12 neuronów.



Wyk. 5-5. Średnie długości krzywej w obrazach badanych słów, dane dla sieci o rozmiarze 15 na 15 neuronów.

Analizując dane w tabelach 5-1 do 5-5 można zaobserwować, iż ze wzrostem rozmiaru sieci rośnie długość krzywych wyznaczanych w poszczególnych obrazach, co jest niejako konsekwencją zastosowania tu najprostszej metryki związanej bezpośrednio z odległością między neuronami w sieci w układzie przypisującym poszczególnym neuronom lokalizacje wyrażane całkowitoliczbowymi wartościami odpowiednich współrzędnych. Oczywiście można by było wprowadzić metryki znormalizowane, które pozwoliłyby lepiej porównywać ze sobą wyniki uzyskiwane dla różnych rozmiarów sieci, jednak nie wydawało się to celowe ze względu na to, że głównym przedmiotem analizy są różnice, jakie występują pomiędzy trajektoriami wizualizującymi sytuacje dla mowy prawidłowej i dla poszczególnych form patologii.

Przedstawione wyniki potwierdzają, iż średnia długość krzywej na obrazach wypowiedzi patologicznych jest dłuższa niż na obrazach wypowiedzi wzorcowych. Uzasadniono w ten sposób, że pomiar długości trajektorii łączącej zwycięskie neurony może być obiektywną metodą oceny złożoności generowanych obrazów. Jedynie dla sieci o rozmiarze 5 na 5 i 7 na 7 neuronów, dla słowa „*rudy*", występuje odstępstwo od tej reguły. Na podstawie tego faktu można uznać, iż sieć złożona z mniej niż 50 neuronów jest niezdolna do efektywnego obrazowania interesujących nas w tej pracy form patologicznych deformacji pojedynczych słów. Spostrzeżenie to jest zgodne z wnio-

skami wyciągniętymi już wcześniej na podstawie wizualnej obserwacji formy obrazów – stwierdzono wtedy, że zwycięska trajektoria w takich obrazach pokrywa znaczną część powierzchni obrazu, co w znacznym stopniu utrudnia wizualną ich ocenę, czyniąc taką postać zobrazowania mało przydatną w kontekście rozważanych tu zadań.

5.2. Obrazy izolowanych głosek

Po serii badań dotyczących wizualizacji z użyciem sieci Kohonena obrazów wypowiedzi (prawidłowych i patologicznych) całych słów, podjęto także badania mające na celu sprawdzenie, czy na podobnej zasadzie nie można by było charakteryzować i wizualizować mniejszych fragmentówsygnału mowy. Oczekiwano przy tym, że obrazy uzyskiwane poprzez wizualizację pojedynczych głosek (oczywiście odpowiednio dobranych do badanego w pracy problemu) będą na tyle prostsze od wcześniej rozważanych, że pozwoli to na lepsze różnicowanie zapisów prawidłowych i patologicznych oraz na ich łatwiejszą interpretację (jakościową, na podstawie kształtu, oraz ilościową, na podstawie długości krzywej). Opisane niżej wyniki badań potwierdziły te oczekiwania.

5.2.1. Sieć bez sąsiedztwa

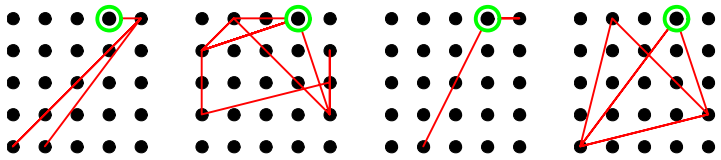
W opisanym wyżej obrazowaniu pojedynczych słów zastosowanie sieci Kohonena bez sąsiedztwa dało znacznie lepsze rezultaty niż sieć z sąsiedztwem, zatem w pierwszej kolejności zbadano możliwość zastosowania właśnie sieci bez sąsiedztwa do obrazowania pojedynczych głosek. W rozdziale tym zaprezentowane zostaną uzyskiwane z pomocą sieci Kohonena obrazy pojedynczych głosek „*s*” i „*sz*” wyodrębnionych z wypowiedzi słów „*pies*” oraz „*szybko*”. Badania przeprowadzono dla czterech rozmiarów sieci: 25, 100, 225 oraz 400 neuronów. Każda badana sieć posiadała, tak jak poprzednio, topologię kwadratu.

W trakcie relacjonowanych tu badań zaobserwowano bardzo ciekawy efekt, związany z obrazowaniem wypowiedzi wzorcowych. Mianowicie w ponad 80% przypadków treningu sieci (bez względu na rozmiar sieci oraz liczbę kroków treningu) wszystkie obrazy wypowiedzi wzorcowych reprezentowane były przez **jeden** wspólny punkt. Jest to logiczne – przy prawidłowej artykulacji obraz fonetyczno-akustyczny jednej głoski powinien być w istocie jednakowy, zatem w sieci Kohonena głosce takiej odpowiada jeden neuron, będący zwycięzcą dla wszystkich próbek sygnału (dla wszystkich widm chwilowych) wchodzących w obręb tej jednej głoski. Prawidłowości tej pozba-

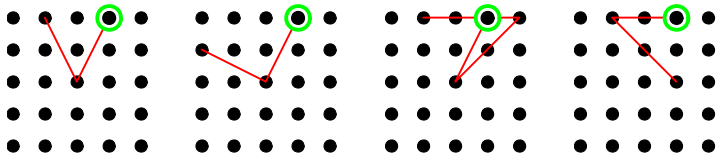
wione są wypowiedzi patologiczne i jest to cecha umożliwiającaich natychmiastowe rozpoznawanie.

Poniżej zaprezentowane zostaną obrazy wypowiedzi patologicznych uzyskane przy pomocy sieci wytrenowanych w powyższy sposób. Nie zamieszczono przy tym obrazów wypowiedzi wzorcowych ponieważ, tak jak wspomniano, wszystkie te obrazy (w obrębie jednej grupy) reprezentowane były przez jeden punkt. Na obrazach trajektorii uzyskanych dla wypowiedzi patobgicznych za pomocą kółka ◉ zaznaczono położenie tego pojedynczego punktu reprezentującego wypowiedzi wzorcowe.

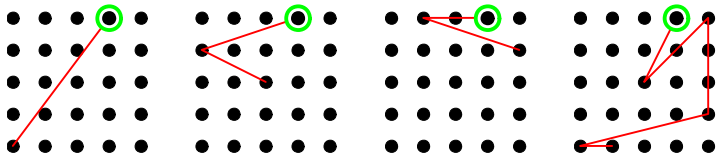
Na rys. 5-51 do 5-54 przedstawiono obrazy głoski „*s*", uzyskane przy pomocy sieci składającej się z 25 neuronów trenowanej przez 200 kroków.



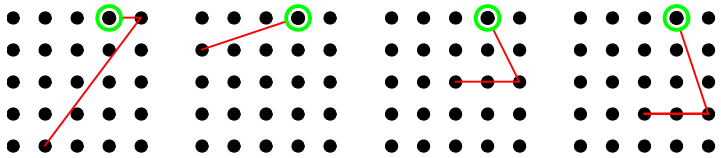
Rys. 5-51. Obrazy głoski „*s*" – rozszczep podniebienia wtórnego częściowy.



Rys. 5-52. Obrazy głoski „*s*" – rozszczep podniebienia wtórnego całkowity.

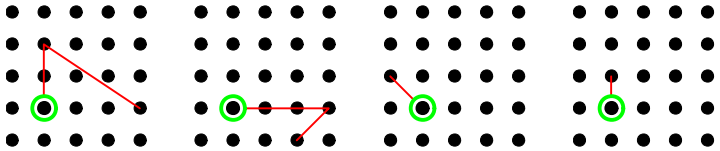


Rys. 5-53. Obrazy głoski „*s*" – rozszczep podniebienia pierwotnego i wtórnego całkowity obustronny.

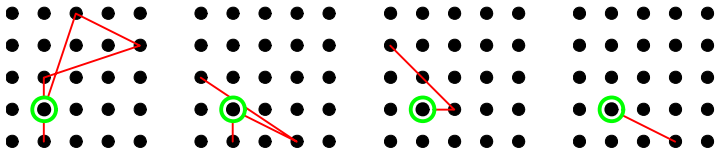


Rys. 5-54. Obrazy głoski „*s*" – rozszczep podniebienia pierwotnego i wtórnego całkowity lewostronny lub prawostronny.

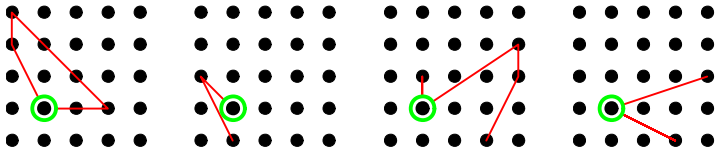
Kolejne rysunki przedstawiają obrazy głoski „*sz*” wygenerowane przez sieć o rozmiarze 5 na 5 neuronów trenowaną przez 300 kroków.



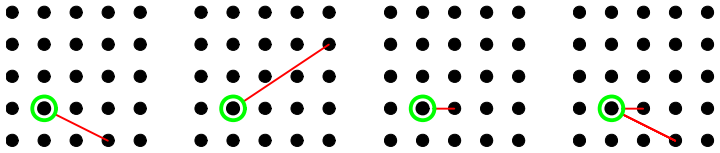
Rys. 5-55. Obrazy głoski „*sz*” – rozszczep podniebienia wtórnego częściowy.



Rys. 5-56. Obrazy głoski „*sz*” – rozszczep podniebienia wtórnego całkowity.

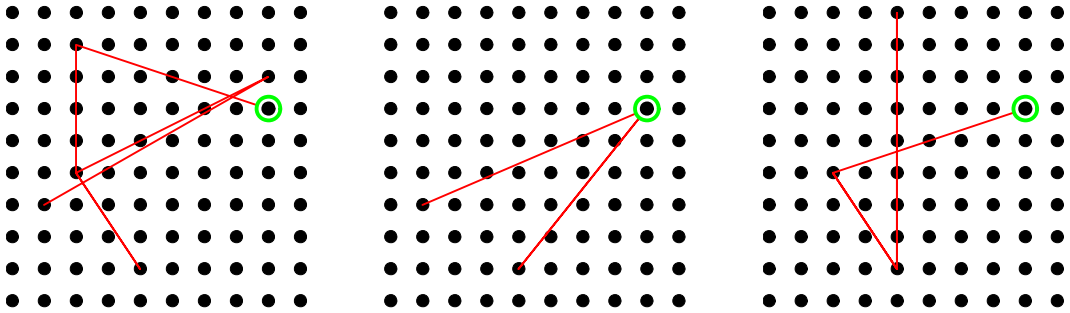


Rys. 5-57. Obrazy głoski „*sz*” – rozszczep podniebienia pierwotnego i wtórnego całkowity obustronny.

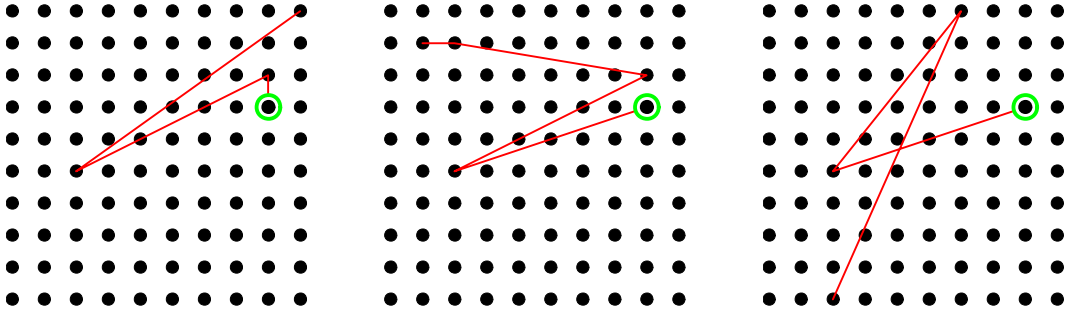


Rys. 5-58. Obrazy głoski „*sz*” – rozszczep podniebienia pierwotnego i wtórnego całkowity lewostronny lub prawostronny.

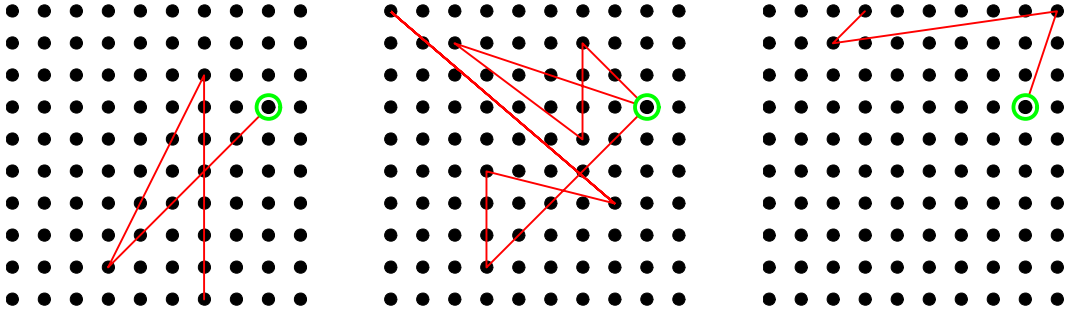
Następną grupę stanowią analogiczne obrazy uzyskane przy pomocy sieci składającej się ze 100 neuronów. Obrazy głoski „*s*” uzyskano przy pomocy sieci trenowanej przez 400 kroków, a obrazy głoski „*sz*” przy pomocy sieci trenowanej przez 200 kroków.



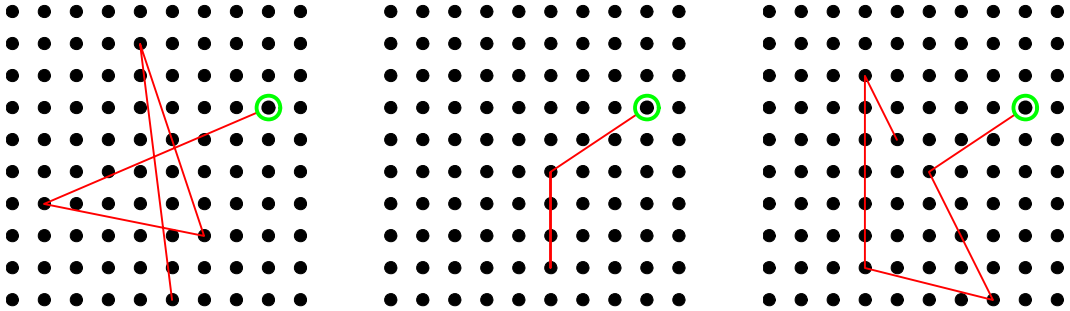
Rys. 5-59. Obrazy głoski „*s*" – rozszczep podniebienia wtórnego częściowy.



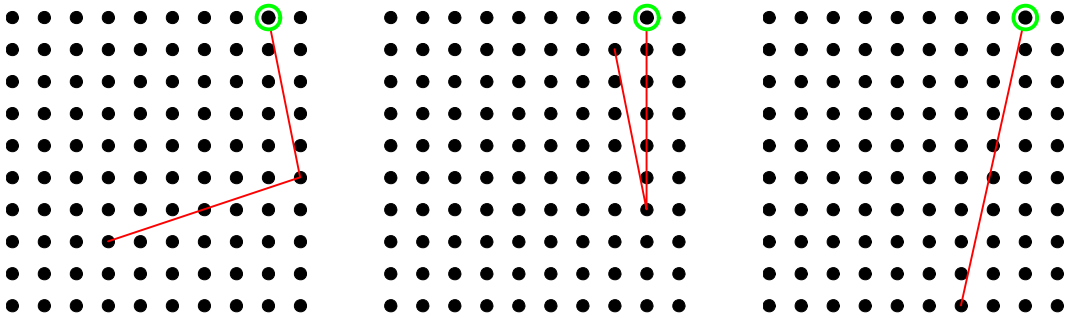
Rys. 5-60. Obrazy głoski „*s*" – rozszczep podniebienia wtórnego całkowity.



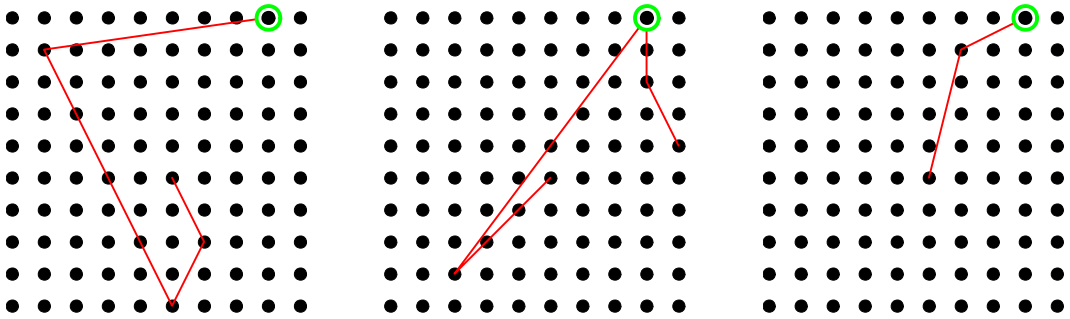
Rys. 5-61. Obrazy głoski „*s*" – rozszczep podniebienia pierwotnego i wtórnego całkowity obustronny.



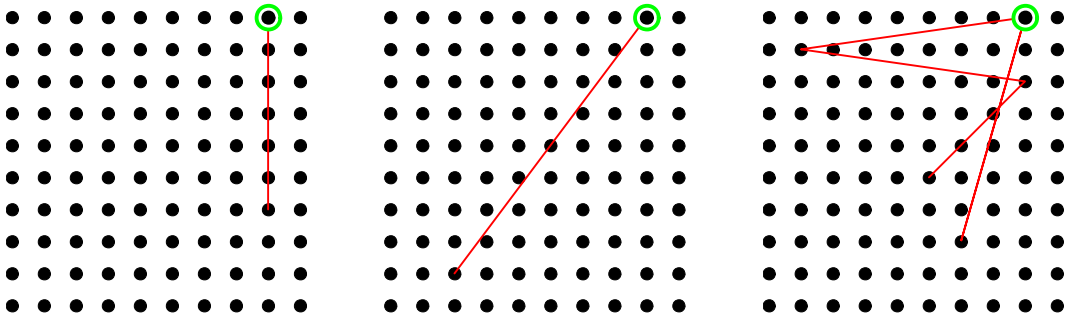
Rys. 5-62. Obrazy głoski „*s*" – rozszczep podniebienia pierwotnego i wtórnego całkowity lewostronny lub prawostronny.



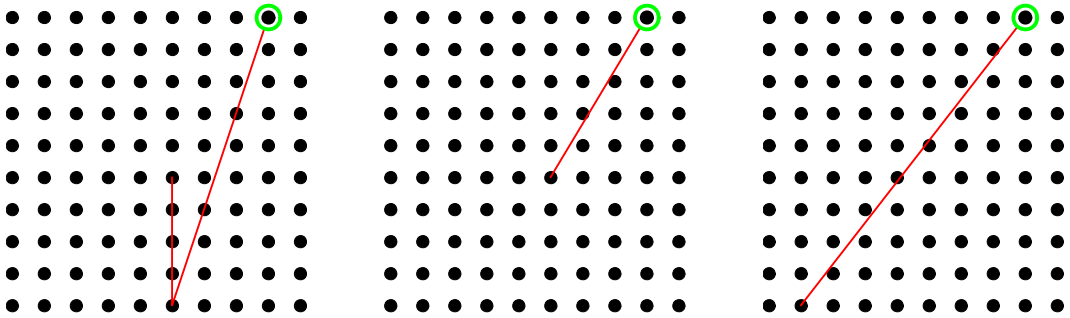
Rys. 5-63. Obrazy głoski „*sz*" – rozszczep podniebienia wtórnego częściowy.



Rys. 5-64. Obrazy głoski „*sz*" – rozszczep podniebienia wtórnego całkowity.

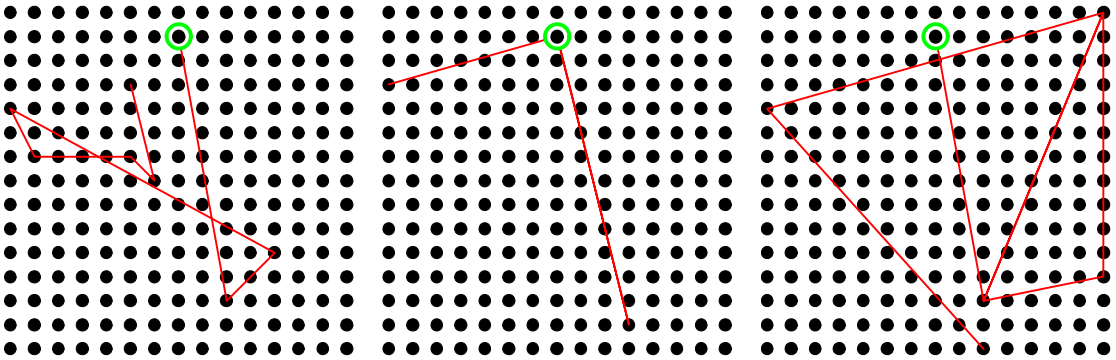


Rys. 5-65. Obrazy głoski „*sz*" – rozszczep podniebienia pierwotnego i wtórnego całkowity obustronny.

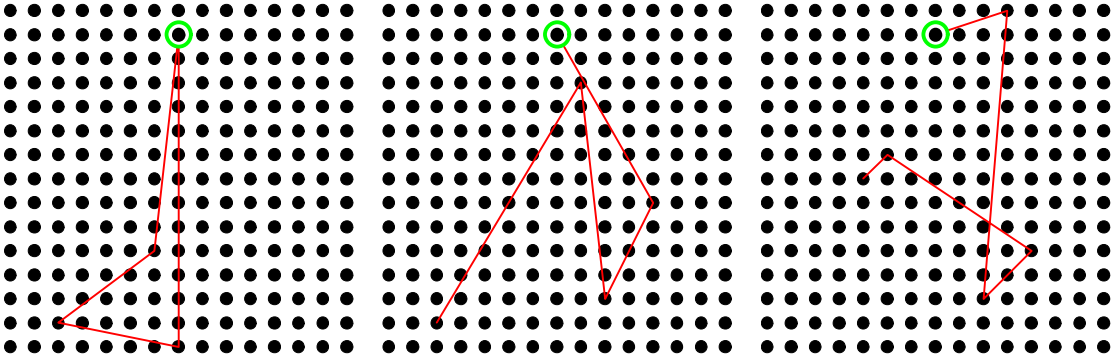


Rys. 5-66. Obrazy głoski „*sz*" – rozszczep podniebienia pierwotnego i wtórnego całkowity lewostronny lub prawostronny.

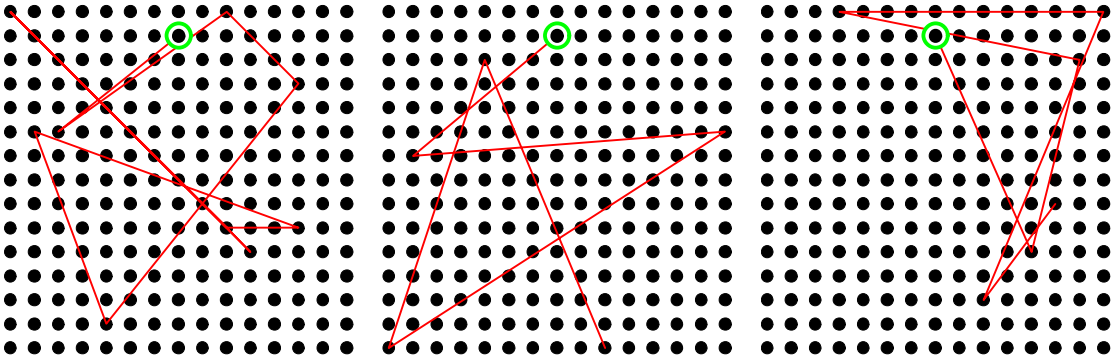
Na rys. 5-67 do 5-74 przedstawiono obrazy uzyskane przy pomocy sieci składającej się z 225 neuronów. W przypadku obrazów głoski „*s*" sieci trenowano przez 400 kroków, natomiast w przypadku głoski „*sz*" przez 200 kroków.



