

Wiktor Jassem, Wim van Dommelen

AKCENT JĘZYKA POLSKIEGO
W PERCEPCJI
SYNTETYCZNEGO SYGNAŁU MOWY

1/1989

P.269



WARSZAWA 1989

Praca wpłynęła do Redakcji dnia 14 kwietnia 1988 r.



56744



Na prawach rękopisu

Instytut Podstawowych Problemów Techniki PAN

Nakład 170 egz. Ark.wyd. 1,5 Ark.druk. 2

Oddano do drukarni w styczniu 1989 r.

Nr zamówienia 171/89

Warszawska Drukarnia Naukowa, Warszawa,
ul.Sniadeckich 8

Wiktor Jassem
Pracownia Fonetyki Akustycznej
Instytut Podstawowych Problemów Techniki
PAN

Wim van Dommelen
Institut für Phonetik und
digitale Sprachverarbeitung
der Christian-Albrechts Universität
Kiel

AKCENT JĘZYKA POLSKIEGO W PERCEPCJI SYNTETYCZNEGO SYGNAŁU MOWY ¹

Streszczenie

Wypowiedź /widzi¹eteras/ wymówiono w trzech wersjach: z akcentem na (I) pierwszej, (II) drugiej oraz (IV) czwartej sylabie. Korzystając z pakietu programów przetwarzania mowy wykonującego analizę i syntezę za pomocą predykcji liniowej, wymieniono wzajemnie w dwóch doświadczeniach odsłuchowych parametr częstotliwości podstawowej pierwotnych wypowiedzi. W dalszych testach uśredniano kilku metodami przebiegi F_0 oraz zastępowano je bądź przez szum, bądź przez stałą wartość częstotliwości podstawowej. Doświadczenia odsłuchowe przeprowadzono z grupą (1) studentów niemieckich, (2) studentów polskich oraz (3) polskich pracowników naukowych i naukowo-technicznych. Słuchacze grupy (1), nie znający języka polskiego, ale specjalizujący się w fonetyce, lokalizowali akcent niemal dokładnie tak samo, jak ci słuchacze polscy, których reakcje były systematyczne. Decydujący wpływ na percepcję miejsca akcentu miał przebieg parametru F_0 , a inne parametry umożliwiały w bardzo ograniczonym zakresie lokalizację akcentu, gdy parametr F_0 zastąpiono szumem. Zastąpienie zmiennej wartości F_0 przez wartość stałą uniemożliwiło prawie całkowicie identyfikację sylaby akcentowanej.

¹ Praca została wykonana częściowo w ramach CPBP O2.13, a częściowo w ramach planu badawczego Instytutu Fonetyki i Cyfrowego Przetwarzania Mowy Uniwersytetu w Kilonii.

1. Wstęp

Istota akcentu należy do najbardziej kontrowersyjnych problemów współczesnej fonetyki doświadczalnej. Akademicki podręcznik fonetyki doświadczalnej, wciąż jeszcze uważany za wzorcowy (Ladefoged 1975), zawiera następującą definicję akcentu: "Sylaba akcentowana jest wymawiana z większym nakładem energii niż sylaba nieakcentowana" (str. 222, tłum. WJ). Autor *explicit*e uważa powyższe stwierdzenie za definicję akcentu. W słowniku terminologicznym w tejże pracy również podany jest opis akcentu jako pojawienie się "zwiększonej energii w określonej sylabie" (str. 281, tłum. WJ). Ladefoged nie przytacza żadnych, choćby przykładowych danych doświadczalnych na poparcie tej - dość zresztą często przez lingwistów przyjmowanej - definicji.

W trzech ogólnych opisach wymowy polskiej, które ukazywały się w okresie ostatnich 30 lat, Wierzchowska (1960, 1967 i 1980) konsekwentnie określa akcent polski jako "dynamiczny" ("ekspiratoryczny"), polegający na zwiększeniu "siły" (wysiłku głosowego) w tej części wypowiedzi, która jest akcentowana (1965 str. 157, 1967 str. 17, 1980 str. 133-134). Operując parametrami akustycznymi należy oczekiwać, iż akcentowana sylaba, *caeteris paribus*, wykaże pomiarowo większy poziom ciśnienia niż nieakcentowana.