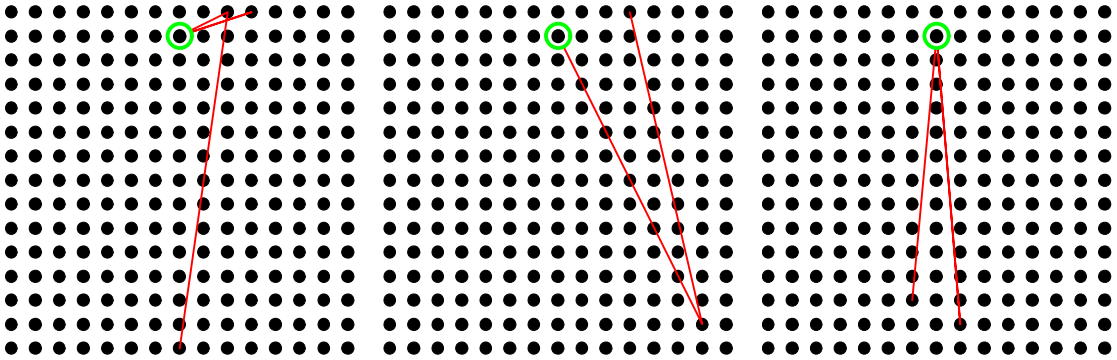
Rys. 5-67. Obrazy głoski „*s*" – rozszczep podniebienia wtórnego częściowy.



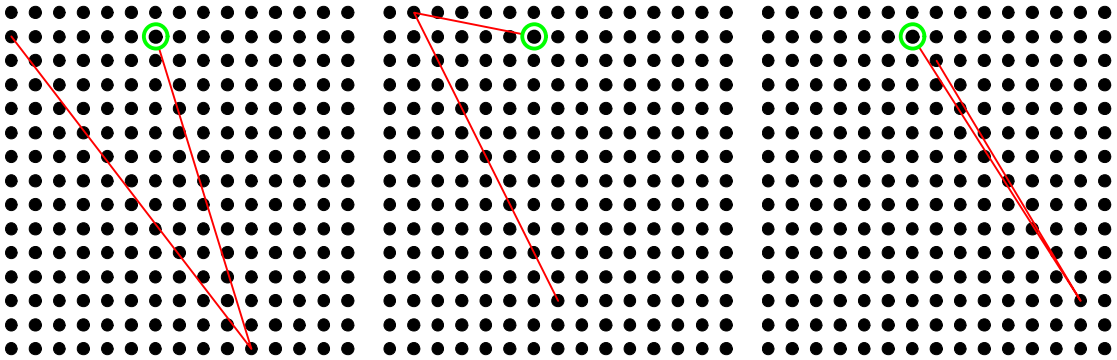
Rys. 5-68. Obrazy głoski „*s*" – rozszczep podniebienia wtórnego całkowity.



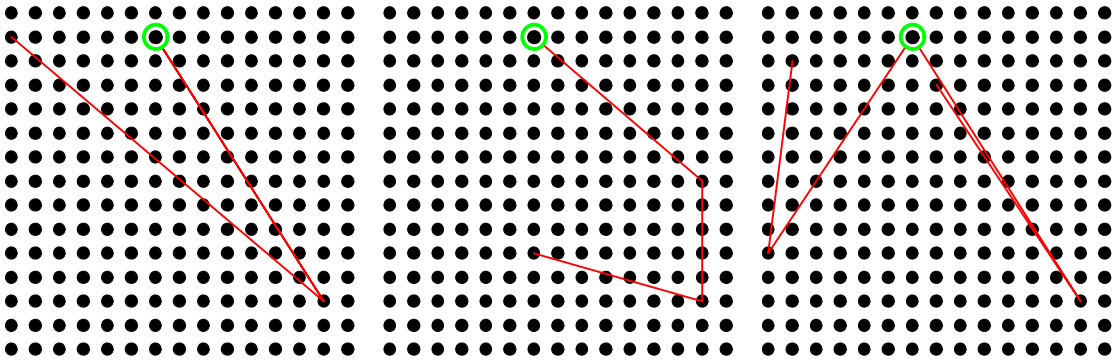
Rys. 5-69. Obrazy głoski „*s*" – rozszczep podniebienia pierwotnego i wtórnego całkowity obustronny.



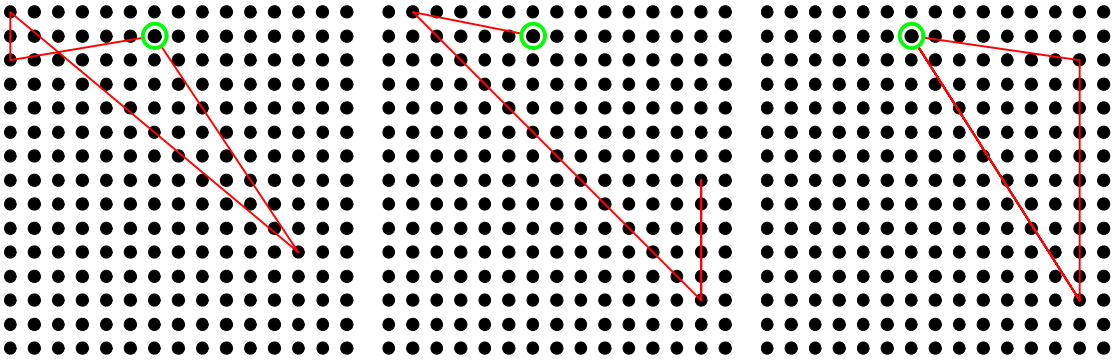
Rys. 5-70. Obrazy głoski „*s*" – rozszczep podniebienia pierwotnego i wtórnego całkowity lewostronny lub prawostronny.



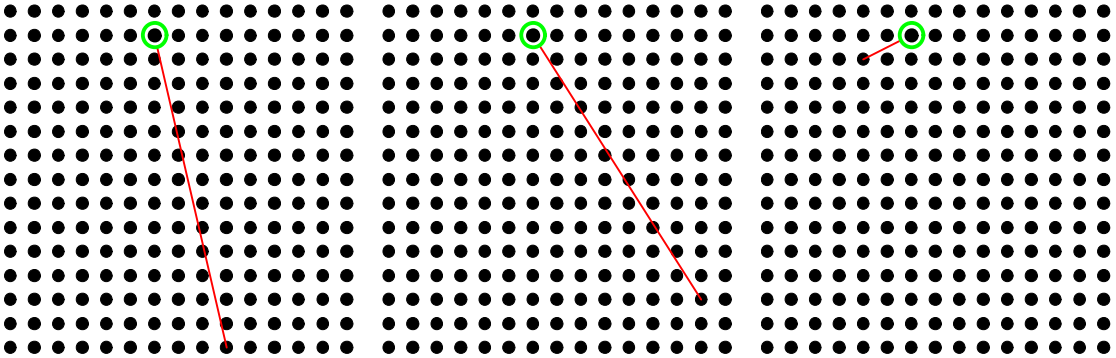
Rys. 5-71. Obrazy głoski „*sz*" – rozszczep podniebienia wtórnego częściowy.



Rys. 5-72. Obrazy głoski „*sz*" – rozszczep podniebienia wtórnego całkowity.

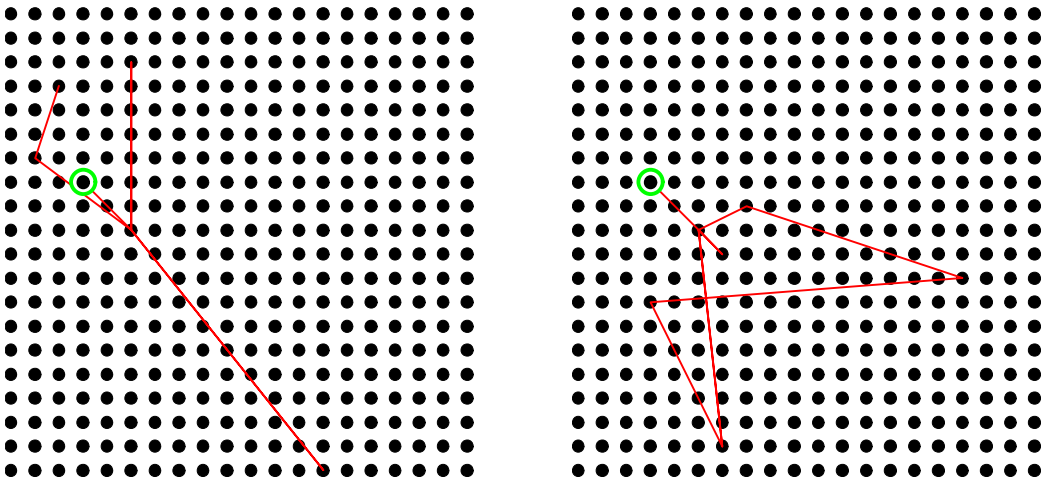


Rys. 5-73. Obrazy głoski „*sz*” – rozszczep podniebienia pierwotnego i wtórnego całkowity obustronny.

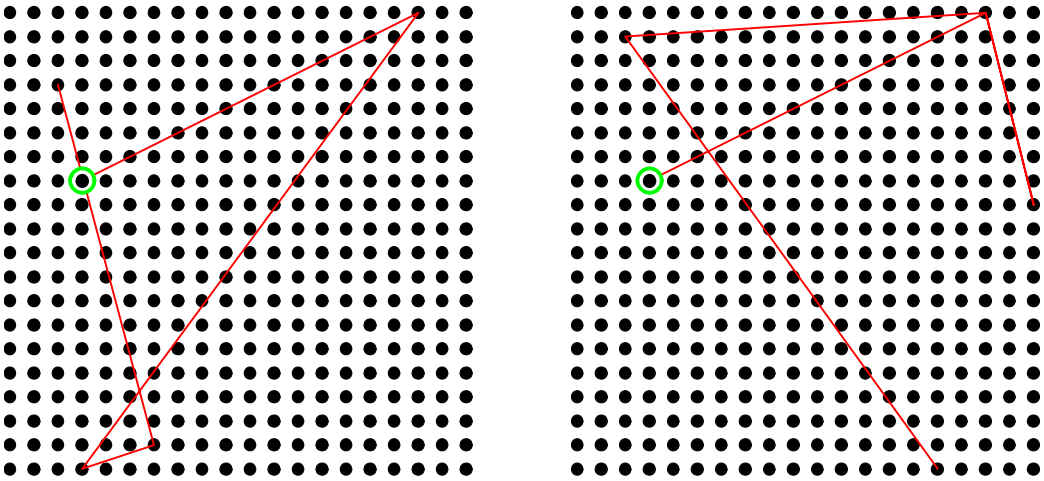


Rys. 5-74. Obrazy głoski „*sz*” – rozszczep podniebienia pierwotnego i wtórnego całkowity lewostronny lub prawostronny.

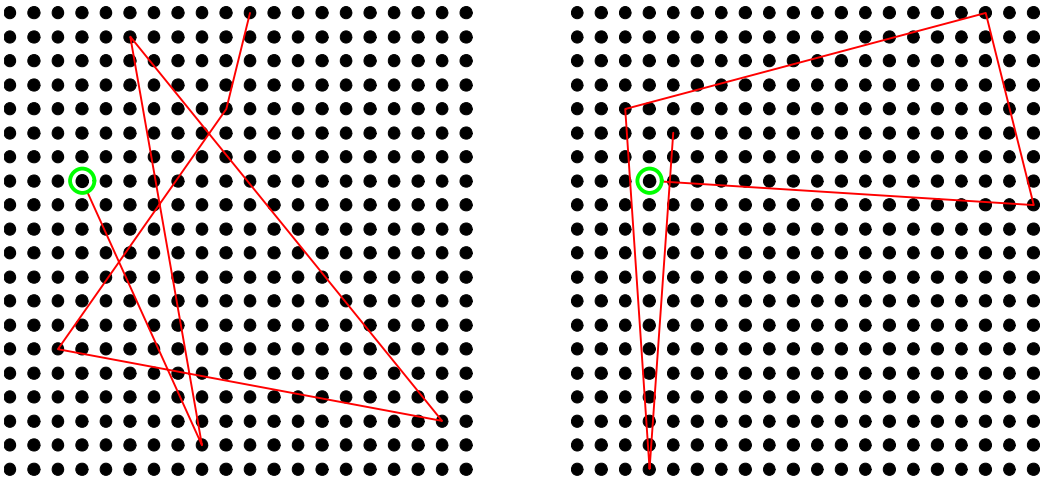
Ostatnią prezentowaną grupę stanowią obrazy głosek „*s*” i „*sz*” uzyskane przy pomocy sieci o rozmiarze 20 na 20 neuronów, trenowanej odpowiednio przez 300 i 500 kroków.



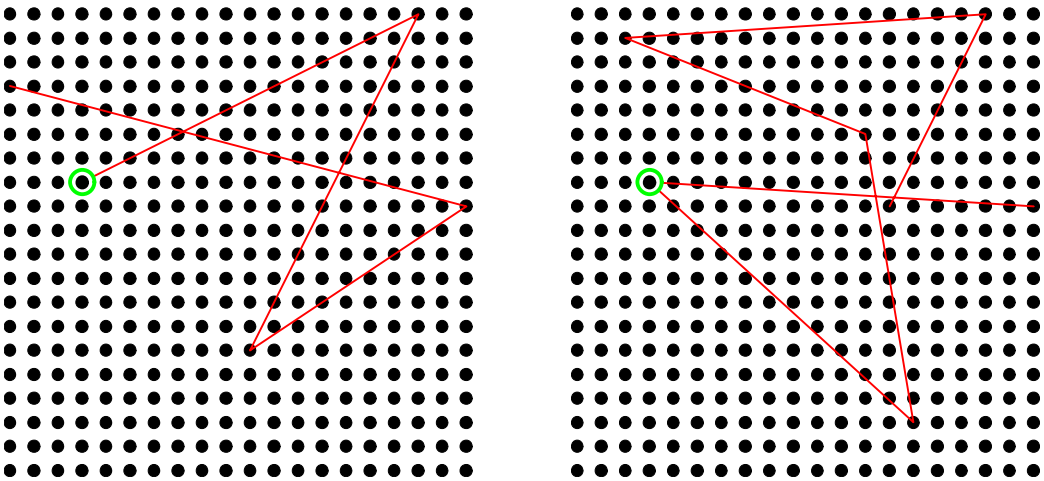
Rys. 5-75. Obrazy głoski „*s*” – rozszczep podniebienia wtórnego częściowy.



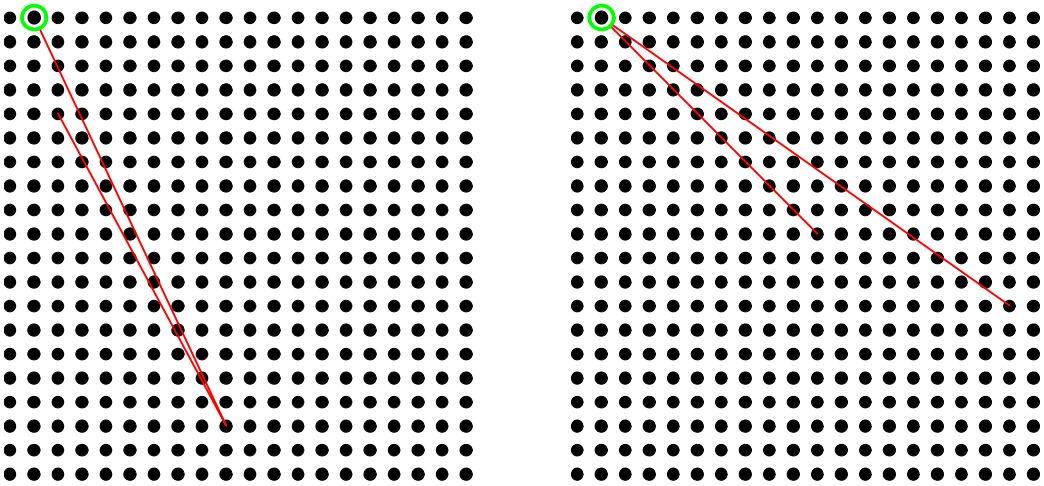
Rys. 5-76. Obrazy głoski „*s*” – rozszczep podniebienia wtórnego całkowity.



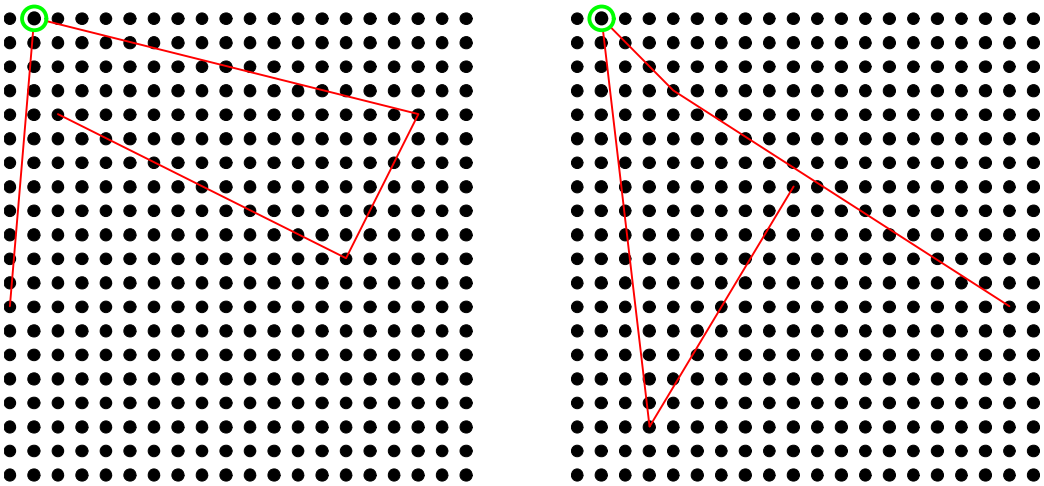
Rys. 5-77. Obrazy głoski „*s*” – rozszczep podniebienia pierwotnego i wtórnego całkowity obustronny.



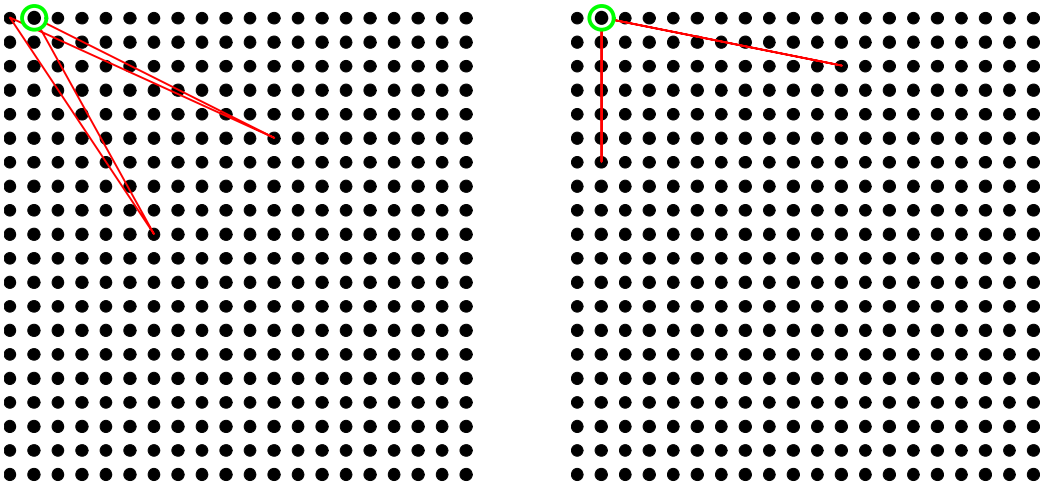
Rys. 5-78. Obrazy głoski „*s*” – rozszczep podniebienia pierwotnego i wtórnego całkowity lewostronny lub prawostronny.



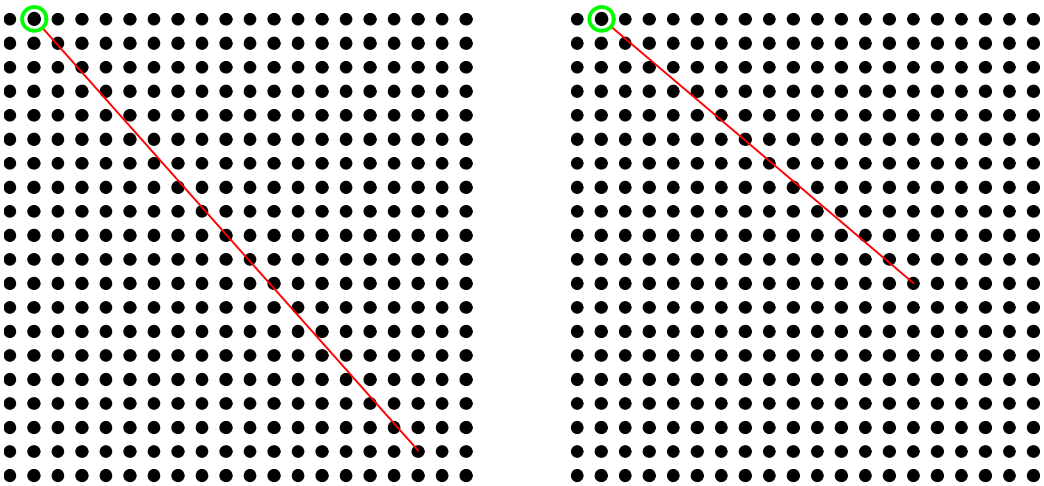
Rys. 5-79. Obrazy głoski „*sz*" – rozszczep podniebienia wtórnego częściowy.



Rys. 5-80. Obrazy głoski „*sz*" – rozszczep podniebienia wtórnego całkowity.



Rys. 5-81. Obrazy głoski „*sz*" – rozszczep podniebienia pierwotnego i wtórnego całkowity obustronny.



Rys. 5-82. Obrazy głoski „*sz*" – rozszczep podniebienia pierwotnego i wtórnego całkowity lewostronny lub prawostronny.

5.2.2. Dyskusja uzyskanych wyników dla izolowanych głosek w sieci bez sąsiedztwa

Obserwując powyższe obrazy można dojść do wniosku, że rezultaty uzyskane dla pojedynczych głosek są znacznie lepsze niż w przypadku obrazowania całych wyrazów. Na podstawie tych obrazów możliwe jest, w bardzo prosty sposób, wizualne odróżnienie wypowiedzi patologicznych od wzorcowych. Należy jednak obiektywnie stwierdzić, że nie wszystkie obrazy wypowiedzi patologicznych reprezentowane były przez krzywą, niektóre obrazy były identyczne z obrazami wypowiedzi wzorowych. Podobnego efektu nie odnotowano nigdy w odniesieniu do obrazów całych słów, natomiast pojawia się on niekiedy w odniesieniu do pojedynczych głosek. Wynika to jednak z prostego i oczywistego faktu. Wywołana czynnikiem morfologicznym (rozszczep podniebienia) patologia sygnału mowy badanych dzieci ujawnia się niezawodnie przy złożonym i trudnym zadaniu artykulacyjnym, jakim jest zadanie wypowiedzenia całego słowa. Natomiast przy artykulacji poszczególnych głosek może dochodzić do działań kompensacyjnych, w wyniku których jakość sygnału niektórych prób pojedynczej głoski, wypowiadanej przez dziecko z wadą wymowy, może nie odbiegać znacząco od standardu, pomimo zdecydowanie patologicznego charakteru całej wypowiedzi, w której ta głoska jest osadzona. Najlepiej ten fakt obrazują wyniki przeprowadzonych badań, które zamieszczono w tabelach 5-6 do 5-13. Dla każdego rozmiaru sieci i wielkości liczby kroków treningu dokonano kilkudziesięciu treningów sieci, zamieszczając w tabelach najlepsze z uzyskanych rezułatów. Podejście takie wydaje się być całkowicie

usprawiedliwione, ponieważ odpowiada ono przypadkowi treningu sieci pod kątem rozpoznawania jednej, konkretnej patologii. Należy zaznaczyć, iż rozpatrywano jedynie takie przypadki, w których obrazy wszystkich wypowiedzi wzorcowych były reprezentowane przez jeden wspólny punkt.

	Kroki treningu							
	100	**200**	**300**	**400**	**500**	**1000**	**1500**	**2000**
25 neuronów	92%	92%	77%	77%	85%	85%	77%	85%
100 neuronów	92%	92%	92%	92%	92%	92%	92%	92%
225 neuronów	100%	92%	92%	100%	85%	92%	100%	100%
400 neuronów	100%	100%	100%	100%	92%	100%	92%	100%

Tab. 5-6. Procentowa ilość obrazów głoski „*s*" reprezentowanych przez krzywą, dane dla rozszczepu podniebienia wtórnego częściowego.

	Kroki treningu							
	100	**200**	**300**	**400**	**500**	**1000**	**1500**	**2000**
25 neuronów	80%	80%	80%	80%	80%	80%	80%	80%
100 neuronów	87%	87%	80%	80%	87%	87%	80%	80%
225 neuronów	80%	80%	87%	80%	80%	87%	87%	80%
400 neuronów	80%	80%	87%	87%	80%	80%	80%	80%

Tab. 5-7. Procentowa ilość obrazów głoski „*s*" reprezentowanych przez krzywą, dane dla rozszczepu podniebienia wtórnego całkowitego.

	Kroki treningu							
	100	**200**	**300**	**400**	**500**	**1000**	**1500**	**2000**
25 neuronów	100%	93%	100%	93%	93%	93%	93%	93%
100 neuronów	100%	100%	100%	100%	100%	100%	100%	100%
225 neuronów	100%	100%	100%	100%	100%	100%	100%	100%
400 neuronów	100%	100%	100%	100%	100%	100%	100%	100%

Tab. 5-8. Procentowa ilość obrazów głoski „*s*" reprezentowanych przez krzywą, dane dla rozszczepu podniebienia pierwotnego i wtórnego całkowitego obustronnego.

	Kroki treningu							
	100	**200**	**300**	**400**	**500**	**1000**	**1500**	**2000**
25 neuronów	100%	100%	100%	100%	100%	100%	100%	100%
100 neuronów	100%	100%	100%	100%	100%	100%	100%	100%
225 neuronów	100%	100%	100%	100%	100%	100%	100%	100%
400 neuronów	100%	100%	100%	100%	100%	100%	100%	100%

Tab. 5-9. Procentowa ilość obrazów głoski „*s*" reprezentowanych przez krzywą, dane dla rozszczepu podniebienia pierwotnego i wtórnego całkowitego lewostronnego lub prawostronnego.

	Kroki treningu							
	100	**200**	**300**	**400**	**500**	**1000**	**1500**	**2000**
25 neuronów	54%	62%	62%	54%	62%	62%	62%	62%
100 neuronów	69%	69%	69%	69%	69%	69%	62%	62%
225 neuronów	69%	69%	69%	69%	69%	69%	69%	69%
400 neuronów	69%	69%	69%	69%	69%	69%	69%	69%

Tab. 5-10. Procentowa ilość obrazów głoski „*sz*" reprezentowanych przez krzywą, dane dla rozszczepu podniebienia wtórnego częściowego.

	Kroki treningu							
	100	**200**	**300**	**400**	**500**	**1000**	**1500**	**2000**
25 neuronów	73%	80%	80%	80%	80%	80%	80%	80%
100 neuronów	87%	87%	87%	87%	87%	87%	93%	80%
225 neuronów	93%	87%	87%	93%	93%	87%	87%	87%
400 neuronów	93%	87%	93%	100%	100%	87%	87%	93%

Tab. 5-11. Procentowa ilość obrazów głoski „*sz*" reprezentowanych przez krzywą, dane dla rozszczepu podniebienia wtórnego całkowitego.

	Kroki treningu							
	100	**200**	**300**	**400**	**500**	**1000**	**1500**	**2000**
25 neuronów	53%	53%	53%	53%	53%	53%	53%	53%
100 neuronów	53%	60%	60%	67%	53%	53%	53%	53%
225 neuronów	60%	67%	60%	60%	60%	53%	53%	53%
400 neuronów	67%	60%	67%	60%	60%	60%	60%	53%

Tab. 5-12. Procentowa ilość obrazów głoski „*sz*" reprezentowanych przez krzywą, dane dla rozszczepu podniebienia pierwotnego i wtórnego całkowitego obustronnego.

	Kroki treningu							
	100	200	300	400	500	1000	1500	2000
25 neuronów	46%	46%	46%	46%	46%	54%	46%	46%
100 neuronów	62%	77%	69%	62%	69%	53%	53%	53%
225 neuronów	77%	69%	69%	69%	77%	69%	62%	62%
400 neuronów	77%	77%	69%	77%	77%	69%	69%	69%

Tab. 5-13. Procentowa ilość obrazów głoski „*sz*" reprezentowanych przez krzywą, dane dla rozszczepu podniebienia pierwotnego i wtórnego całkowitego lewostronnego lub prawostronnego.

Analizując powyższe tabele można dojść do wniosku, że najlepsze efekty obrazowania uzyskuje się dla sieci trenowanej od 100 do 500 kroków. Zwiększenie liczby kroków treningu powyżej 500 nie powoduje wyraźnej poprawy wyników, a w większości przypadków pogarsza je (być może efekt ten można wiązać z "przeuczeniem" sieci). Kolejną zaobserwowaną prawidłowością jest poprawa wyników wraz ze wzrostem rozmiaru sieci. Szczególnie widoczne jest to na przykładzie danych dla 25 i 100 neuronów. Jednak podobnie jak w przypadku długości treningu korzyści ze wzrostu wybranego parametru mają charakter ograniczony – po osiągnięciu pewnej wartości granicznej dalszy wzrost rozmiaru sieci nie powoduje znaczącej poprawy wyników.

Przedstawione w tabelach dane można interpretować jako miary skuteczności dyskryminacji wypowiedzi z konkretnej grupy patologicznej. Wprawdzie widać wyraźną dysproporcję pomiędzy wynikami uzyskanymi dla głoski „*s*" i „*sz*", jednak ogólne rezultaty są zadawalające. Łatwo przy tym zauważyć, że na podstawie obrazów tych dwóch głosek możliwa jest dyskryminacja wypowiedzi patologicznych ze 100% skutecznością – dla rozszczepu podniebienia wtórnego całkowitego dokonywana jest ona na podstawie obrazów głoski „*sz*", natomiast dla wszystkich pozostałych przypadków na podstawie obrazów głoski „*s*" [Kap00c].

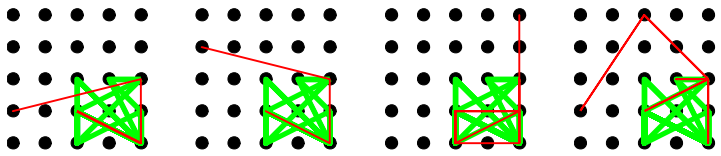
5.2.3. Sieć z wprowadzonym sąsiedztwem

W poprzednim podrozdziale przedstawiono rezultaty obrazowania izolowanych głosek, uzyskane przy pomocy sieci Kohonena **bez sąsiedztwa**. Analizując te dane stwierdzono, że wyniki uzyskane dla wypowiedzi głoski „*sz*" nie są zadawalające. Prawdopodobnie spowodowane jest to faktem, iż sieć dokonuje zbyt dużego uogólnienia nauczonego wzorca, w wyniku czego bardzo często wszystkie wektory z pojedynczej wypowiedzi patologicznej rozpoznawane są także przez pojedynczy neuron, co

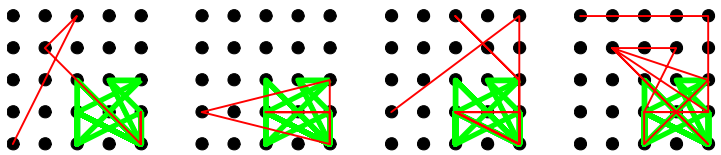
obniża czułość i selektywność metody. Postawiono hipotezę, że rozwiązaniem tego problemu może być zastosowanie sieci Kohonena z wprowadzonym sąsiedztwem. W paragrafie tym przedstawiono obrazy uzyskane przy weryfikacji tej hipotezy. Wszystkie obrazy uzyskano przy pomocy odpowiednio uczonych sieci Kohonena z funkcją sąsiedztwa opisaną wzorem (4.2).

Tak jak poprzednio obrazy podzielone zostały, ze względu na rozmiar sieci, na 4 grupy. Uzyskane w ten sposób obrazy wypowiedzi wzorcowych charakteryzowały się tym, iż nie były już (niestety) zredukowane do pojedynczego punktu, lecz ze względu na wyraźnie większą selektywność sieci z sąsiedztwem – miały również kształt pewnej trajektorii o niezerowej długości. Niemniej na skutek tego, że różnorodność sygnałów mowy prawidłowej jest wyraźnie mniejsza od różnorodności mowy patologicznej – obserwowano, że trajektorie te były skupione w jednym, małym obszarze sieci. Ponieważ niemożliwe było w tej sytuacji wyznaczenie wspólnego kształtu dla wszystkich obrazów wypowiedzi wzorcowych, zdecydowano się na przedstawienie obrazów wypowiedzi patologicznych na tle **sumarycznego** obrazu wypowiedzi wzorcowych. Taki sumaryczny obraz powstał poprzez nałożenie na siebie wszystkich obrazów wypowiedzi wzorcowych, na prezentowanych obrazach zaznaczono go przy pomocy zielonej linii. Obrazy takie są prezentowane poniżej.

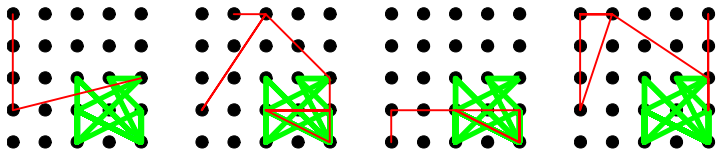
Pierwszą grupę stanowią obrazy uzyskane przy pomocy sieci o rozmiarze 5 na 5 neuronów. Rys. 5-83 do 5-86 przedstawiają obrazy głoski „*s*” uzyskane przy pomocy sieci trenowanej przez 500 kroków, natomiast na rys. 5-87 do 5-90 zamieszczono obrazy głoski „*sz*” uzyskane po 300 krokach treningu



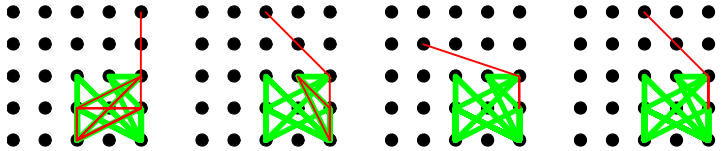
Rys. 5-83. Obrazy głoski „*s*” – rozszczep podniebienia wtórnego częściowy.



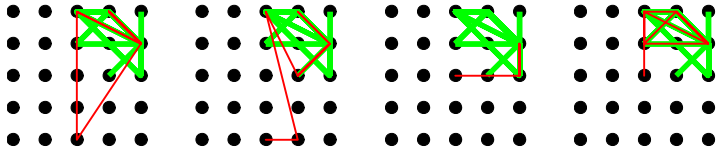
Rys. 5-84. Obrazy głoski „*s*” – rozszczep podniebienia wtórnego całkowity.



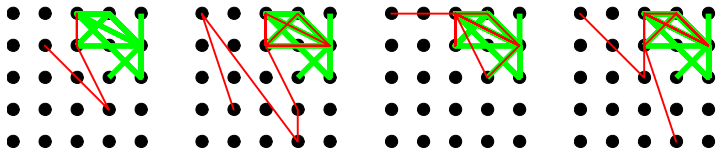
Rys. 5-85. Obrazy głoski „*s*” – rozszczep podniebienia pierwotnego i wtórnego całkowity obustronny.



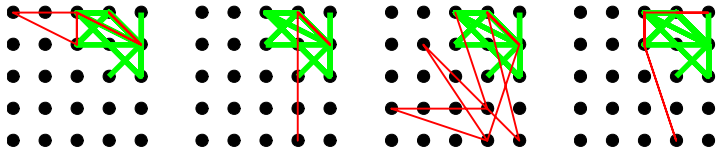
Rys. 5-86. Obrazy głoski „*s*” – rozszczep podniebienia pierwotnego i wtórnego całkowity lewostronny lub prawostronny.



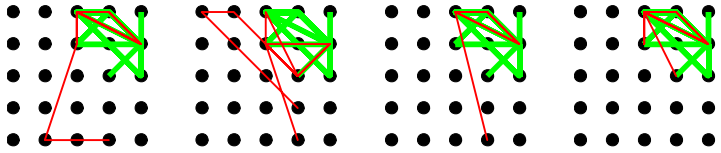
Rys. 5-87. Obrazy głoski „*sz*” – rozszczep podniebienia wtórnego częściowy.



Rys. 5-88. Obrazy głoski „*sz*” – rozszczep podniebienia wtórnego całkowity.

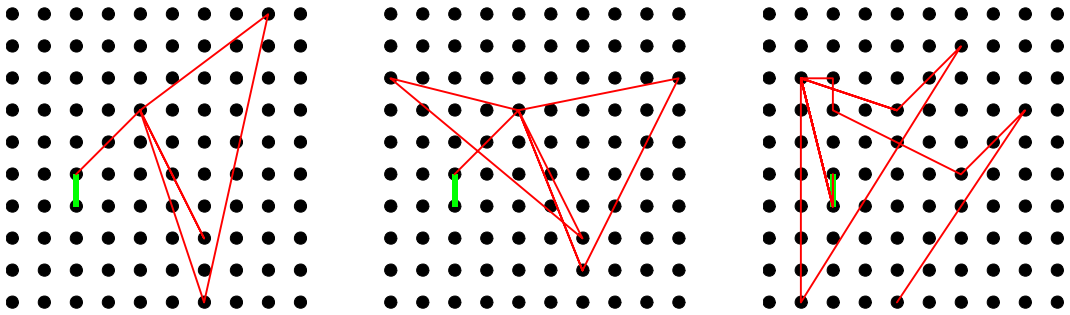


Rys. 5-89. Obrazy głoski „*sz*” – rozszczep podniebienia pierwotnego i wtórnego całkowity obustronny.

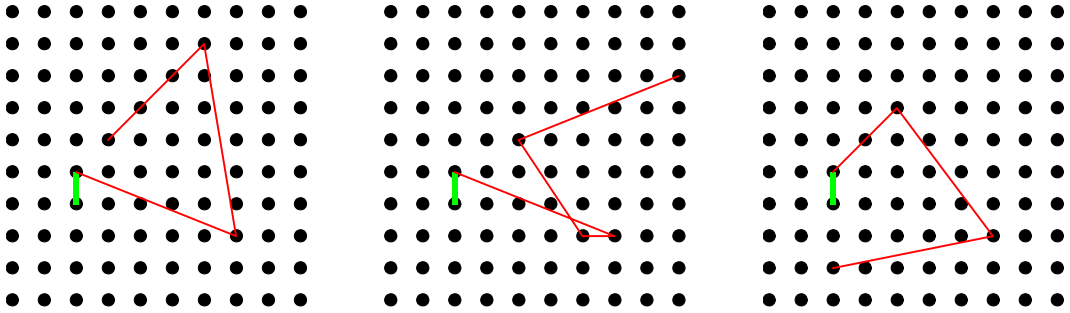


Rys. 5-90. Obrazy głoski „*sz*” – rozszczep podniebienia pierwotnego i wtórnego całkowity lewostronny lub prawostronny.

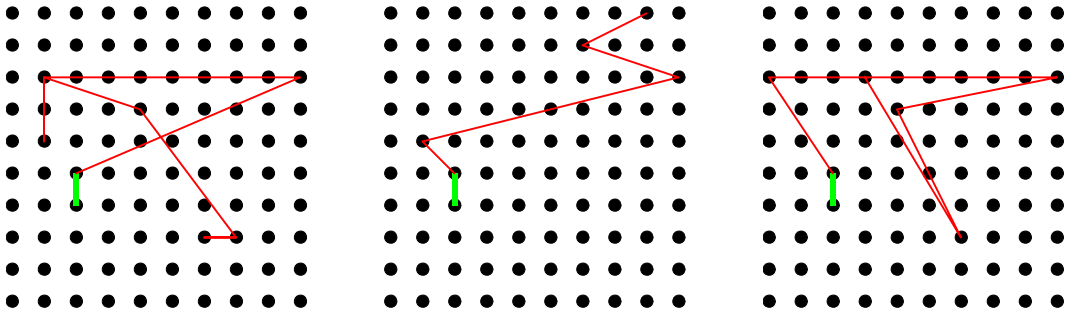
Kolejne rysunki przedstawiają obrazy głosek „*s*” i „*sz*” wygenerowane przez sieć składającą się ze 100 neuronów. Sieć trenowano odpowiednio przez 100 i 300 kroków.



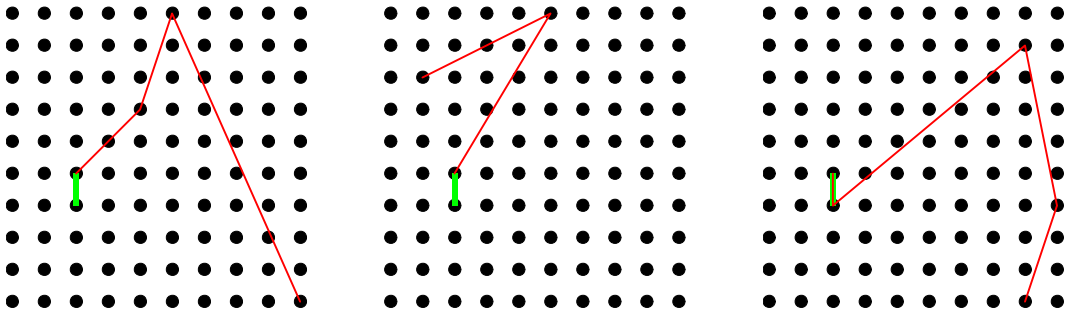
Rys. 5-91. Obrazy głoski „*s*” – rozszczep podniebienia wtórnego częściowy.



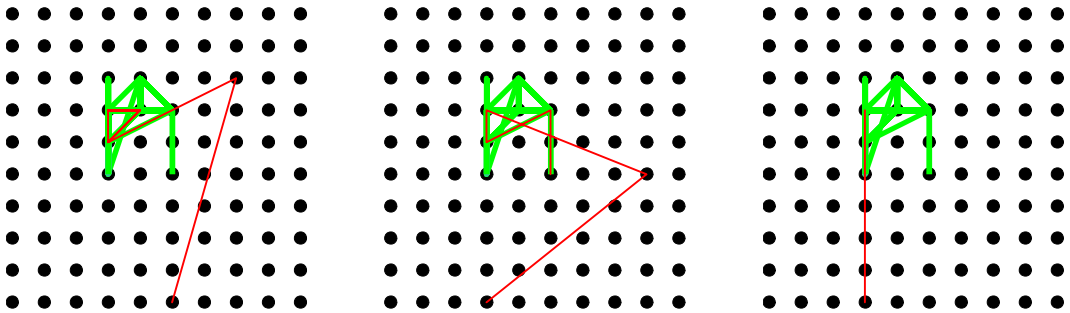
Rys. 5-92. Obrazy głoski „*s*” – rozszczep podniebienia wtórnego całkowity.



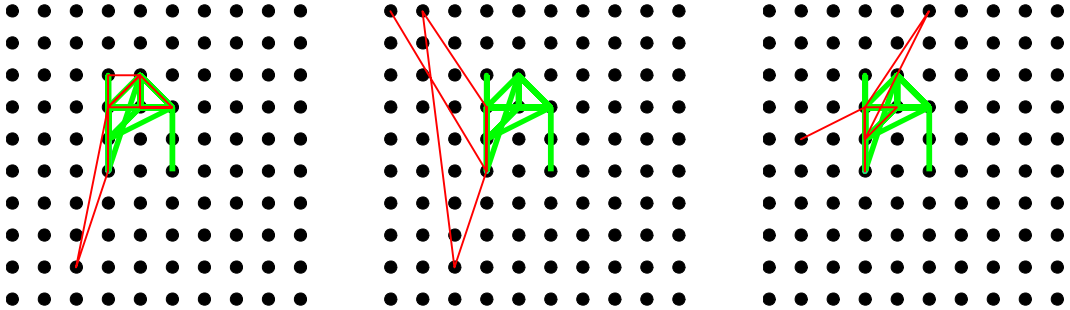
Rys. 5-93. Obrazy głoski „*s*” – rozszczep podniebienia pierwotnego i wtórnego całkowity obustronny.



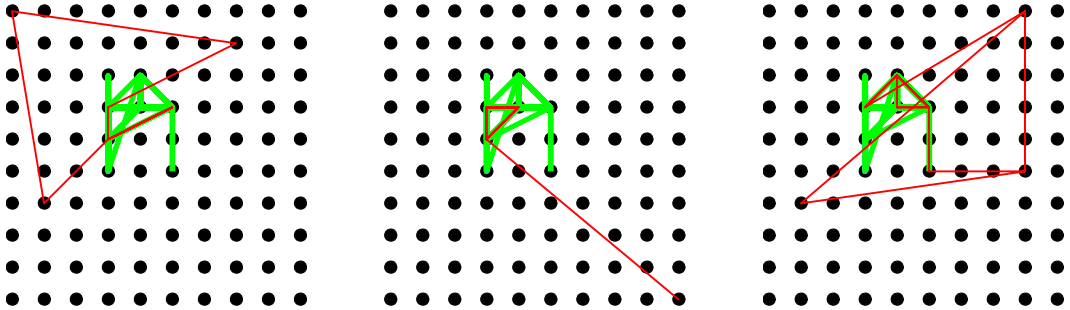
Rys. 5-94. Obrazy głoski „*s*” – rozszczep podniebienia pierwotnego i wtórnego całkowity lewostronny lub prawostronny.



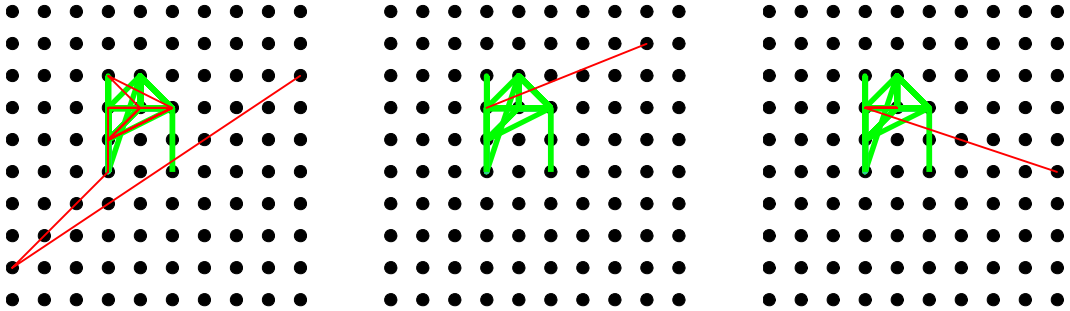
Rys. 5-95. Obrazy głoski „*sz*" – rozszczep podniebienia wtórnego częściowy.