Już ok. 30 lat temu jeden z autorów niniejszej pracy (Jassem 1959 i 1962) próbował wykazać na podstawie pomiarów trzech zmiennych parametrów fonetyczno-akustycznych: częstotliwości podstawowej, obwiedni czasowej poziomu ciśnienia oraz iloczasu segmentalnego, że o miejscu akcentu w języku polskim decyduje określony przebieg F_0 (niekoniecznie lokalne maksimum). Następnie zespół Jassem, Morton i Steffen-Batogowa (1968) przedstawił wyniki doświadczeń wykonanych na materiale syntetycznego sygnału mowopodobnego, które wskazywały na dominującą rolę parametru F_0 w percepcji akcentu przez polskich słuchaczy. Jednak 10 lat później Dobrogowska (1978) stwierdziła na podstawie pomiarów przebiegu F_0 , że ten parametr nie jest jednoznacznie związany z akcentem mowy. Sprawa istoty akcentu polskiego określonego w terminach akustycznych pozostawała zatem problemem otwartym. Niniejsza praca podejmuje to zagadnienie raz jeszcze, ponieważ jest ono obecnie szczególnie

aktualne ze względu na prowadzone prace nad elektroniczną syntezą i automatycznym rozpoznawaniem mowy polskiej.

Wynikiem ubocznym (ale istotnym) niżej przedstawionych doświadczeń było stwierdzenie konieczności ustalenia jednorodności reakcji słuchaczy w eksperymentach percepcyjnych, których przedmiotem są fonetyczno-akustyczne cechy sygnału mowy.

2. Akcent potencjalny i akcent realny

W literaturze fonetycznej stosowane są tradycyjnie terminy "akcent wyrazowy" oraz "akcent zdaniowy". Nazwy te, odnoszące się do wyróżniania określonej sylaby w obrębie wyrazu oraz zdania, należy uznać za niewłaściwe przede wszystkim dlatego, że wyraz i zdanie nie są elementami fonetycznymi, lecz morfologicznymi i składniowymi, należąc do innej płaszczyzny analizy. W pracy Jassema i Gibbona (1980) przedstawiono propozycję wprowadzenia do analizy fonetycznej i fonologicznej terminów angielskich "stress" oraz "accent" w sensie nowych, ale nawiązujących do prac Bolingera (1955 a,b; 1958) definicji. Są one niezależne od pojęć wyrazu i zdania, ale pozostają z nimi w określonych związkach - ściślej odpowiednio z wyrazem oraz z frazą intonacyjną. Akcent potencjalny określono jako miejsce, w którym zachodzi możliwość wyróżnienia sylaby, zaś akcent realny jako rzeczywiste jej wyróżnienie, przy czym dla języka angielskiego akcent realny stanowią w płaszczyźnie fonetyczno-akustycznej określone relacje iloczynowe oraz określone przebiegi parametru F_0 . Akcent realny sygnalizowany wyłącznie relacjami iloczynowymi jest podrzędny względem akcentu sygnalizowanego również specyficznymi przebiegami częstotliwości podstawowej.

W niniejszej pracy będziemy stosować pojęcie akcentu potencjalnego i realnego zgodnie z propozycjami przedstawionymi w pracy Jassema i Gibbona (1980).

3. Materiał doświadczalny

Ciąg segmentów, który w transkrypcji fonematycznej można zapisać /vidi, itē, eteras/ został wymówiony przez WJ (jednego z autorów) w studio o silnie ograniczonym pogłosie i zapisany na taśmie magnetofonowej przy zastosowaniu aparatury wysokiej jakości, w trzech wersjach różniących się miejscem realnego

akcentu tonalnego, który padał w wersji I na pierwszą sylabę, w wersji II na drugą, a w wersji IV na czwartą sylabę. Wypowiedziom tym odpowiadać może bądź ortograficzna postać Widzi cię teraz, bądź Widzicie teraz. W obu wersjach występują dwa akcenty potencjalne: /vi/ i /te/ lub /t̥e/ i /te/.