Rys. 5-96. Obrazy głoski „*sz*" – rozszczep podniebienia wtórnego całkowity.

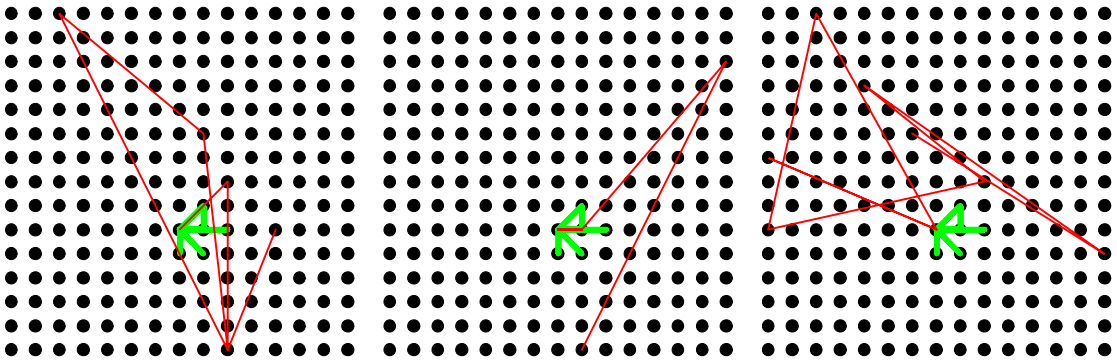


Rys. 5-97. Obrazy głoski „*sz*" – rozszczep podniebienia pierwotnego i wtórnego całkowity obustronny.

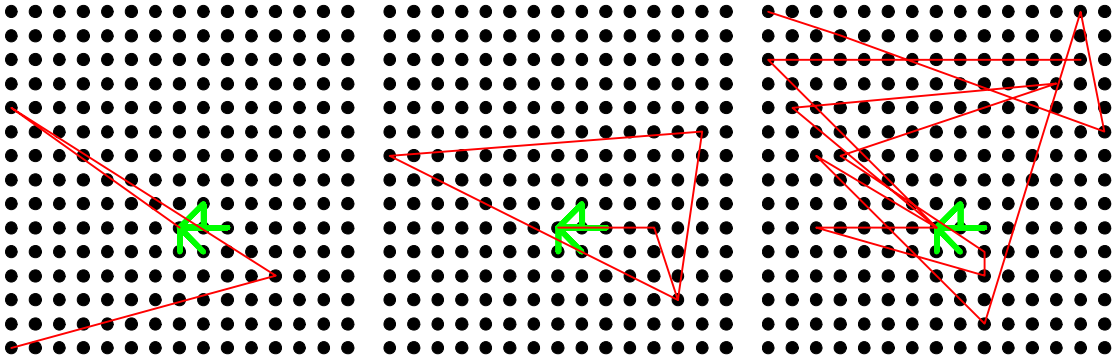


Rys. 5-98. Obrazy głoski „*sz*" – rozszczep podniebienia pierwotnego i wtórnego całkowity lewostronny lub prawostronny.

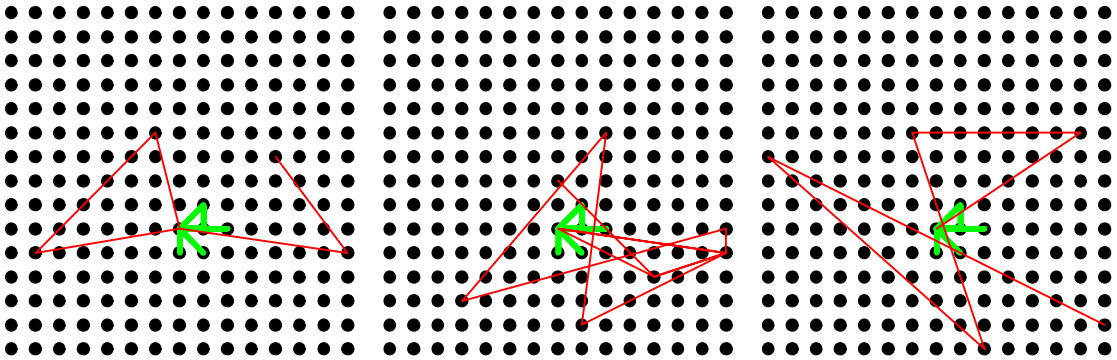
Na rys. 5-99 do 5-106 przedstawiono obrazy uzyskane przy pomocy sieci składającej się z 225 neuronów. Obrazy głoski „*s*" uzyskano przy pomocy sieci trenowanej przez 100 kroków, a obrazy głoski „*sz*" przy pomocy sieci trenowanej przez 200 kroków.



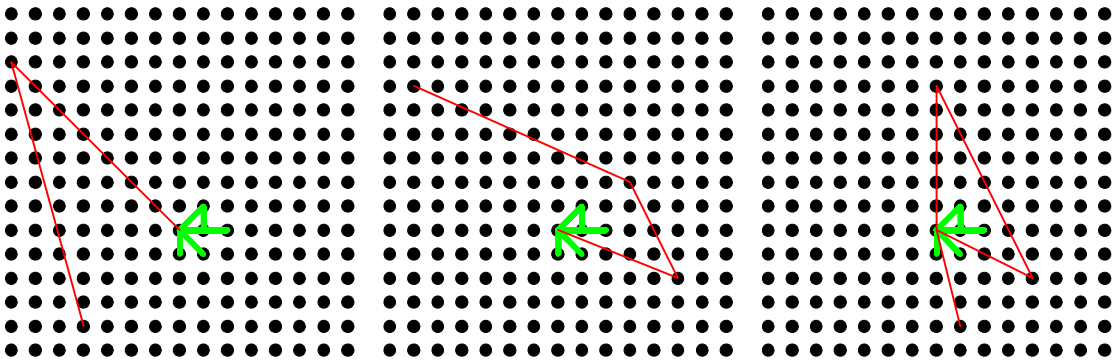
Rys. 5-99. Obrazy głoski „*s*" – rozszczep podniebienia wtórnego częściowy.



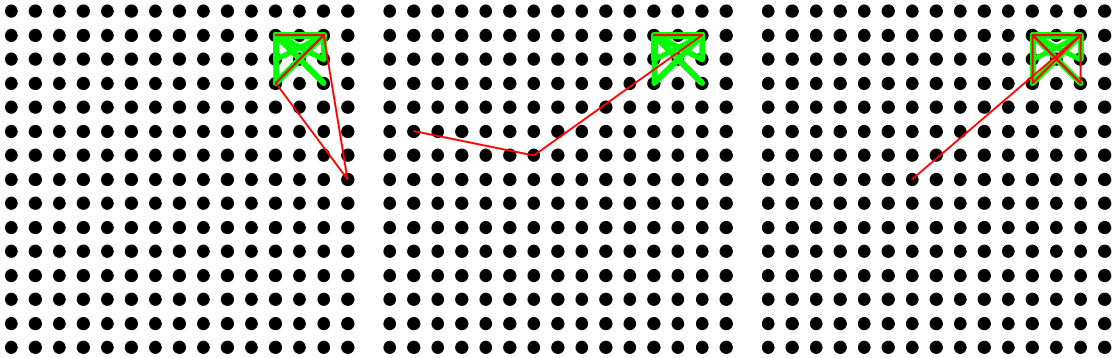
Rys. 5-100. Obrazy głoski „*s*" – rozszczep podniebienia wtórnego całkowity.



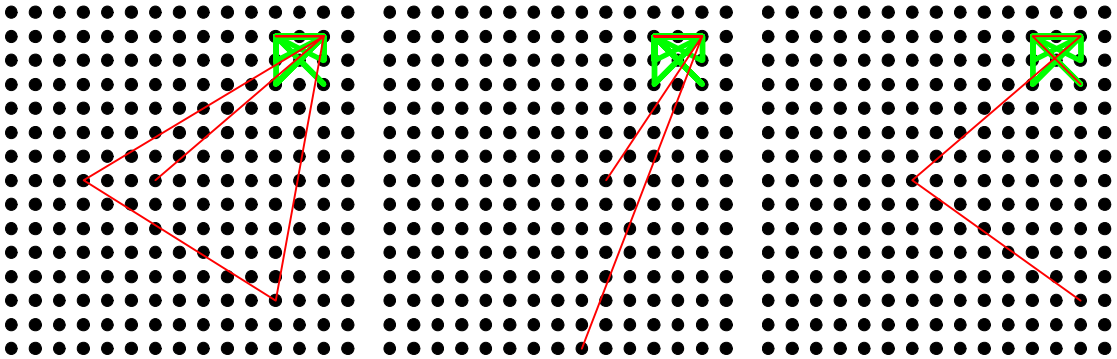
Rys. 5-101. Obrazy głoski „*s*" – rozszczep podniebienia pierwotnego i wtórnego całkowity obustronny.



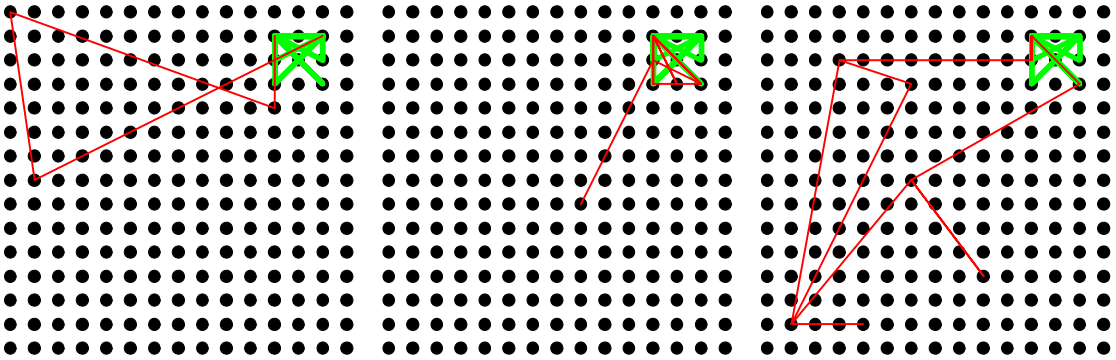
Rys. 5-102. Obrazy głoski „*s*” – rozszczep podniebienia pierwotnego i wtórnego całkowity lewostronny lub prawostronny.



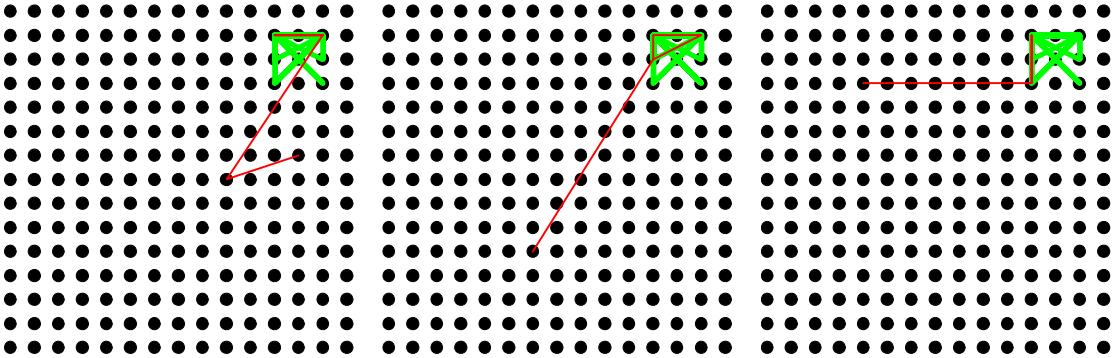
Rys. 5-103. Obrazy głoski „*sz*” – rozszczep podniebienia wtórnego częściowy.



Rys. 5-104. Obrazy głoski „*sz*” – rozszczep podniebienia wtórnego całkowity.

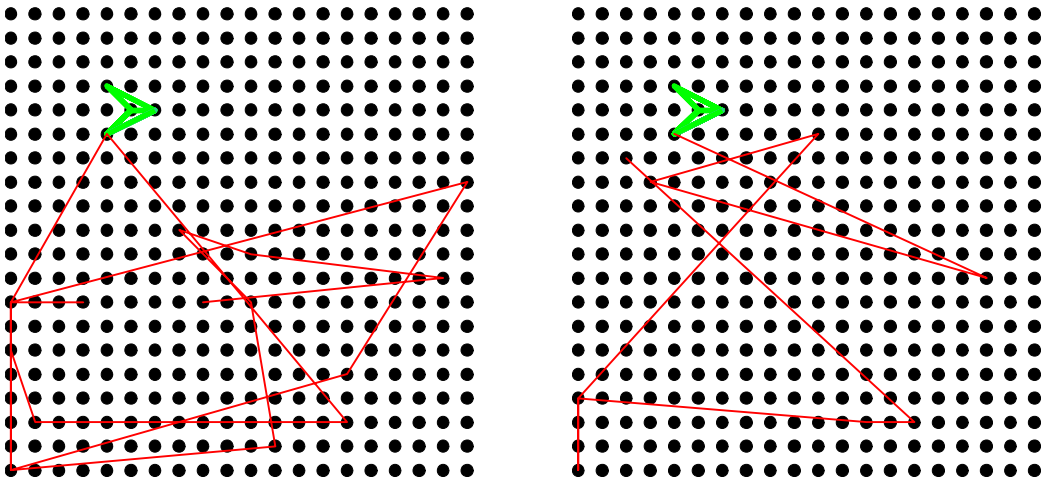


Rys. 5-105. Obrazy głoski „*sz*” – rozszczep podniebienia pierwotnego i wtórnego całkowity obustronny.

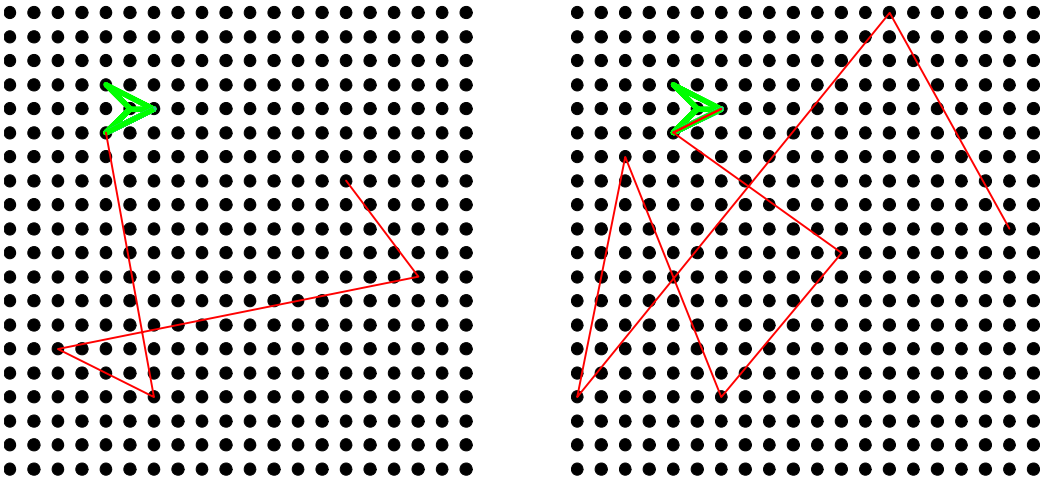


Rys. 5-106. Obrazy głoski „*sz*” – rozszczep podniebienia pierwotnego i wtórnego całkowity lewostronny lub prawostronny.

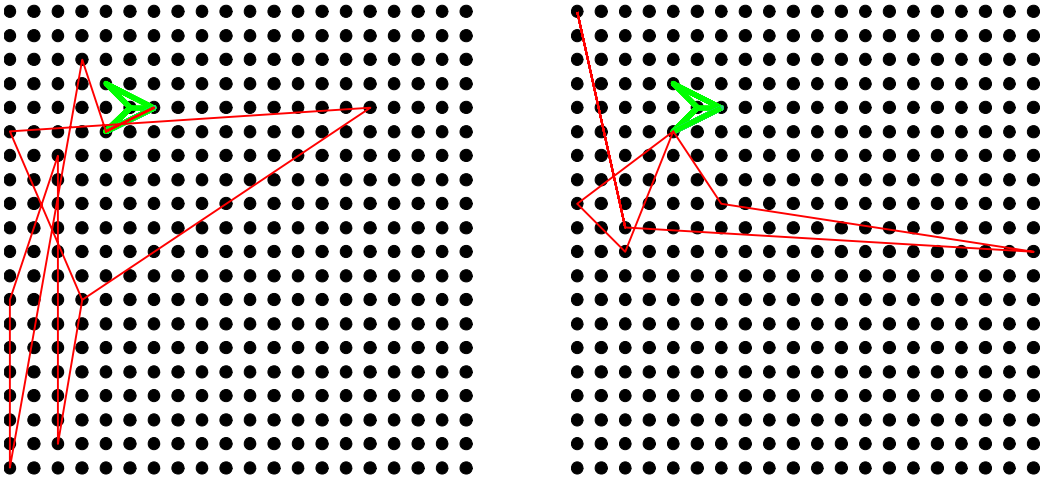
Ostatnią prezentowaną grupę stanowią obrazy głosek uzyskane przy pomocy sieci o rozmiarze 20 na 20 neuronów. W przypadku obydwu głosek przeprowadzono 100 kroków treningu.



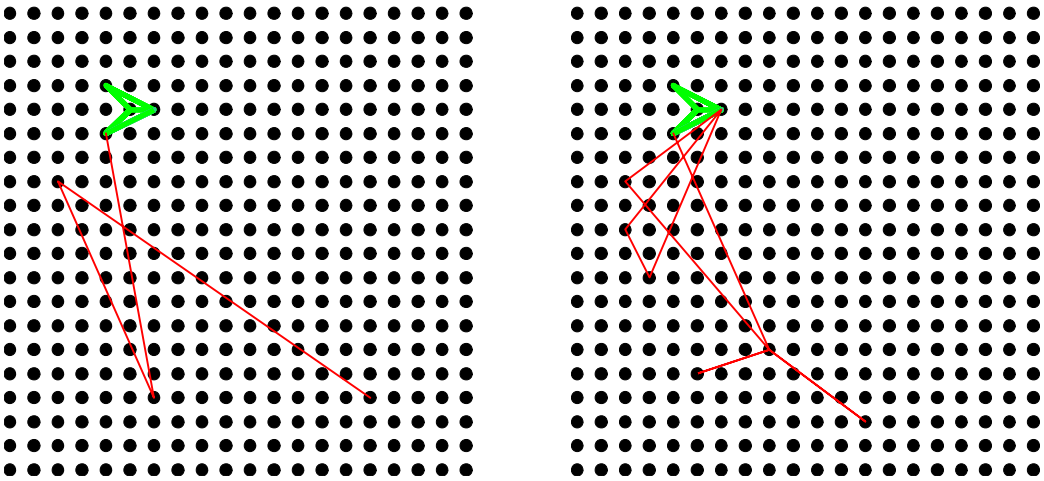
Rys. 5-107. Obrazy głoski „*s*” – rozszczep podniebienia wtórnego częściowy.



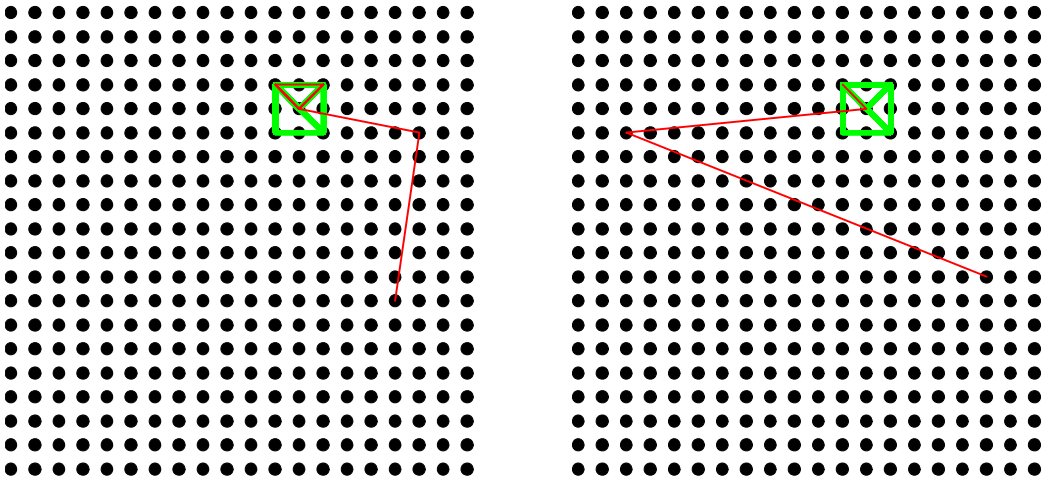
Rys. 5-108. Obrazy głoski „*s*” – rozszczep podniebienia wtórnego całkowity.



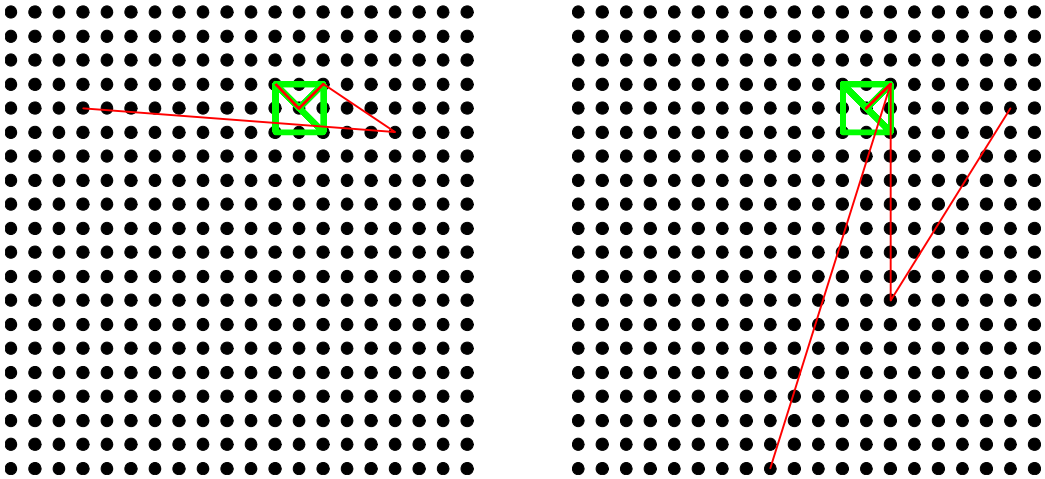
Rys. 5-109. Obrazy głoski „*s*” – rozszczep podniebienia pierwotnego i wtórnego całkowity obustronny.



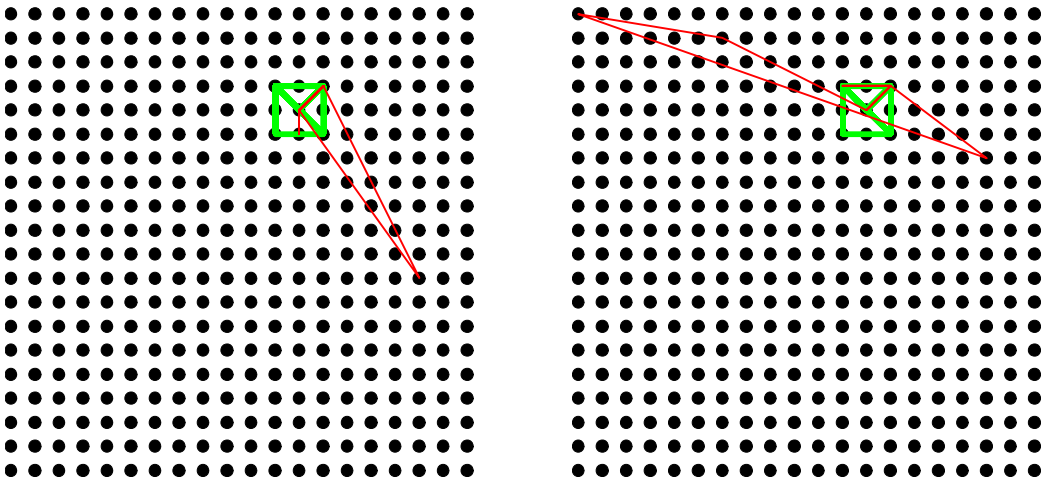
Rys. 5-110. Obrazy głoski „*s*” – rozszczep podniebienia pierwotnego i wtórnego całkowity lewostronny lub prawostronny.



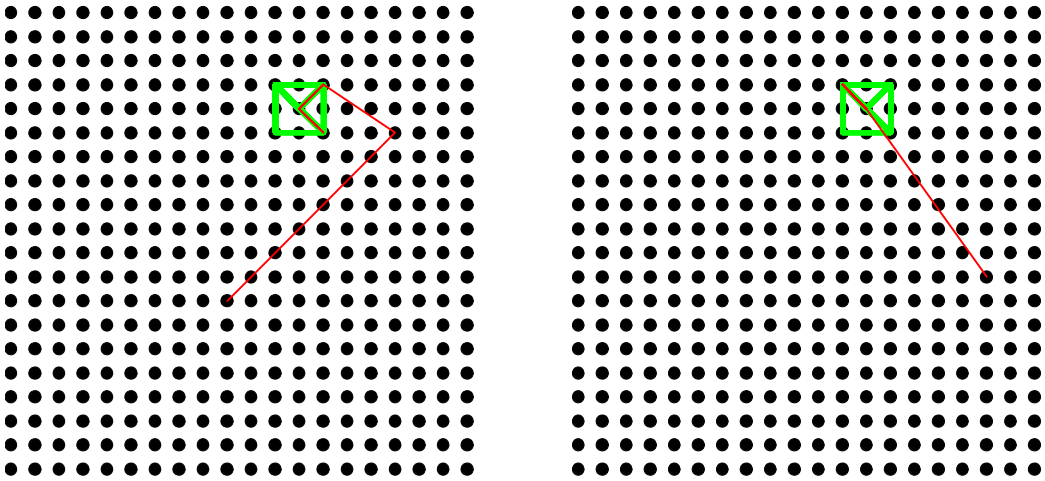
Rys. 5-111. Obrazy głoski „*sz*" – rozszczep podniebienia wtórnego częściowy.



Rys. 5-112. Obrazy głoski „*sz*" – rozszczep podniebienia wtórnego całkowity.



Rys. 5-113. Obrazy głoski „*sz*" – rozszczep podniebienia pierwotnego i wtórnego całkowity obustronny.



Rys. 5-114. Obrazy głoski „*sz*" – rozszczep podniebienia pierwotnego i wtórnego całkowity lewostronny lub prawostronny.

5.2.4. Dyskusja uzyskanych wyników dla izolowanych głosek w sieci z wprowadzonym sąsiedztwem

Nawet powierzchowne obejrzenie obrazów na podanych wyżej rysunkach pokazuje, że zaproponowana technika wizualizacji jest skuteczna. Podobnie jak w przypadku sieci trenowanej bez uwzględnienia czynnika sąsiedztwa, również te wyżej przytoczone obrazy pozwalają na łatwe odseparowanie wypowiedzi wzorcowych od patologicznych. Można zauważyć, podobnie jak w przypadku obrazowania całych wyrazów, że długość krzywej na obrazach wypowiedzi wzorcowej jest znacznie krótsza niż na obrazach wypowiedzi patologicznej. Dla sieci o odpowiednio dużym rozmiarze różnica ta jest tak znacząca, iż pozawala na 100% dyskryminację wypowiedzi patologicznych [Kap99a, Kap99b]. Potwierdzają to pomiary średniej długości krzywej reprezentującej daną wypowiedź. W tab. 5-14 do 5-23 zamieszczono wartości średniej długości krzywej w zależności od rozmiaru sieci i liczby kroków treningu.

Warto odnotować fakt, że w tym przypadku nie trenowano sieci pod kątem rozpoznawania pojedynczej patologii. Dla raz wytrenowanej sieci wyliczano średnią długość krzywej jednocześnie dla wypowiedzi wzorcowych oraz wszystkich 4 kategorii wypowiedzi patologicznych. Aby podkreślić, iż wyniki te nie są przypadkowe, lecz jest to ogólna własność prezentowanych obrazów, powtórzono test 10 razy – prezentowane w tabelach dane są wartością średnią z tych 10 testów.

	Kroki treningu							
	100	200	300	400	500	1000	1500	2000
25 neuronów	5,5	7,4	9,0	10,1	8,8	11,9	13,2	13,1
100 neuronów	3,9	6,4	7,8	8,7	11,1	14,3	14,8	17,3
225 neuronów	6,7	8,7	8,3	11,7	10,3	15,3	17,1	21,4
400 neuronów	4,1	6,9	8,7	11,4	11,6	14,6	17,6	19,6

Tab. 5-14. Średnia długość krzywej w obrazach głoski „*s*", dane dla wypowiedzi wzorcowej.

	Kroki treningu							
	100	200	300	400	500	1000	1500	2000
25 neuronów	4,7	5,9	6,2	7,8	10,3	16,8	19,8	21,6
100 neuronów	22,3	21,3	24,4	23,3	23,8	29,9	29,5	31,8
225 neuronów	32,1	34,5	30,7	34,2	33,4	37,0	36,1	39,6
400 neuronów	41,1	45,7	42,6	40,3	39,6	41,7	43,9	42,3

Tab. 5-15. Średnia długość krzywej w obrazach głoski „*s*", dane dla rozszczepu podniebienia wtórnego częściowego.

	Kroki treningu							
	100	200	300	400	500	1000	1500	2000
25 neuronów	12,2	14,6	13,0	15,6	15,9	18,8	18,8	20,4
100 neuronów	20,2	23,2	24,4	24,2	25,6	28,2	30,3	32,8
225 neuronów	28,3	31,3	31,5	33,1	31,6	39,7	39,2	39,9
400 neuronów	42,5	39,3	40,9	37,5	44,1	42,7	46,5	45,3

Tab. 5-16. Średnia długość krzywej w obrazach głoski „*s*", dane dla rozszczepu podniebienia wtórnego całkowitego.

	Kroki treningu							
	100	200	300	400	500	1000	1500	2000
25 neuronów	12,3	14,9	15,5	18,1	18,1	20,0	21,7	24,1
100 neuronów	24,5	25,3	26,7	26,0	28,3	32,5	32,5	40,5
225 neuronów	36,2	39,6	34,7	37,7	35,2	43,6	41,9	46,5
400 neuronów	51,9	46,6	51,6	46,2	48,5	49,1	51,4	51,6

Tab. 5-17. Średnia długość krzywej w obrazach głoski „*s*", dane dla rozszczepu podniebienia pierwotnego i wtórnego całkowitego obustronnego.

	Kroki treningu							
	100	200	300	400	500	1000	1500	2000
25 neuronów	16,3	18,3	16,1	19,3	20,2	19,7	21,9	23,7
100 neuronów	27,9	29,4	31,1	28,4	30,8	31,9	35,0	34,9
225 neuronów	39,5	43,0	40,9	39,9	40,6	43,6	43,1	47,5
400 neuronów	54,6	56,0	50,6	51,1	50,0	53,3	53,7	53,8

Tab. 5-18. Średnia długość krzywej w obrazach głoski „*ś*", dane dla rozszczepu podniebienia pierwotnego i wtórnego całkowitego lewostronnego lub prawostronnego.

	Kroki treningu							
	100	200	300	400	500	1000	1500	2000
25 neuronów	4,7	5,9	6,2	7,8	10,5	10,6	9,7	12,5
100 neuronów	8,2	6,3	6,9	9,0	9,6	12,0	14,5	16,2
225 neuronów	7,8	7,1	8,1	9,8	9,6	15,0	13,6	15,0
400 neuronów	5,6	8,4	8,0	10,0	10,9	15,2	16,8	18,6

Tab. 5-19. Średnia długość krzywej w obrazach głoski „*sz*", dane dla wypowiedzi wzorcowej.

	Kroki treningu							
	100	200	300	400	500	1000	1500	2000
25 neuronów	6,4	8,5	10,4	10,4	13,3	15,7	17,2	21,4
100 neuronów	11,2	11,6	13,0	12,7	17,0	18,0	23,8	26,0
225 neuronów	15,4	14,7	14,8	16,1	18,1	24,6	26,1	31,8
400 neuronów	15,7	15,5	17,6	18,6	18,7	26,6	31,4	36,6

Tab. 5-20. Średnia długość krzywej w obrazach głoski „*sz*", dane dla rozszczepu podniebienia wtórnego częściowego.

	Kroki treningu							
	100	200	300	400	500	1000	1500	2000
25 neuronów	8,4	10,9	13,7	14,3	15,6	16,6	20,5	23,2
100 neuronów	16,2	16,0	20,0	19,9	21,1	23,2	28,1	33,5
225 neuronów	23,3	20,7	22,4	23,8	24,4	31,2	32,6	38,0
400 neuronów	28,1	25,0	24,4	26,2	27,7	34,0	39,9	43,6

Tab. 5-21. Średnia długość krzywej w obrazach głoski „*sz*", dane dla rozszczepu podniebienia wtórnego całkowitego.

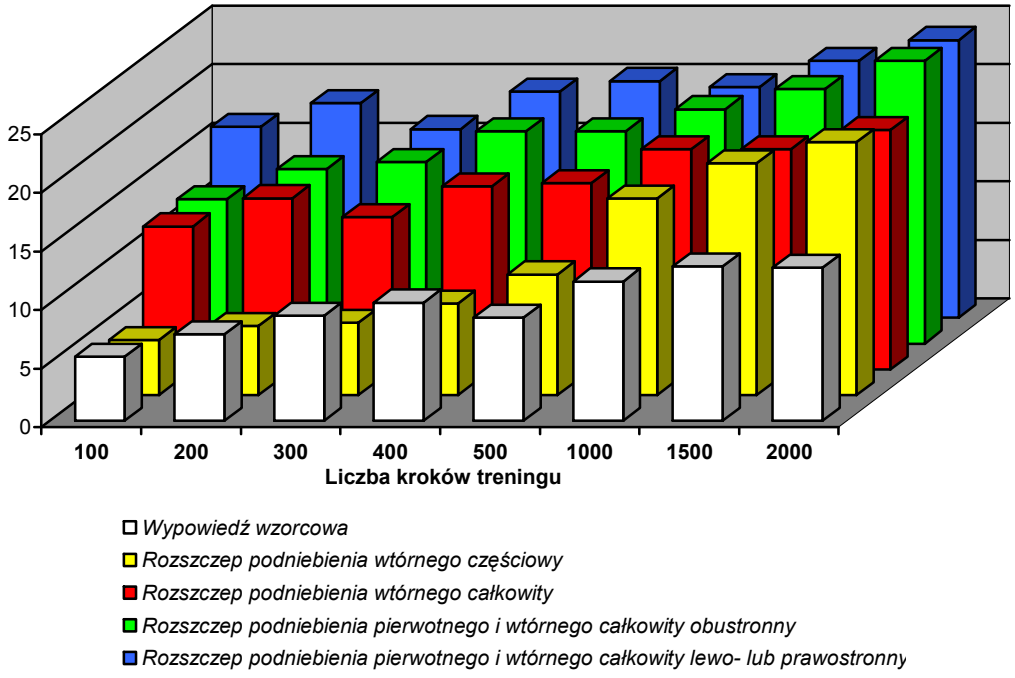
	Kroki treningu							
	100	200	300	400	500	1000	1500	2000
25 neuronów	8,4	11,1	12,0	13,6	14,5	17,4	18,6	25,3
100 neuronów	15,9	15,5	18,3	16,7	20,5	23,6	29,3	32,8
225 neuronów	23,0	21,9	22,7	24,2	24,4	30,8	34,5	37,9
400 neuronów	27,0	25,2	25,0	27,6	26,6	35,0	43,4	47,9

Tab. 5-22. Średnia długość krzywej w obrazach głoski „*sz*”, dane dla rozszczepu podniebieniapierwotnego i wtórnego całkowitego obustronnego.

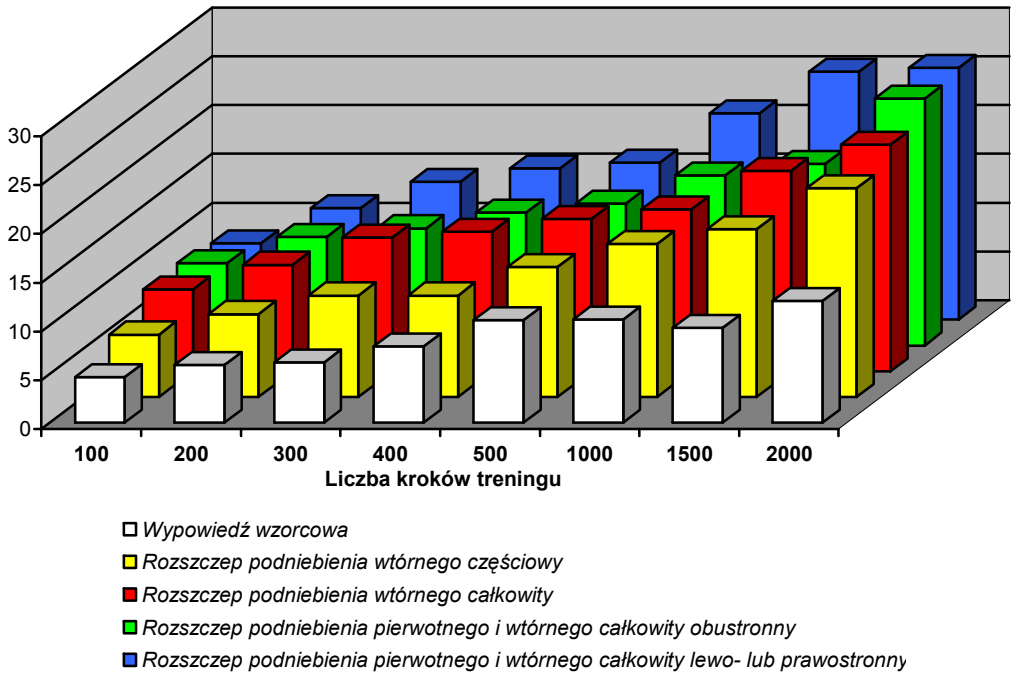
	Kroki treningu							
	100	200	300	400	500	1000	1500	2000
25 neuronów	7,8	11,4	14,1	15,5	16,1	21,1	25,4	25,8
100 neuronów	16,0	15,8	19,4	18,3	20,0	23,0	29,0	36,2
225 neuronów	21,3	18,7	23,1	23,6	24,9	29,7	34,7	39,5
400 neuronów	23,6	21,2	22,8	25,5	29,0	34,6	40,4	43,5

Tab. 5-23. Średnia długość krzywej w obrazach głoski „*sz*”, dane dla rozszczepu podniebienia pierwotnego i wtórnego całkowitego lewostronnego lub prawostronnego.

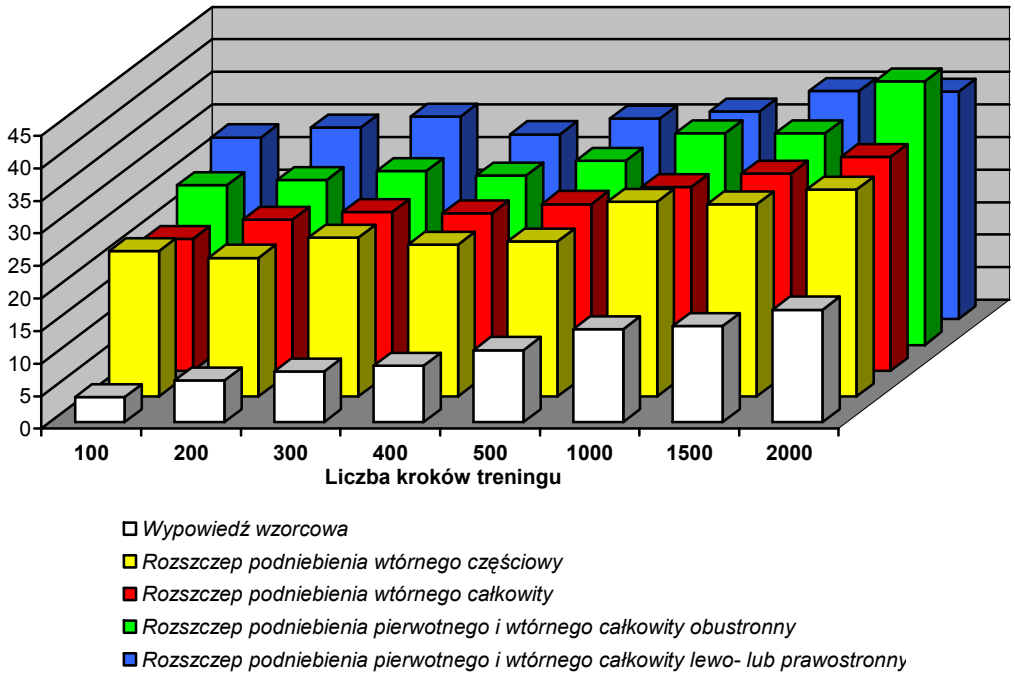
Dla łatwiejszej oceny uzyskanych wyników dane z tabel 5-14 do 5-23 zaprezentowano również w formie graficznej na wykresach 5-6 do 5-13.



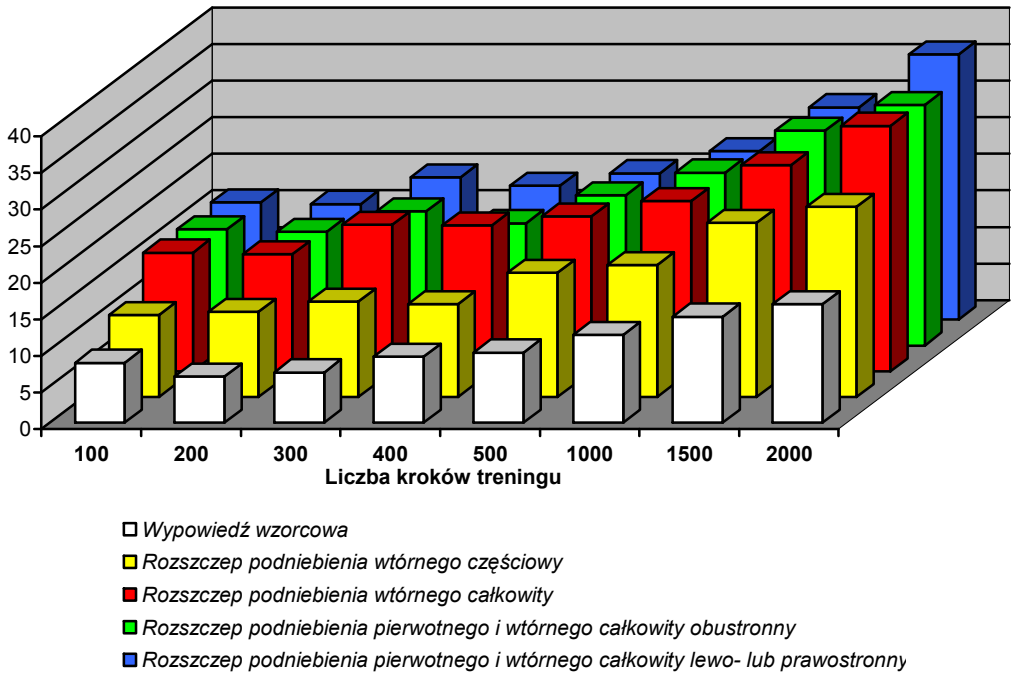
Wyk. 5-6. Średnia długość krzywej w obrazach głoski „*s*” w zależności od liczby kroków treningu (dane dla sieci o rozmiarze 25 neuronów).



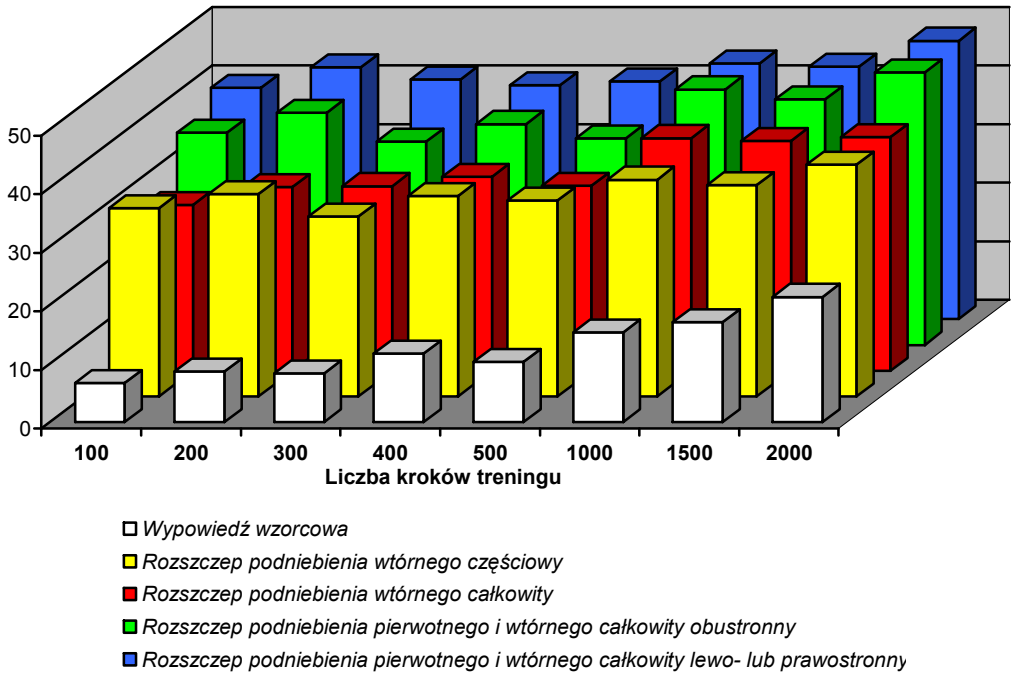
Wyk. 5-7. Średnia długość krzywej w obrazach głoski „*sz*" w zależności od liczby kroków treningu (dane dla sieci o rozmiarze 25 neuronów).



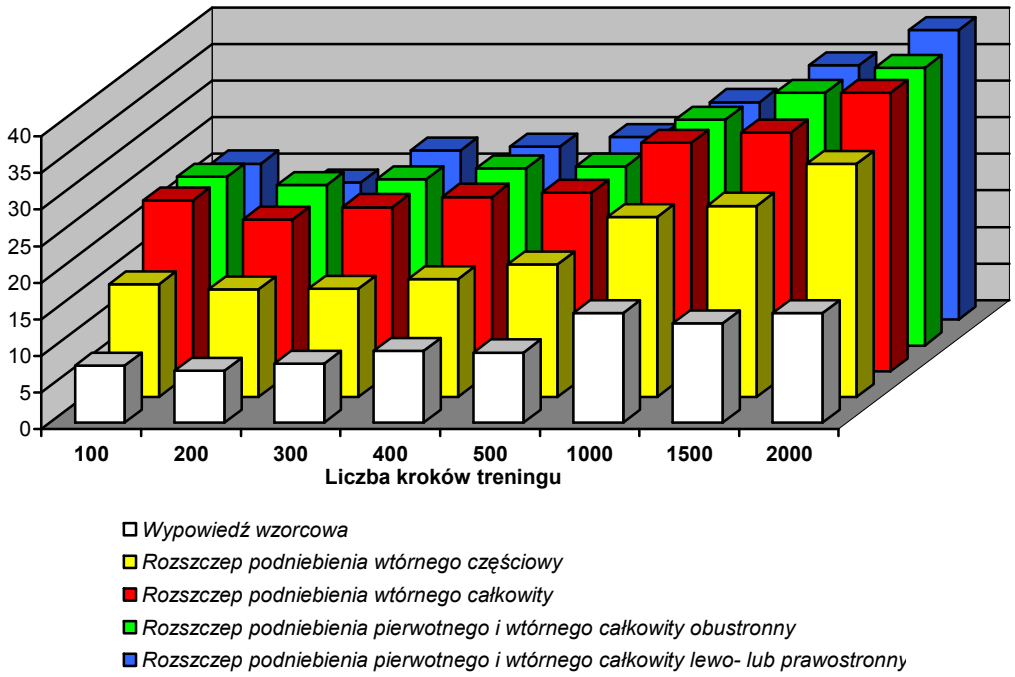
Wyk. 5-8. Średnia długość krzywej w obrazach głoski „*s*" w zależności od liczby kroków treningu (dane dla sieci o rozmiarze 100 neuronów).