Korzystając z programów analizy i syntezy mowy stosowanych w Instytucie Fonetyki i Cyfrowego Przetwarzania Mowy Uniwersytetu w Kilonii, opartych na kodowaniu liniowo-predykcyjnym (Rathjen 1979; Schafer-Vincent 1972; Barry et al. 1982), przeprowadzono syntezę wypowiedzi stanowiącą materiał doświadczalny sześciu eksperymentów percepcyjnych. W obrębie każdego eksperymentu każdy bodziec poprzedzony był sygnałem sinusoidalnym o częstotliwości 200 Hz i czasie trwania 100 ms, a po każdym bodźcu następowała przerwa czterosekundowa, która była w doświadczeniu percepcyjnym wykorzystywana przez słuchaczy na podjęcie określonej decyzji. Dla orientacji słuchaczy dodatkowe, krótkie sygnały sinusoidalne zapisano w odstępach co 10 bodźców.

Eksperyment 1

Spreparowano trzy bodźce: I - resyntezę wypowiedzi I z zachowaniem wartości wszystkich parametrów analizy, I_{II} - resyntezę wypowiedzi I z zachowaniem wszystkich parametrów z wyjątkiem F₀, które - po nieznacznym dopasowaniu liniowym - przeniesiono z wypowiedzi II, oraz I_{IV} - resyntezę wypowiedzi I z zachowaniem wartości wszystkich parametrów z wyjątkiem F₀, które - po nieznacznym dopasowaniu liniowym - przeniesiono z wypowiedzi IV. Każdy bodziec (I_I, I_{II}, I_{IV}) został skopiowany 10-krotnie, dając łącznie 30 bodźców. Rys. 1/2 przedstawia zapis wyników analizy LPC wypowiedzi II (z akcentem realnym na drugiej sylabie). Uzyskanymi parametrami fonetycznymi są: F₀, częstotliwość oraz szerokość wstęgi i poziom pięciu lokalnych maksimum obwiedni widma krótkoterminowego (formantów) tudzież obwiednia czasowa poziomu ciśnienia.

Eksperyment 2

Parametr F₀ we wszystkich trzech pierwotnych wypowiedziach zastąpiono szumem białym. Wykonano po 10 kopii każdego syntetycznego bodźca I_S, II_S, IV_S. Stąd łączna liczba 30

bodźców w eksperymencie.

Eksperyment 3

Bodźce były tutaj analogiczne do występujących w eksperymencie 1, z tym, że dodatkowo przebieg F_0 z wypowiedzi I oraz IV nałożono na wypowiedź II, F_0 z wypowiedzi I i II nałożono na wypowiedź IV, a ponadto zresyntetyzowano bez zmian parametrów wypowiedzi II i IV uzyskując w rezultacie tych manipulacji 9 różnych bodźców: $I_I, I_{II}, I_{IV}, II_I, II_{II}, II_{IV}, IV_I, IV_{II}, IV_{IV}$. Wykonano po 10 kopii na łączną liczbę 90 bodźców.

Eksperyment 4

Pierwotne wypowiedzi ułożono w następujące pary:

(1) $I_I : I_{II}$, (2) $I_I : I_{IV}$ oraz (3) $I_{II} : I_{IV}$. W obrębie każdej pary zresyntetyzowano ponadto po trzy wersje pośrednie, stanowiące stopniowe przejście od pierwszego do drugiego członu pary. Powstały w ten sposób trzy zbiory bodźców, po pięć w każdym zbiorze: dwa skrajne (np. I_I, I_{IV}) oraz trzy pośrednie. Elementy każdego z tych zbiorów oznaczono, w sposób możliwie prosty mnemotechnicznie, jak następuje: $I_I, (I_I:I_{II})_2, (I_I:I_{II})_3, (I_I:I_{II})_4, I_{II}$ oraz analogicznie dla pozostałych dwu piątek. Jak poprzednio liczba rzymska wskazuje na to, która sylaba była akcentowana w pierwotnej wypowiedzi, przy czym liczba podstawowa odnosi się do wszystkich parametrów syntezy prócz F_0 , a liczba w indeksie do wypowiedzi, z której transplutowano przebieg F_0 . Pośrednie trzy przebiegi otrzymano syntetycznie w ten sposób, że dla każdego przedziału czasowego LPC ("framy") obliczono wartość F_0 według wzorów:

$$F_0(I_I:I_{II})_2 = F_0(I_I + \frac{1}{4} I_I - I_{II}),$$

$$F_0(I_I:I_{II})_3 = F_0(I_I + \frac{1}{2} I_I - I_{II}),$$

$$F_0(I_I:I_{II})_4 = F_0(I_I + \frac{3}{4} I_I - I_{II}) \text{ i analogicznie}$$

dla $F_0(I_I:I_{IV})_{2,3,4}$ oraz $F_0(I_{II}:I_{IV})_{2,3,4}$.

Eksperyment 5

Ten eksperyment stanowi rozszerzenie eksperymentu 2 z tym, że dodatkowo parametr obwiedni czasowej poziomu w pierwotnej

wypowiedzi II zastąpiono przebiegiem tego parametru z wypowiedzi I, natomiast pierwotny przebieg tego parametru w wypowiedzi IV zastąpiono odpowiednio przebiegiem z pierwotnej wypowiedzi II. Każdy z 5 bodźców: I_S , II_S , IV_S , $II_{S(I)}$, $IV_{S(I)}$ skopiowano 5-krotnie (łącznie 25 bodźców w eksperymencie).

Eksperyment 6

We wszystkich trzech pierwotnych wypowiedziach zastąpiono zmienne przebiegi F_0 stałą wartością $F_0 = \text{const} = 100 \text{ Hz}$. Wykonano po 10 replikacji.

Wszystkie bodźce w obrębie każdego z 6 eksperymentów zrandomizowano korzystając ze standardowego programu generującego liczby losowe.

4. Doświadczenia odsłuchowe

Zresyntetyzowany materiał doświadczalny opisany w poprzednim rozdziale poddano testom odsłuchowym przez trzy grupy słuchaczy:

(1) Grupa I - 8 studentów narodowości niemieckiej na IV i V roku filologii Uniwersytetu w Kilonii, specjalizujących się częściowo w fonetyce. Słuchacze ci mieli za sobą kilkuletni staż obejmujący m.in. słuchowe ćwiczenia fonetyczne, ale nie znali polskiego języka, ani nie posiadali teoretycznej wiedzy z zakresu polskiej akcentuacji;

(2) Grupa II - 9 polskich studentów różnych wydziałów, z których żaden nie miał wykształcenia ani z zakresu fonetyki, ani lingwistyki;

(3) Grupa III - 10 pracowników naukowych i naukowo-technicznych, obejmująca osoby z wykształceniem fonetycznym, jak i bez takiego wykształcenia.

Z jednym wyjątkiem żadna z osób biorących udział w odsłuchach nie miała problemów medycznych z narządem słuchu. Owego jedynego wyjątku dotyczyła osoba w grupie III. Należy ona do podgrupy z (ograniczonym) wykształceniem fonetycznym. Ta osoba (HK) ma średni jednostronny ubytek słuchu, ale jej wyniki nie odbiegały od pozostałych osób w tej podgrupie.

Doświadczenia odsłuchowe przebiegały tak samo we wszystkich trzech grupach.