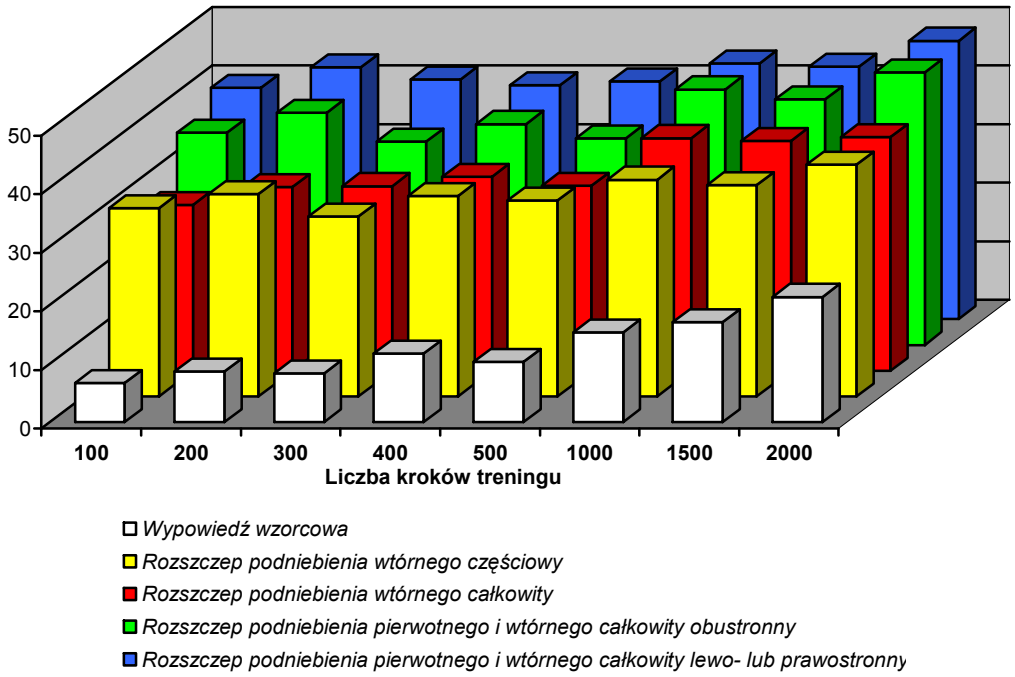
Wyk. 5-9. Średnia długość krzywej w obrazach głoski „*sz*” w zależności od liczby kroków treningu (dane dla sieci o rozmiarze 100 neuronów).



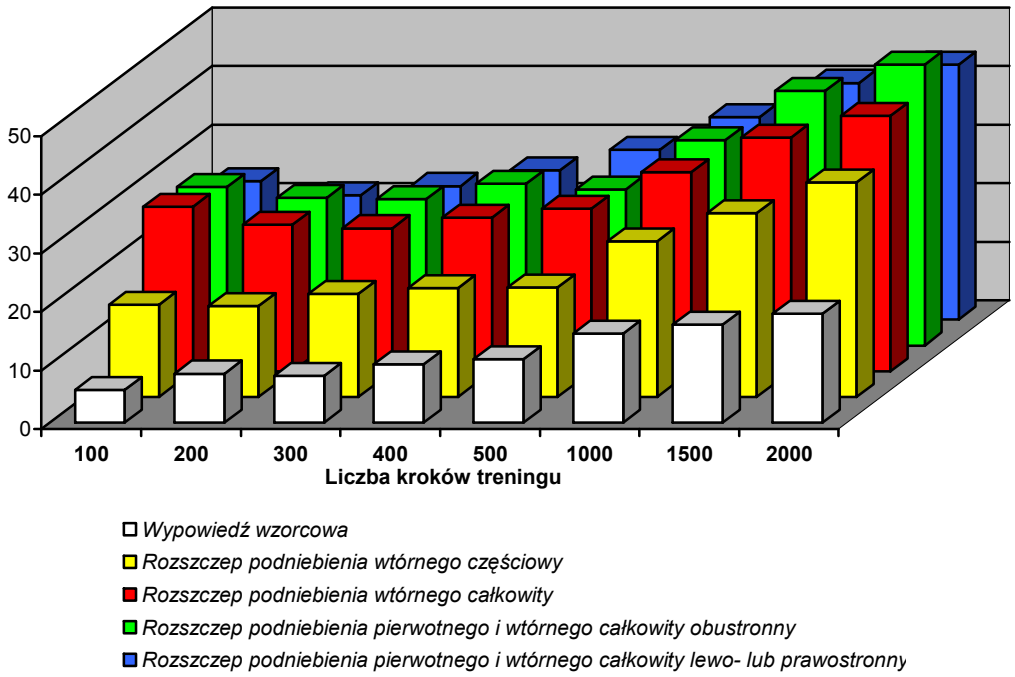
Wyk. 5-10. Średnia długość krzywej w obrazach głoski „*s*” w zależności od liczby kroków treningu (dane dla sieci o rozmiarze 225 neuronów).



Wyk. 5-11. Średnia długość krzywej w obrazach głoski „*sz*" w zależności od liczby kroków treningu (dane dla sieci o rozmiarze 225 neuronów).



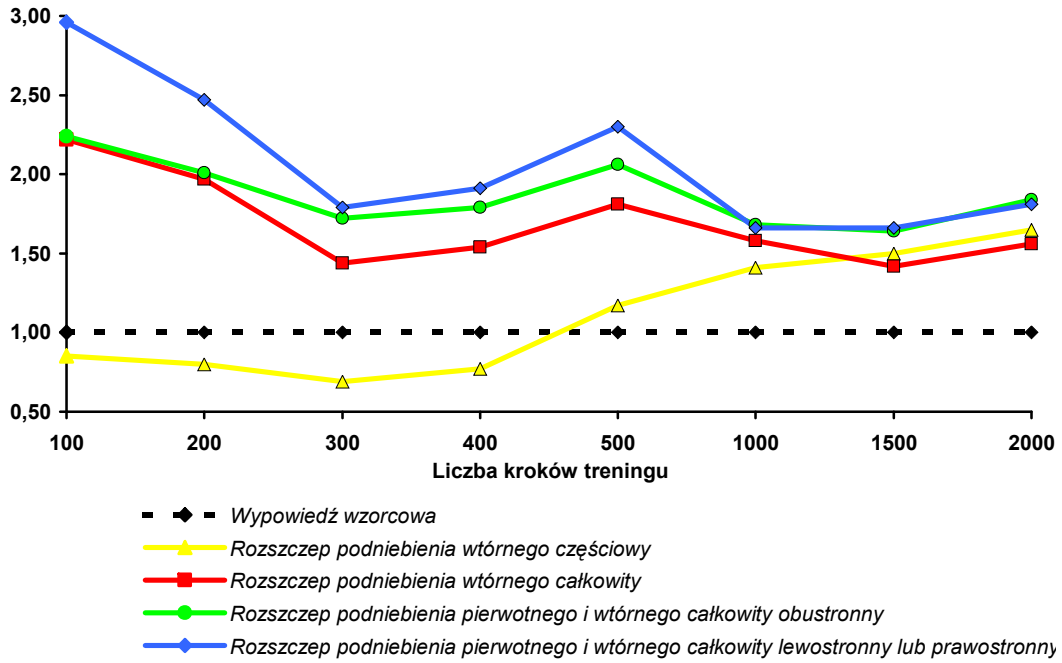
Wyk. 5-12. Średnia długość krzywej w obrazach głoski „*s*" w zależności od liczby kroków treningu (dane dla sieci o rozmiarze 400 neuronów).



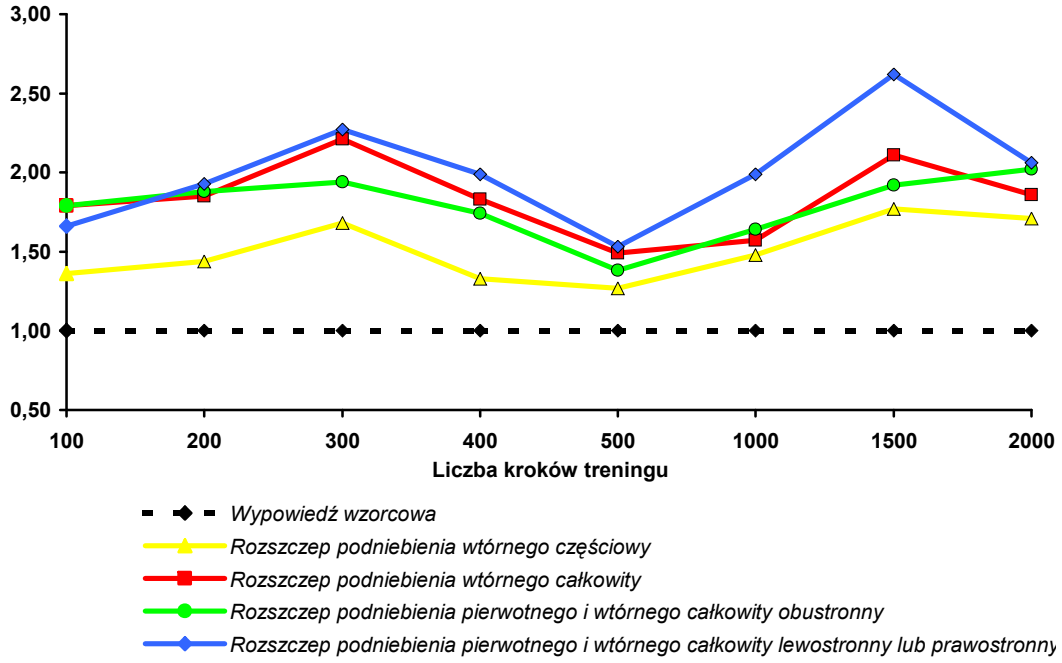
Wyk. 5-13. Średnia długość krzywej w obrazach głoski „*sz*" w zależności od liczby kroków treningu (dane dla sieci o rozmiarze 400 neuronów).

5.2.5. Dobór optymalnych parametrów stosowanej metody

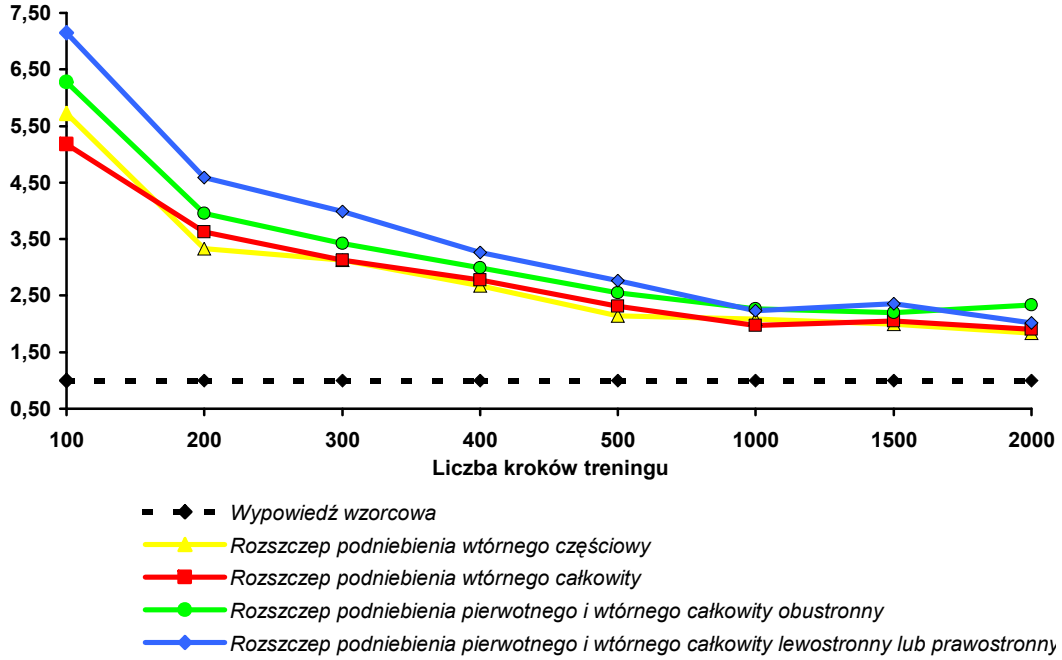
Na podstawie danych zamieszczonych w tabelach 5-14 do 5-23 podjęto próbę ustalenia, jakie są optymalne parametry (rozmiar sieci, długość procesu uczenia) rozważanej metody. W tym celu skonstruowano wykresy, na których przedstawiono stosunek średniej długości krzywej zmierzonej dla danego schorzenia do średniej długości krzywej uzyskanej dla wypowiedzi wzorcowej. Wykresy 5-14 do 5-21 przedstawiają obliczony stosunek długości oddzielnie dla każdej głoski i rozmiaru sieci.



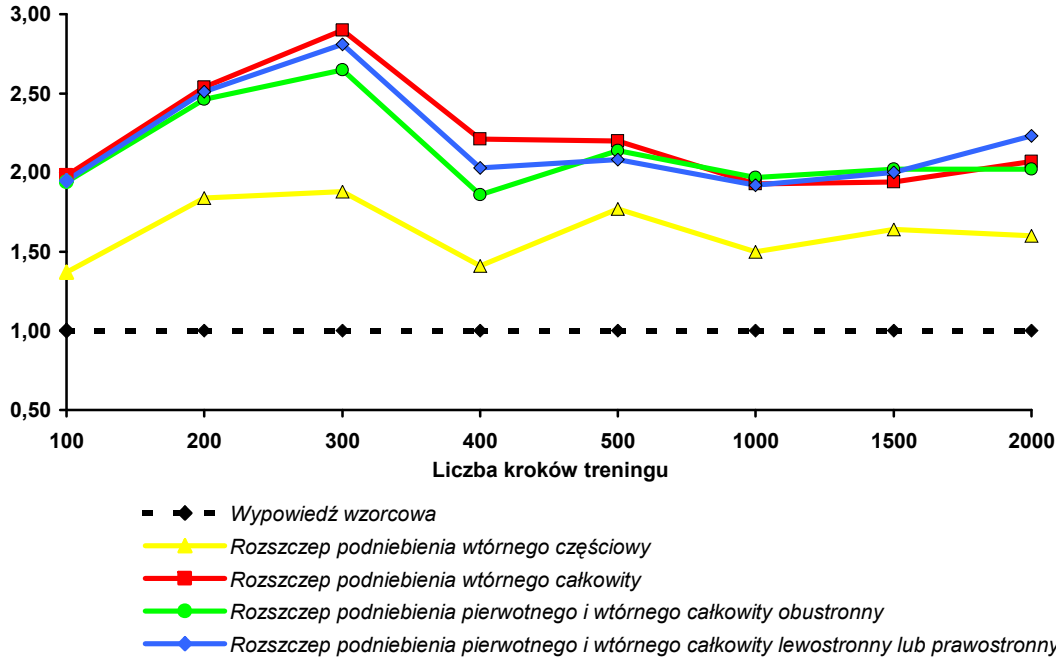
Wyk. 5-14. Względna długość krzywej w obrazach głoski „*s*” w zależności od liczby kroków treningu (dane dla sieci o rozmiarze 25 neuronów).



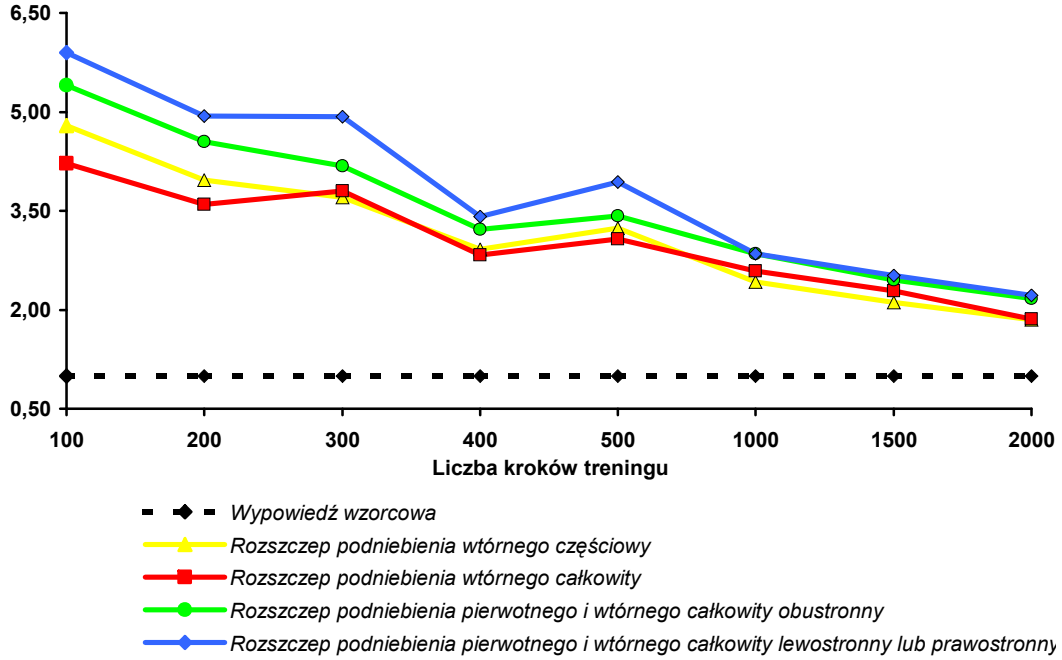
Wyk. 5-15. Względna długość krzywej w obrazach głoski „*sz*” w zależności od liczby kroków treningu (dane dla sieci o rozmiarze 25 neuronów).



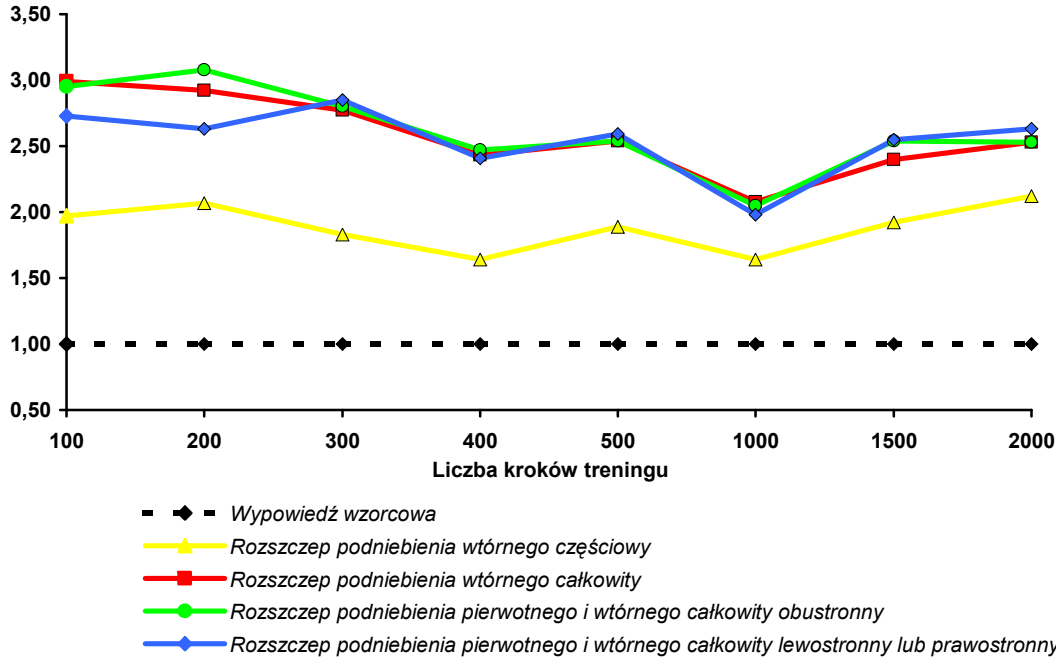
Wyk. 5-16. Względna długość krzywej w obrazach głoski „*s*” w zależności od liczby kroków treningu (dane dla sieci o rozmiarze 100 neuronów).



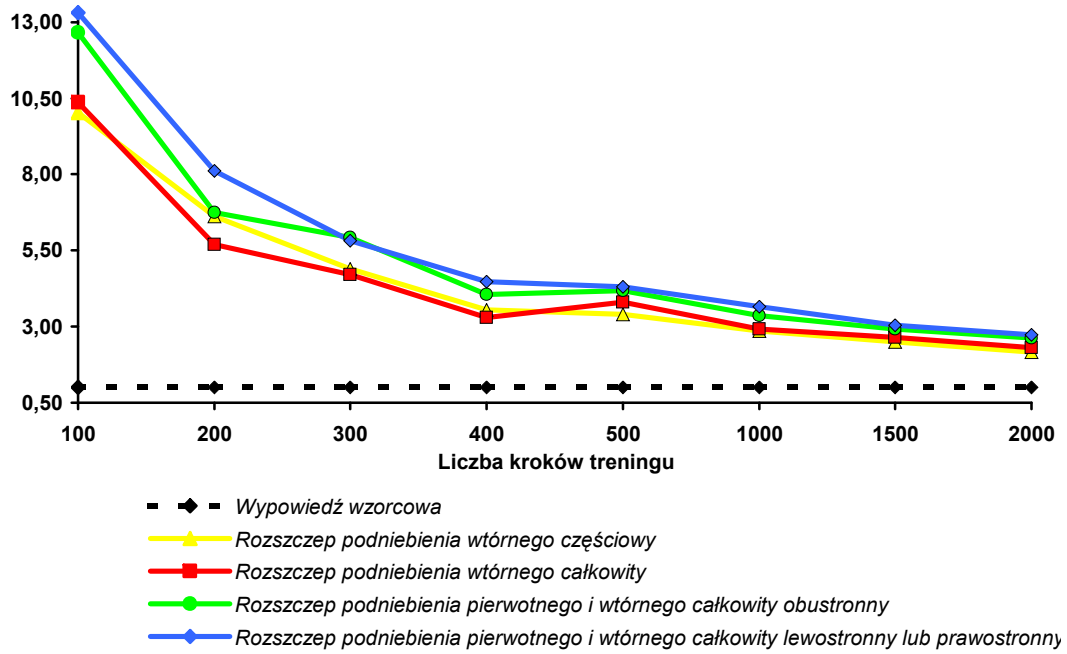
Wyk. 5-17. Względna długość krzywej w obrazach głoski „*sz*” w zależności od liczby kroków treningu (dane dla sieci o rozmiarze 100 neuronów).



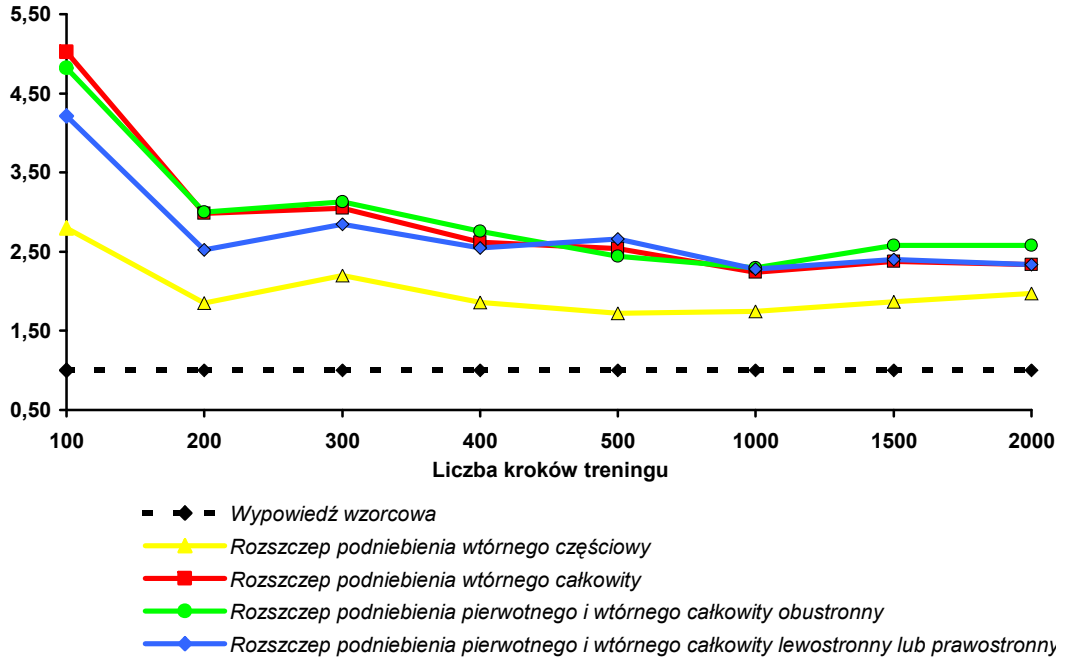
Wyk. 5-18. Względna długość krzywej w obrazach głoski „*s*” w zależności od liczby kroków treningu (dane dla sieci o rozmiarze 225 neuronów).



Wyk. 5-19. Względna długość krzywej w obrazach głoski „*sz*” w zależności od liczby kroków treningu (dane dla sieci o rozmiarze 225 neuronów).



Wyk. 5-20. Względna długość krzywej w obrazach głoski „*s*” w zależności od liczby kroków treningu (dane dla sieci o rozmiarze 400 neuronów).



Wyk. 5-21. Względna długość krzywej w obrazach głoski „*sz*” w zależności od liczby kroków treningu (dane dla sieci o rozmiarze 400 neuronów).

Na podstawie powyższych wykresów można wyciągnąć wnioski dotyczące optymalnego rozmiaru sieci oraz minimalnej liczby kroków treningu, przy której sieć najlepiej dyskryminuje wypowiedzi patologiczne. W przypadku wyboru sieci o rozmiarze 5 na 5 neuronów można generalnie stwierdzić, że uzyskane wyniki nie są zadowalające. Mianowicie można zaobserwować, że dla głoski „*s*", w przypadku rozszczepu podniebienia wtórnego częściowego, dla liczby kroków treningu mniejszej od 500, średnia długość krzywych jest **mniejsza** od średniej długości krzywych dla wypowiedzi patologicznych. Dyskredytuje to całkowicie omawiany przypadek i skłania do jego eliminacji z rozważań. W przypadku ustalenia rozmiaru sieci na 10 na 10 neuronów, otrzymane rezultaty są już w zasadzie zadowalające, aczkolwiek różnica stosunków długości krzywych dla wypowiedzi wzorcowej i wypowiedzi z częściowym rozszczepem podniebienia głoski „*sz*" nie jest zbyt duża, co może powodować, w niektórych przypadkach, uznanie wypowiedzi patologicznej za wzorcową.

Wzrost rozmiaru sieci powoduje powiększenie tego stosunku, a tym samym zwiększenie skuteczności dyskryminacji. Największą różnicę można zaobserwować w przypadku sieci zbudowanej z 400 neuronów. Dlatego wydaje się zasadnym stwierdzenie, iż **im większa jest sieć** służąca do obrazowania pojedynczych głosek **tym otrzymane rezultaty są lepsze**. Na marginesie należy wspomnieć, iż wniosek ten jest zgodny z rezultatami badań przeprowadzonych dla sieci bez sąsiedztwa.

Jeżeli chodzi o oszacowanie minimalnej liczby kroków treningu, to należy stwierdzić, iż **nadmierne zwiększenie liczby iteracji nie przynosi poprawy rezultatów**. Zjawisko powyższe spowodowane jest faktem, iż w przypadku powiększenia liczby kroków treningowych sieć zaczyna rozróżniać coraz więcej szczegółów w wypowiedziach wzorcowych, co bynajmniej nie sprzyja prawidłowej dyskryminacji. Następuje wtedy wzrost długości wzorcowej trajektorii, a co za tym idzie – niekorzystne zmniejszenie stosunku długości trajektorii odpowiadającym wypowiedziom patologicznym do długości trajektorii wzorcowych. Najlepiej widoczne jest to w przypadku sieci o rozmiarze 20 na 20 neuronów. Taką wartością graniczną w przypadku głoski „s" jest 100-200 kroków treningu, natomiast w przypadku głoski „*sz*" 100-300 kroków treningu. Wyjątek stanowią rezultaty uzyskane dla głoski „*sz*" i sieci złożonej z 25 neuronów, jednak (jak stwierdzono to wcześniej) sieć taka jest za mała do poprawnego obrazowania pojedynczych głosek.

5.3. Wnioski końcowe

Z przedstawionych w tym paragrafie rezultatów badań wynika w sposób jednoznaczny, iż **obrazy izolowanych głosek charakteryzują się większą czytelnością niż obrazy pojedynczych słów**. Dodatkowym argumentem przemawiającym za takim wnioskiem jest fakt, że w przypadku obrazowania głosek możliwe jest wskazanie pewnego wspólnego zobrazowania dla wszystkich obrazów wypowiedzi wzorcowych. W zależności od wprowadzenia w sieci pojęcia sąsiedztwa jest to albo pojedynczy punkt, lub pewien zbiór punktów skupionych w niewielkim obszarze obrazu. W każdym z tych przypadków można w sposób bardzo czytelny zaprezentować obraz badanej wypowiedzi na tle tego uogólnionego obrazu wypowiedzi wzorcowych. Należy zaznaczyć, że niewątpliwą zaletą zaproponowanej metody wizualizacji jest fakt, iż **bez względu na długość badanej wypowiedzi zostanie ona przekształcona na obraz o stałym rozmiarze**, co pozwala na łatwą wizualizację i wygodną interpretację pozyskiwanych obrazów. Pozwala to również na bardzo proste porównywanie kilku obrazów, np. przez nałożenie ich na siebie (własność ta została wykorzystana przy tworzeniu sumarycznego obrazu głosek wypowiedzi wzorcowej).

Trzeba jednak także z żalem odnotować pewne niepowodzenia, napotkane przy realizacji zadań wynikających z tezy prezentowanej pracy. Trzeba obiektywnie przyznać, że rezultaty uzyskane dla obrazów pojedynczych słów nie są tak dobre, by na ich podstawie można było już teraz zaproponować metodę wizualizacji danych akustycznych odznaczającą się użytecznością praktyczną. Prowadząc obszerne i wyczerpujące badania wykazano, że tylko sieć bez sąsiedztwa jest w stanie wygenerować obrazy pozwalające na odróżnienie (na podstawie oceny długości oraz kształtu krzywej) wypowiedzi patologicznych i wzorcowych, przy czym efekty są tu wyraźnie gorsze, niż uzyskiwane dla izolowanych głosek. Wprawdzie na obrazach tych można wskazać pewien wspólny kształt trajektorii – zarówno poprawnych, jak i patologicznych, jednak jest on na tyle ogólny, że niemożliwe jest jego wyraźne wyodrębnienie i umieszczenie w tle obrazu badanego sygnału.

Ogólnie wszystkie obrazy sygnału mowy uzyskane z pomocą proponowanej w pracy metody (zarówno te uzyskane dla pojedynczych słów jak i te – znacznie lepsze – otrzymane dla izolowanych głosek) charakteryzują się tym, iż krzywa łącząca zwycięskie neurony jest w przypadku mowy patologicznej dłuższa niż w przypadku mowy prawidłowej. Własność ta dowodzi, że obrazy takie mogą stanowić dane wejściowe do dalszego etapu analizy sygnału (np. w systemie automatycznego rozpoznawania mowy patologicznej), lecz ich główna użyteczność wiąże się z faktem, że sygnał mowy pato-

logicznej, wizualizowany zgodnie z podanymi zasadami, może być podstawą oceny sytuacji i podejmowania decyzji przez lekarza – w sposób znacznie bardziej skuteczny, niż sygnał słuchany i oceniany subiektywnie na podstawie danych psychoakustycznych. Warto w tym miejscu przytoczyć dodatkowo znany fakt, że po okresie fascynacji automatyczną diagnostyką medyczną na całym świecie powraca się obecnie do technik, które ułatwiają lekarzom orientację w stanie pacjenta, jednak nie wyręczają ich w procesie oceny i podejmowania decyzji, gdyż przy wszystkich (bezspornych) zaletach automatycznej diagnostyki – to lekarz i tylko lekarz ponosi pełną odpowiedzialność za przyjmowane rozstrzygnięcia diagnostyczne i ich skutki. Właściwości opisanej tu metody wizualizacji sygnału mowy patologicznej **idealnie** pasują do tej nowej „filozofii" dzielenia zadań pomiędzy człowieka (lekarza) i wspomagający jego działania system komputerowy.

Przechodząc do „drobniejszych" wniosków z przeprowadzonych badań można stwierdzić, że małe sieci nie nadają się do poprawnej wizualizacji sygnałów mowy, bez względu na to czy obrazowane są słowa czy głoski. Wraz ze wzrostem rozmiaru sieci poprawia się zdolność dyskryminacji sygnałów patologicznych i wzorcowych. Można przyjąć (na podstawie przeprowadzonych badań), że minimalną wartością graniczną stopnia złożoności używanej sieci jest około 50 neuronów.

Jeżeli chodzi o liczbę kroków treningu, przy której sieć najlepiej dyskryminuje wypowiedzi patologiczne, to dla przypadku obrazowania głosek można tą wartość określić w miarę precyzyjnie. Wynosi ona około 200-300 kroków i w głównej mierze zależy od rodzaju głoski, a mniej od rozmiaru sieci. W przypadku obrazowania słów niemożliwe jest określenie tak precyzyjnej wartości, przeprowadzone eksperymenty wykazały, iż najlepszy efekt uzyskuje się dla 5000 do 20000 kroków treningu.

6. Podsumowanie

W pracy przedstawiono oryginalną metodę obrazowania sygnałów akustycznych zapisanych w postaci czasowo-częstotliwościowej. Metoda ta polega na zastosowaniu sieci neuronowej Kohonena do transformacji wielowymiarowego zjawiska, jakim jest sygnał mowy, do postaci dwuwymiarowego obrazu. Obraz taki powstaje poprzez wyznaczenie, dla każdego elementu wizualizowanego sygnału, zwycięskiego neuronu w sieci Kohonena, a następnie połączenie kolejnych lokalizacji tych „zwycięskich" neuronów. Elementami sygnału, dla których wyznacza się odpowiedzi sieci, są widma chwilowe kolejnych próbek czasowych sygnału, wyznaczone w procesie jego wstępnego (sprzętowego) przetwarzania. Jak z tego wynika, proces tworzenia (z pomocą sieci Kohonena) trajektorii charakteryzującej badaną wypowiedź, jest procesem dynamicznym – kolejne punkty (odpowiadające kolejnym „zwycięskim" neuronom), będą się pojawiały i będą łączone linią łamaną w pewnej kolejności i w pewnym tempie, zależnym także od dynamiki wypowiedzi. Można oczekiwać, że będzie to dodatkowym czynnikiem ułatwiającym analizę powstającego wykresu oraz ocenę, czy mamy tu do czynienia z próbką mowy prawidłowej czy patologicznej. Jednak w tej pracy abstrahowano od dynamicznego charakteru rozważanej formy wizualizacji i poprzestawano na ocenie ostatecznego kształtu powstającej linii łamanej, który okazał się także całkowicie wystarczający do tego aby różnicować mowę prawidłową od patologicznej i aby dokonywać pewnych ocen stopnia patologicznego zniekształcenia mowy celem np. wyboru właściwej techniki wykonania zabiegu operacyjnego korygującego uszkodzenie ciała lub wadę rozwojową będącą źródłem pojawiania się efektu mowy patologicznej.

Materiał badawczy stanowiły wypowiedzi dzieci z wadą rozwojową w zakresie narządów mowy (rozszczepem podniebienia) podzielone na 4 kategorie patologiczne; piątą kategorię tworzyły wypowiedzi wzorcowe. W obrębie każdej kategorii wyróżniono próbki sygnału, który stanowiły pojedyncze słowa oraz (w drugiej grupie badań) izolowane głoski, które były przedmiotem dalszych prac. Głównym celem badań było znalezienie takiego odwzorowania sygnału, które pozwala na wzrokową ocenę stopnia deformacji sygnału – w korelacji z procesami patologicznymi i zjawiskami patomorfo-

logicznymi, będącymi przyczyną deformacji sygnału. Dodatkowo przeprowadzono badania mające na celu określenie ogólnych własności generowanych obrazów, a także poszukiwano na drodze empirycznej optymalnych parametrów stosowanej sieci neuronowej w zależności od rodzaju obrazowanych fragmentów mowy (słowa, głoski). Optymalizacji podlegały elementy struktury sieci (rozmiar, zastosowanie – lub pominięcie – sąsiedztwa) a także elementy funkcjonalne (długość procesu uczenia).

W celu przeprowadzenia wspomnianych badań stworzono specjalny program komputerowy implementujący sieć neuronową Kohonena i realizujący przedstawiony w pracy sposób wizualizacji sygnałów akustycznych. Stworzono też narzędzia badawcze umożliwiające analizę uzyskiwanych obrazów, dzięki czemu można było ustalać, które rozwiązania w zakresie struktury i funkcji badanej sieci neuronowej lepiej spełniają postawione zadania. Wszystkie zamieszczone w pracy obrazy i tabele zostały wygenerowane przy pomocy tego oprogramowania.

W treści pracy pokazano, że sieć Kohonena, po odpowiednim treningu, może transformować sygnał dźwiękowy do postaci dwuwymiarowych obrazów, których kształt jest związany ze stopniem odkształcenia badanego sygnału w stosunku do wzorcowego przebiegu sygnału mowy prawidłowej. Powstałe w ten sposób obrazy mogą być po prostu przedmiotem wizualizacji i w ten sposób mogą służyć lekarzom jako narzędzie wspomagające w obiektywnej ocenie stopnia deformacji mowy, a w ślad za tym w podejmowaniu decyzji dotyczących diagnostyki, terapii, protezowania, selekcji zabiegów chirurgicznych, ukierunkowania rehabilitacji czy wreszcie prognozy stanu pacjenta po leczeniu. Nie jest to jednak jedyna możliwość, gdyż trajektorie generowane przez sieci Kohonena mogą także podlegać automatycznej klasyfikacji i rozpoznawaniu.

W pracy zaproponowano obiektywną metodę oceny takich obrazów, polegającą na pomiarze całkowitej długości krzywej reprezentującej daną wypowiedź. Przeprowadzone eksperymenty wykazały, iż całkowita długość krzywej w obrazach wypowiedzi patologicznych jest dłuższa niż w obrazach wypowiedzi wzorcowych. Własność ta może stanowić przesłankę do konstrukcji systemu automatycznej klasyfikacji i rozpoznawania wypowiedzi patologicznych.

Jako dalsze kierunki badań należy wskazać poprawienie sposobu wizualizacji całych słów oraz opracowanie metody pozwalającej na rozpoznawanie rodzaju patologii. Lepsze obrazowanie pojedynczych słów można uzyskać poprzez wizualizację, przy pomocy oddzielnych sieci, pojedynczych głosek, a następnie połączenie tych cząstkowych obrazów w jeden obraz reprezentujący całe słowo. W tym celu należy opracować metodę automatycznego podziału słowa na mniejsze fragmenty, przy czym z innych badań (nad rozpoznawaniem treści wypowiedzi) wynika, że problem segmentacji mowy

bywa bardziej złożony, niż problem jej rozpoznawania. Jako rozwiązanie tego problemu można zaproponować system hybrydowy złożony z dwóch sieci neuronowych, z których pierwsza dokonuje podziału słowa na mniejsze części, a druga (sieć Kohonena) wizualizuje te mniejsze fragmenty wypowiedzi.

Jeżeli chodzi o opracowanie metody pozwalającej na rozpoznawanie rodzaju patologii, to metoda taka może być oparta na wspomnianym wcześniej pomiarze długości trajektorii w obrazie. W takim przypadku ważna jest nie tylko maksymalizacja stosunku długości krzywej dla wypowiedzi patologicznych do długości dla wypowiedzi wzorcowych, ale także zróżnicowanie wartości tego stosunku dla różnych typów schorzeń odpowiedzialnych za generacje wypowiedzi patologicznych. Zróżnicowanie takie może stanowić następnie podstawę do konstrukcji systemu rozpoznającego w sposób automatyczny typ schorzenia dla danej wypowiedzi patologicznej. Alternatywnym podejściem do rozwiązania tego problemu jest zastosowanie jako danych treningowych próbek patologicznych. W takim przypadku trajektorie na obrazach wypowiedzi patologicznych należących do tej samej kategorii co dane treningowe powinny mieć najmniejszą długość. Dzięki temu możliwe byłoby ich oddzielenie od pozostałych wypowiedzi (wzorcowych lub charakteryzujących się patologią innego rodzaju). Na podstawie zamieszczonych w rozdziale 5 wykresów można przypuszczać, iż zaproponowane rozwiązania mają szansę powodzenia, jednak wymagają jeszcze dopracowania.

Podsumowując należy dodatkowo stwierdzić, że wprawdzie badania w tej pracy opierano na analizie sygnału mowy, jednak zaproponowana metoda wizualizacji może być zastosowana do obrazowania także dowolnego innego sygnału akustycznego lub procesu wibroakustycznego – na przykład dla potrzeb diagnostyki maszyn. Stwierdzenie takie jest prawdziwe, ponieważ wszystkie metody opracowane dla sygnału trudniejszego, bardziej złożonego, podlegającego bardziej krytycznej ocenie ze strony odbiorców (bo słuchacze są szczególnie wrażliwi na deformacje sygnału mowy), dają się zastosować dla sygnałów najogólniej mówiąc łatwiejszych – a przynajmniej część technicznych sygnałów akustycznych posiada parametry i charakterystyki bezspornie mniej skomplikowane, niż parametry sygnału mowy.

Uzyskanie i wypróbowanie efektywnego narzędzia nowej formy prezentacji wizualnej sygnału mowy patologicznej stanowiło najważniejszy wynik tej pracy, a cała jej zawartość świadczy, że postawioną na początku tezę – udało się wykazać.

Literatura

[Abu89] Abu-Mostafa Y., Psaltis D.: *Optical Neural Computers*. Scientific American, 1989

[Ata97] Atal B.S.: *Automatic Recognition of Speaker from Their Voices*. Proceedings of the IEEE Vol. 64, No. 4, April 1997

[Bas79] Basztura Cz., Jurkiewicz J., Tyburcy E.: *Zastosowanie fonetycznej funkcji mowy do segmentacji ciągłego sygnału mowy.* Archiwum Akustyki, t. 14, z. 2, 1979

[Bas87] Basztura Cz.: *Funkcje podobieństwa obrazów akustycznych jako wskaźniki obiektywnej oceny jakości transmisji mowy.* Archiwum Akustyki, t. 22, z. 3, 1987

[Bas88] Basztura Cz.: *Źródła, sygnały i obrazy akustyczne.* Wydawnictwo Komunikacji i Łączności, Warszawa 1988

[Bas91] Basztura Cz., Majewski W., Barycki W.: *Evaluation of word in simple automatic speech recognition system.* Archives of Acoustic, v. 16, 1991

[Bas93] Basztura Cz.: *Jak rozmawiać z komputerem ludzkim głosem.* Wydawnictwo Format, Wrocław 1993

[Bas95] Basztura Cz., Baściuk K.: *Aplikacyjne i obiektywne uwarunkowania identyfikacji i weryfikacji głosów w badaniach fonoskopijnych*. I Krajowa Konferencja Głosowa Komunikacja Człowiek-Komputer, Wrocław, październik 1995

[Bas96] Basztura Cz.: *Komputerowe systemy diagnostyki akustycznej.* Wydawnictwo Naukowe PWN, Warszawa 1996

[Bas99] Basztura Cz., Drygajło A.: *Wybrane zagadnienia badań fonoskopijnych.* W: Speech And Language Technology, Vol. 3, edited by: Jassem W., Basztura Cz., Demenko G., Jassem K., Poznań 1999

[Bax95] Baxt W.G.: *Application of artificial neural networks to clinical medicine.* Lancet, vol. 346, pp. 1135-1138, 1995

[Ben92] Bengio Y., De Mori R., Flammia G., Kompe R.: *Global Optimization of a Neural Network – Hidden Markov Model Hybrid*. IEEE Trans. On Neural Networks, vol. 3, no. 2, March 1992

[Ber90] Bergman A.S.: *Auditory Scene Analysis: The Perceptual Organization of Sound.* Bradford Books, MIT Press, Cambridge, Mass., 1990

[Bie64] Bień B.: *Wpływ górnych protez na wymowę.* Rozprawa doktorska, AM Zabrze, 1964

[Bis95] Bishop C.: *Neural networks for pattern recognition.* Clarendon Press, Oxford, 1995

[Bra99] Brachmański S., Staroniewicz P.: *Fonetyczna struktura materiału stosowanego w subiektywnych pomiarach jakości mowy.* W: Speech And Language Technology, Vol. 3, edited by: Jassem W., Basztura Cz., Demenko G., Jassem K., Poznań 1999

[Bri89] Bridle J.S.: *Probabilistic interpretation of feedforward classification network outputs, with relationship to statistical pattern recognition*. W: Neurocomputing: Algorithms, Architectures and Applications, edited by: Fougelman-Souile F., Herault J., Springer Verlag, 1989

[Bri90] Bridle J.S.: *Training stochastic model recognition algorithms as network can lead to maximum mutual information estimation of parameters*. W: Advances in Neural Information Processing Systems, edited by: Touretzky D.S., Morgan Kaufman, 1990

[Bro88] Broomhead D.S., Lowe D.: *Multivariable functional interpolation and adaptive networks*. Complex Systems, vol. 2, 1988

[Bur97] Burke H.B.: *Evaluating artificial neural networks for medical applications.* Proceedings of the 1997 International Conference on Neural Networks, pp. 2494-2496, Houston, USA, 1997

[Cam97] Campbell JR., Joseph P.: *Speaker Recognition: A Tutorial*. Proceedings of the IEEE Vol. 85, No. 9, September 1997

[Che95] Che C.W., Lin Q.: *Speaker recognition Using HMM with Experiments on the YOHO database*. Proceeding of EUROSPEECH, Madrid - Spain, September 1995

[Col96] Cole R. A., Mariani J., Uszkoreit H., Zaenen A., Zue V.: *Survey of the State of the Art in Human Language Technology.* http://www.cse.ogi.edu/ /CSLU/HLTsurvey/, 1996

[Com87] Combes J.M., Grossmann A., Tchamitchain P.: *Wavelets, Time-Frequency Methods and Phase Space.* Proceedings of the International Conference, Marseille, 1987

[Cyb89] Cybenko G.: *Approximation by Superpositions of Sigmoidal Function*. Mathematics of Control, Signals and Systems 2, 1989

[Dod97] Doddington G.R.: *Speaker Recognition – Identifying People by Their Voices*. Proceedings of the IEEE Vol. 76, September 1997

[Eng92] Engel Z., Tadeusiewicz R., Tosińska-Okrój H., Wszołek W.: *Wykorzystanie akustycznej analizy czasowo-częstotliwościowej zmienności sygnału mowy do oceny wyników operacyjnego leczenia dzieci z rozszczepem podniebienia.* XXXIX Otwarte Seminarium z Akustyki OSA-92, Kraków 1992

[Eng93] Engel Z.: *Ochrona środowiska przed drganiami i hałasem.* Wydawnictwo Naukowe PWN, Warszawa 1993

[Eng94] Engel Z, Tadeusiewicz R., Tosińska-Okrój H., Wszołek W.: *Analiza zmienności sygnału mowy jako metoda oceny wyników wybranej klasy zabiegów chirurgicznych.* Zeszyty Naukowe AGH, Mechanika, t. 12, z. 1, 1994

[Eng95] Engel Z., Tadeusiewicz R., Wszołek W.: *Estimation of Applicability of the Neural Networks for Automatic Evaluation of Pathological Speech.* 15th International Congress on Acoustics, Trondheim, Norway, June 1995

[Fab99] Fabian P., Migas A., Suszczańska N.: *Zastosowanie analizy morfologicznej i składniowej w procesie rozpoznawania mowy.* W: Speech And Language Technology, Vol. 3, edited by: Jassem W., Basztura Cz., Demenko G., Jassem K., Poznań 1999

[Far89] Farhart N.H.: *Optoelectronic Neural Network and Learning Machines*. IEEE Circuits and Devices Magazine, vol. 5, no. 9. 1989

[Fur81] Furui S.: *Cepstral Analysis Technique for Automatic Speaker Verification*. IEEE Trans. on ASSP, Vol. 29, 1981

[Fur89] Furui S.: *Digital Speech Processing, Synthesis, and Recognition.* Marcel Dekker, New York, 1989

[Fur96] Furui S.: *An overview of speaker recognition technology.* W: Automatic Speech and Speaker Recognition Advanced Topics, edited by: Lee C.H., Soong F.K., Kuldip K. Paliwal, Kluwer Academic Publishers, 1996

[Gaj99a] Gajer M., Kapusta M., Shomali A.: *Zastosowanie niejawnych modeli Markowa w systemach automatycznego rozpoznawania mowy*. Kwartalnik Elektrotechnika i Elektronika - AGH, zeszyt 3, Kraków, 1999

[Gaj99b] M. Gajer, M. Kapusta: *Rozpoznawanie obrazów z wykorzystaniem metod minimalnoodległościowych i sieci neuronowych*. Metrologia i Systemy Pomiarowe, Kwartalnik PAN, Tom VI, Zeszyt 3, 1999

[Gaj99c] Gajer M., Kapusta M., Shomali A.: *Visualisation and recognition of pathological utterances with the usage of the Kohonen neural networks.* V Krajowa Konferencja Zastosowań Matematyki w Biologii i Medycynie. Ustrzyki Górne 13-19 września 1999

[Gil96] Gilkey R., Anderson T.: *Binaural and Spatial Hearing.* Erlbaum, Hillsdale, New Jersey, 1996

[Gra89] Graf H.P., Jackel L.D.: *Analog Electronic Neural Network Circuits*. IEEE Circuits and Devices Magazine, vol. 5, no. 8. 1989.

[Gro76] Grossberg S.: *Adaptive pattern classification and universal recording. Parallel development and coding of neural feature detectors*. Biological Cybernetics, No. 23, 1976

[Gub80] Gubrynowicz R., Mikiel W., Żarnecki P.: *Akustyczna metoda oceny źródła krtaniowego w przypadkach zmian patologicznych fałdów głosowych.* Archiwum Akustyki, t. 15, 1980

[Gub81] Gubrynowicz R., Mikiel W., Żarnecki P.: *Acoustical analysis for the evaluation of laryngeal dysfunction in the case of vocal cord paralysis.* Speech Analysis and Synthesis, v. 5, IPPT PAN, Warszawa 1981

[Hał99] Hałas H.R.: *Automatyczne rozpoznawanie fonemów za pomocą sieci neuronowych*. Rozprawa doktorska, Politechnika Warszawska, 1999

[Ham90] Hampshire J.B., Waibel A.H.: *A Novel Objective Function for Improved Phoneme Recognition Using Time-Delay Neural Networks*. IEEE Trans. on Neural Networks, vol. 1, no. 2, June 1990

[Han94] Hanes M.D., Ahalt S.C., Krishnamurthy A.K.: *Acoustic-to-Phonetic Mapping Using Recurrent Neural Nets*. IEEE Trans. on Neural Networks, vol. 5, no. 4, July 1994

[Hay93] Haykin S., Ukrainec A.: *Neural networks for adaptive signal processing.* W: Adaptive system identification and signal processing algorithms, edited by: Kalouptsidis N., Theodoridis S., Prentice Hall, New York, 1993

[Hay94] Haykin S.: *Neural networks, a comprehensive foundation.* Macmillan College Publishing Company, New York, 1994

[Her93] Hertz J., Krogh A., Palmer R.G.: *Wstęp do teorii obliczeń neuronowych*. Wydawnictwo Naukowo-Techniczne, Warszawa 1993

[Hig91] Higgins A., Bahler L., Porter J.: *Voice Verification Using Randomized Phrase Prompting*. Digital Signal Processing, Vol. 1, 1991

[Hor91] Hornik K.: *Approximation capabilities of multilayer feedforward networks.* Neural Networks, vol. 4, no. 2, pp. 251-257, 1991

[Hou95] Houtsma A.J.M.: *Pitch perception.* W: Hearing, edited by: Moore B.C.J., Academic Press, Orlando, Florida, 1995

[Izw99] Izworski A., Wszołek W.: *Wykorzystanie metod sztucznej inteligencji w diagnostyce i przetwarzaniu patologicznych sygnałów mowy.* W: Speech And Language Technology, Vol. 3, edited by: Jassem W., Basztura Cz., Demenko G., Jassem K., Poznań 1999

[Jas73] Jassem W.: *Podstawy fonetyki akustycznej.* PWN, Warszawa 1973

[Jas79] Jassem W.: *Wielostopniowy model rozpoznawania akustycznego sygnału mowy.* Materiały XXVI Otwartego Seminarium z Akustyki, Wrocław-Oleśnica, 1979

[Kac76] Kacprowski J., Mikiel W., Szewczyk A.: *Akustyczne badania modelowe rozszczepu podniebienia.* Archiwum Akustyki, t. 11, z. 2, 1976

[Kac79] Kacprowski J.: *Obiektywne metody akustyczne w diagnostyce narządu głosu.* Archiwum Akustyki, t. 14, z. 4, 1979

[Kac95] Kacprzak T., Ślot K.: *Sieci neuronowe komórkowe*. Wydawnictwa Naukowe PWN, Łódź 1995

[Kap98] Kapusta M., Shomali A.: *Kierunki rozwoju systemów rozpoznawania mowy wynikające z postępu lingwistyki komputerowej.* Automatyka, Tom 2, Zeszyt 1, AGH - Kraków, 1998

[Kap99a] Kapusta M., Shomali A., Gajer M.: *Wizualizacja mowy patologicznej z zastosowaniem sieci neuronowych Kohonena.* Materiały Konferencyjne I Ogólnopolskie Warsztaty Doktoranckie OWD'99, Istebna-Zaolzie, 17-21 października 1999

[Kap99b] Kapusta M., Gajer M., Shomali A.: *Rozpoznawanie mowy patologicznej na podstawie obrazów głosek szumowych.* Kwartalnik Elektroniki i Telekomunikacji PAN, t. 45, z. 3-4, 1999

[Kap00a] Kapusta M., Gajer M., Shomali A.: *Wykorzystanie techniki obliczeń neuronowych do przetwarzania i rozpoznawania sygnałów mowy.* Miesięcznik PAK (w druku)

[Kap00b] Kapusta M., Gajer M., Shomali A.: *Recognition of Pathological Utterances with the Usage of the Kohonen Neural Networks.* Bulletin of the Polish Academy of Science, Technical Science, vol. 48, no. 1, 2000

[Kap00c] Kapusta M., Gajer M.: *Sieć Kohonena jako klasyfikator mowy patologicznej.* 21 Międzynarodowe Sympozjum Naukowe Studentów i Młodych Pracowników Nauki, Zielona Góra, 8-9 maja 2000

[Keh97] Kehagias A., Petridis V.: *Predictive modular neural networks for time series classifications.* Neural Networks, vol. 10, no. 1, pp. 31-49, 1997

[Koh84] Kohonen T.: *Self-Organisation and Associative Memory*. Springer Verlag, Heidelberg 1984

[Koh90a] Kohonen T.: *The Self-Organizing Map*. Proceedings of the IEEE, special issue on Neural Networks, vol. 78, no. 9, 1990

[Koh90b] Kohonen T.: *Unsupervised learning algorithms*. Proceedings of International Conference "Neural Networks – Biological Computers or Electronic Brains", Lyon 1990

[Kor94] Korbicz J., Obuchowicz A., Uciński D.: *Sztuczne sieci neuronowe, podstawy i zastosowania*. Akademicka Oficyna Wydawnicza PLJ, Warszawa 1994

[Ksi87a] Książkiewicz-Jóźwiak B., Wszołek W.: *Widmowa ocena stopnia deformacji sygnału mowy u pacjentów leczonych protezami całkowitymi.* Prot. Stom. XXXVII, 2, 1987

[Ksi87b] Książkiewicz-Jóźwiak B.: *Określenie optymalnego dla funkcji mowy ustawienia zębów przednich dolnych u pacjentów leczonych protezami całkowitymi.* Rozprawa doktorska, AM, Kraków 1987

[Kup96] Kupść A., Marciniak M., Mykowiecka A.: *Komputerowe przetwarzanie języka naturalnego - wybrane zagadnienia.* Informatyka, 1996, nr 7

[Kur98] Kurek J.: *Neural Net for Robot Manipulation*. Proceedings of ICSC Conference on Neural Computation, Vienna, September 1998

[LeC89] LeCun Y., Jackel L.D., Howard R.E., Hubbard W.: *Handwritten digit recognition: Applications of neural networks chips and automatic learning*. IEEE Communications Magazine, vol. 27, 1989

[LeC94] Le Cerf P., Ma W., Van Compernolle D.: *Multilayer Perceptrons as Labelers for Hidden Markov Models*. IEEE Trans. on ASSP, vol. 2, no. 1, part II, January 1994

[Łab97] Łabiszewska-Jaruzelska F.: *Ortopedia szczękowa, zasady i praktyka.* Wydawnictwo Lekarskie PZWL, Warszawa, 1997

[Maj79] Majewski W.: *Wybrane problemy rozpoznawania mówców.* Materiały XXVI Otwartego Seminarium z Akustyki, Wrocław-Oleśnica, 1979

[Mat93] Matsui T., Furui S.: *Concatenated Phoneme Models for Text Variable Speaker Recognition*. Proceedings of ICASSP, April 1993

[Mat95] Matsui T., Kanno T., Furui S.: *Speaker Recognition Using HMM Composition in Noisy Environments*. Proceeding of EUROSPEECH, Madrid - Spain, September 1995

[Mas92] Mason J.C., Parks P.C.: *Selection of neural networks structures: some approximation theory guidelines.* W: Neural networks for control and systems, edited by: Warwick K., Irwin G.W., Hunt K.J., Peter Peregrinus Ltd., London, 1992

[Mea89] Mead C.: *Analog VLSI and Neural Systems*. Addison Wesley, Reading, 1989

[McC43] McCulloch W.S., Pitts W.: *A logical calculus of the ideas immanent in nervous activity*. Bulletin of Mathematical Biophysics, No. 5, 1943

[Mic95] Micheli-Tzanaoku E.: *Neural networks in biomedical signal processing.* W: The Biomedical Engineering Handbook, edited by: Bronzino J., IEEE Press, Piscataway, NJ, 1995

[Moo95a] Moore B.C.J.: *Hearing.* Academic Press, San Diego, 1995

[Moo95b] Moore B.C.J.: *Perceptual Consequences of Cochlear Damage.* Allen and Unwin, London, 1995

[Moo99] Moore B.C.J.: *Wprowadzenie do psychologii słyszenia*. Wydawnictwo Naukowe PWN, Poznań, 1999

[Mor90] Morgan N., Wooters C., Bourland C., Cohen M.: *Continuous Speech Recognition on the Resource Management Database Using Connectionist Probability Estimation*. Proceedings of International Conference on Spoken Language Processing, Kobe, 1990

[Mor92] Morgan D.P., Scofield Ch.L.: *Neural Networks and Speech Processing.* Kluwer Academic Publishers, 1992

[Now85] Nowakowska W.: *Badania wpływu nazalizacji na strukturę widmową samogłosek i spółgłosek polskich.* Prace IPPT PAN, z. 46, Warszawa, 1985

[Nyg95] Nygaard L.C., Pisoni D.B.: *Speech Perception: New Directions in Research and Theory.* W: Speech, Language and Cognition, edited by: Miler J.L., Eimas P.D., Academic Press, San Diego, 1995

[Orr96] Orr M.: *Introduction to radial basis function networks.* Centre for Cognitive Science, University of Edinburgh, Edinburgh, 1996

[Oso96] Osowski S.: *Sieci neuronowe w ujęciu algorytmicznym.* Wydawnictwa Naukowo-Techniczne, Warszawa, 1996

[Pal92] Pal S.K., Mitra S.: *Multilayer Perceptron, Fuzzy Sets, and Classification*. IEEE Trans. on Neural Networks, vol. 3, no. 5, September 1992

[Pal95] Palmer A.R.: *Neural signal processing*. W: Hearing, edited by: Moore B.C.J., Academic Press, San Diego, 1995

[Pec85] Peciak J.: *Maskowanie i badania jakości metod maskowania sygnałów mowy*. Zeszyty Naukowe, Wyższa Szkoła Marynarki Wojennej, Gdańsk 1985

[Pic88] Pickles J.O.: *An Introduction to the Physiology of Hearing.* Second Edition, Academic Press, London, 1988

[Pla95] Plack C.J., Carlyon R.P.: *Loudness perception and intensity coding.* W: Handbook of Perception and Cognition, Volume 6. Hearing edited by: Moore B.C.J., Academic Press, Orlando, Florida, 1995

[Plo76] Plomp R.: *Aspects of Tone Sensation.* Academic, London, 1976

[Pog89] Poggio T., Girossi F.: *A Theory of Networks for Approximation and Learning*. Technical Report 1140, MIT AI Laboratory, Cambridge CA, 1989

[Psa88] Psaltis D., Park C.H., Hong J.: *Higher Order Associative Memories and Their Optical Implementations*. Neural Networks, vol. 1, 1988

[Psa89] Psaltis D., Brady D., Hsu K.: *Learning in optical neural computers*. International Joint Conference on Neural Networks, vol. 2, Washington DC, 1989

[Rab89] Rabiner L.R.: *A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition*. Proceedings of the IEEE, Vol. 77, No. 2, 1989

[Rab93] Rabiner L.R., Biing-Hwang J.: *Fundamentals of speech recognition.* Prentice Hall PTR, New Jersey 1993

[Ren92] Renals S., Morgan N., Cohen M., Franco H.: *Connectionist Probability Estimation in the DECIPHER Speech Recognition Systems*. IEEE Proceedings of ICASSP, San Francisco, 1992

[Rey95] Reynolds D., Carlson B.: *Text-Independent Speaker verification Using Decoupled and Integrated Speaker and Speech Recognizers*. Proceeding of EUROSPEECH, Madrid - Spain, September 1995

[Rig94] Rigoll G.: *Maximum Mutual Information Neural Networks for Hybrid Connectionist – HMM Speech Recognition Systems*. IEEE Trans. on ASSP, vol. 2, no. 1, part II, January 1994

[Rii97] Riis S.K., Krogh A.: *Hidden Neural Networks: A Framework for HMM/NN Hybrids*. IEEE Proceedings of ICASSP, Munich, 1997

[Rob94] Robinson A.J.: *An Application of Recurrent Neural Nets to Phone Probability Estimation*. IEEE Trans. on Neural Networks, vol. 5, no. 2, March 1994

[Ron95] Ronco E., Gawthrop P.: *Modular Neural Networks: a state of the art.* Technical Report CSC-95026, Centre for System and Control, University of Glasgow, Glasgow, 1995

[Ros91] Rosenberg A.E., Soong F.K. *Recent Research in Automatic Speaker Recognition*. W: Advances in Speech Signal Processing edited by: Furui S., Sondhi M., Marcel Dekker, New York, 1991

[Ros97] Rosenberg A.E.: *Automatic Speaker Verification: A Review*. Proceedings of the IEEE Vol. 64 No. 4, April 1997

[Rum86] Rumelhart D., Hinton G.E., Williams R.J.: *Learning Representation by Back-Propagating Errors*. Nature 323, 1986

[Rum91] Rumelhart D.: *Neural Networks – a Parallel Distributed Processing Perspective*. Proceedings of Neural Network Summer School – Theory, Design and Applications, Cambridge 1991

[Sap66] Sapożkow M.A.: *Sygnał mowy w telekomunikacji i cybernetyce.* WNT, Warszawa, 1966

[Sho99] Shomali A.: *Rozpoznawanie mówcy na podstawie długookresowego histogramu amplitud sygnału mowy.* Rozprawa doktorska, AGH - Kraków, 1999

[Sin92] Singer E., Lippmann R.: *A Speech Recognizer Using Radial Basis Function Neural Networks in a HMM Framework*. Proceedings of International Conference on ASSP, vol. 1, 1992

[Spe91] Specht D.F.: *A general regression neural network.* IEEE Transactions on Neural Networks, vol. 2, no. 6, 1991

[Str94] Ström N.: *Experiments with a New Method for Fast Speaker Adaptation in Continuous Speech Recognition.* Working Papers of the 9th Swedish Phonetics Conference, Sweden, 1994

[Suz89] Suzuki Y., Atlas L.: *A Study of Regular Architectures for Digital Implementation of Neural Networks*. Proceedings of IEEE International Symposium on Circuits and Systems, Portland 1989

[Świ97] Świercz M.: *Zastosowanie sieci neuronowych do modelowania wybranych systemów biomedycznych.* Politechnika Białostocka, Rozprawy naukowe nr 51, Białystok, 1997

[Tad75] Tadeusiewicz R.: *KART – 1 – konwerter do bezpośredniego wczytywania informacji akustycznej do maszyny cyfrowej.* Archiwum Akustyki, t. 10, 1975

[Tad78] Tadeusiewicz R.: *Głosowa łączność człowieka z maszyną cyfrową.* Zeszyty Naukowe AGH, Automatyka, Nr 709 z. 22, 1978

[Tad83] Tadeusiewicz R.: *Ocena przydatności wybranych metryk w minimalno odległościowych metodach rozpoznawania mowy.* Archiwum Akustyki, t. 18, 1983

[Tad85a] Tadeusiewicz R., Książkiewicz-Jóźwiak B., Wszołek W.: *Akustyczne metody porównywania i oceny sygnału mowy w protetyce chirurgii szczękowo-twarzowej.* OSA-85, Kraków 1985

[Tad85b] Tadeusiewicz R.. Książkiewicz-Jóźwiak B., Wszołek W.: *Widmowe kryteria oceny stopnia deformacji sygnału mowy w protetyce stomatologicznej.* VII Konferencja Naukowo-Szkoleniowa, Biocybernetyka i Inżynieria Biomedyczna, Gdańsk 1985

[Tad88a] Tadeusiewicz R.: *Sygnał mowy.* Wydawnictwo Komunikacji i Łączności, Warszawa 1988

[Tad88b] Tadeusiewicz R., Wszołek W.: *Miary podobieństwa sygnałów akustycznych w zastosowaniu do oceny mowy patologicznej.* OSA, Białowieża 1988

[Tad91] Tadeusiewicz R., Flasiński M.: *Rozpoznawanie obrazów*. PWN, Warszawa 1991

[Tad93] Tadeusiewicz R.: *Sieci Neuronowe*. Akademicka Oficyna Wydawnicza RM, Warszawa, 1993

[Tad94] Tadeusiewicz R.: *Problemy biocybernetyki*. Wydawnictwo Naukowe PWN, Warszawa, 1994

[Tad97] Tadeusiewicz R., Izworski A., Wszołek W.: *Pathological Speech Evaluation Using the Artificial Intelligence Methods*. World Congress on Medical Physics and Biomedical Engineering, Nice, France, September 1997

[Tad98a] Tadeusiewicz R., Wszołek W., Izworski A.: *Application of Neural Networks in Diagnosis of Pathological Speech*. International ICSC/IFAC Symposium On Neural Computation, Vienna, Austria, September 1998

[Tad98b] Tadeusiewicz R.: *Elementarne wprowadzenie do techniki sieci neuronowych z przykładowymi programami*. Akademicka Oficyna Wydawnicza PLJ, Warszawa, 1998

[Tad99] Tadeusiewicz R., Wszołek W., Izworski A., Wszołek T.: *Methods of Pathological Speech Visualization*. Proceedings of The First Joint Meeting of BMES & EMBS Serving Humanity, Advancing Technology, Atlanta, USA, 13-16 Oct. 1999

[Thi95] Thimm G., Fiesler E.: *Neural network initialization*. W: From Natural to Artificial Neural Computation, edited by: Mıra J., Sandoval F., IWAN, Malaga, 1995

[Tis91] Tishby N.Z.: *On the Application of Mixture AR Hidden Markov Models to Text Independent Speaker Recognition*. IEEE Trans. on Signal Processing, Vol. 39, No. 3, March 1991

[Vem94] Vemuri V.R., Rogers R.D.: *Artificial neural networks: forecasting time series*. IEEE Computer Society Press, Los Alamitos, CA, 1994

[Vie93] Viemeister N.F., Plack C.J.: *Time analysis*. W: Human Psychophysics, edited by: Yost W.A., Popper A.N., Fay R.R., Springer-Verlag, New York, 1993

[Wag87] Wagner K., Psaltis D.: *Multilayer optical neural networks*. Applied Optics, vol. 26, 1987

[Wai89a] Waibel A., Hanazawa T., Hinton G., Shikano K., Lang K.J.: *Phoneme Recognition Using Time-Delay Neural Networks*. IEEE Trans. on ASSP, vol. 37, no. 3, March 1989

[Wai89b] Waibel A., Sawai H., Shikano K.: *Modularity and Scaling in Large Phonetic Neural Networks*. IEEE Trans. on ASSP, vol. 37, no. 12, December 1989

[Wei90] Weigand A.S., Rumelhart D.E., Huberman B.A.: *Back-propagation weight elimination and time series prediction.* Proceedings of the Connectionist Models Summer School, pp. 105-116, 1990

[Wid60] Widrow B., Hoff M.E.: *Adaptive switching circuits*. IRE WESCON Convention Record, New York 1960

[Wid88] Widrow B., Winter R., Baxter R.: *Layered neural nets for pattern recognition*. IEEE Trans. on ASSP, vol. 36, 1988

[Wsz91] Wszołek W., Zabawa P.: *Zastosowanie metod rozpoznawania obrazów w badaniu zniekształceń sygnału mowy.* Zeszyty Naukowe AGH, Mechanika, t. 10, z. 2, 1991

[Wsz94] Wszołek W.: *Analiza i ocena obrazów dźwiękowych w procesach wibroakustycznych.* Rozprawa doktorska, AGH, Kraków 1994

[Wu94] Wu L., Niranjan M., Fallside F.: *Fully Vector-quantized Neural Network-Based Code-Excited Non-linear Predictive Speech Coding*. IEEE Trans. on Speech and Audio Processing, vol. 2, October 1994

[Zav94] Zavaliagkos G., Zhao Y., Shwartz R., Makhoul J.: *A Hybrid Segmental Neural Net/Hidden Markov Model System for Continuous Speech Recognition*. IEEE Trans. on ASSP, vol. 2, no. 1, part II, January 1994

[Zub94] Zuben F., Netto M.A.: *Exploring the nonlinear dynamics behavior of artificial neural networks*. IEEE Proceedings ICNN, Orlando, 1994