Każdy słuchacz otrzymał arkusze, na których każdy bodziec

oznaczono kolejnym numerem. Przy każdym numerze nakreślono pięć kresek poziomych oznaczających kolejne sylaby:

1. _____
2. _____
3. _____
:
40. _____
:
:

Słuchaczy poproszono, aby zaznaczyli sylaby, które słyszą jako akcentowane, według następujących wytycznych: Jeśli w danym bodźcu tylko jedna sylaba jest akcentowana, nad odpowiadającą tej sylabie kreską należy postawić krzyżyk. Jeśli dwie sylaby odbierane są jako akcentowane, słuchacz ma wybór. O ile jedna sylaba słyszana jest jako silniej akcentowana, to daną sylabę należy zaznaczyć krzyżykiem, a pozostałą sylabę, słabiej akcentowaną - pojedynczą kreską pionową. O ile dwie sylaby odbierane są jako równo silnie akcentowane, obie zaznacza się krzyżykiem. Krzyżyk interpretowany był następnie jako akcent główny, a kreska pionowa jako akcent poboczny. O ile słuchacz nie mógł podjąć decyzji, był proszony o postawienie krzyżyka na dowolnej sylabie w sposób przypadkowy.

5. Wyniki doświadczenia

5.1. Jednorodność ocen

Szczegóły przeprowadzonej analizy jednorodności ocen zamieszczone są w Dodatku A. Tutaj ograniczymy się do stwierdzenia, że grupa I stawiała akcenty na tyle konsekwentnie i zgodnie - mimo randomizacji - że uznano ją w całości za jednorodną. W grupie II 3 osoby zachowywały się konsekwentnie, a pozostałe chaotycznie (oceny były przypadkowe dla replikacji tych samych bodźców), wobec czego tylko 3 osoby wzięto do zbioru finalnego. Te trzy osoby, podobnie jak pozostałe, nie odbywały uprzednio - o czym wspominaliśmy wyżej - żadnego treningu fonetycznego, wobec czego drastyczną różnicę w ich wynikach w stosunku do wyników pozostałych osób w tej grupie należy przypisać takim dodatkowym czynnikom psychicznym jak motywacja, koncentracja,

zainteresowanie itp. W grupie III również w sposób drastyczny wyróżniły się dwie podgrupy: słuchaczy konsekwentnych (7 osób) i niekonsekwentnych (3 osoby). Tym razem jednak pierwszą podgrupę stanowili bądź specjaliści fonetycy lub lingwiści, bądź osoby skądinąd obyte z odsłuchowymi doświadczeniami fonetycznymi. Pozostałe osoby, w tej grupie takich kwalifikacji nie posiadały.

W ten sposób z pierwotnych trzech grup słuchaczy I, II i III powstały dwie: G i P, obie homogeniczne w sensie reakcji na bodźce. Grupa G obejmowała wszystkich słuchaczy niemieckich (8 osób), zaś grupa P - systematycznie reagujących słuchaczy polskich (10 osób).

5.2. Statystyczna analiza ocen systematycznych

Zmienną losową stanowi w naszym przypadku percepcyjna ocena stopnia akcentu. Występuje ona w trzech wartościach: akcent główny, oznaczony w dalszym ciągu przez * * * , akcent poboczny, oznaczony przez * * oraz brak akcentu, oznaczony przez * . Należy zbadać rozkład tej zmiennej w zależności od trzech warunków: (1) Rodzimego języka słuchacza - dwa stany: G oraz P, (2) pozycji sylaby w wypowiedzi - pięć stanów: sylaba 1, ..., 5, oraz (3) modyfikacji parametru F_0 w resyntezie - 3, 4 lub 5 stanów w zależności od sposobu modyfikacji ("transplantacji") parametru w poszczególnych eksperymentach 1, ..., 6, jak to opisano w poprzednim rozdziale.

Z punktu widzenia warunków (1) i (2) wyniki odsłuchów potraktowano oddzielnie dla każdego stanu, natomiast warunek (3) uznano za zmienną niezależną i zbadano rozkład wartości zmiennej losowej względem poszczególnych stanów warunku (3) stosując zmodyfikowany test χ^2 dla badania niezależności w tablicach kontyngencji. Układ tablic kontyngencji ilustrujemy poniższymi przykładami:

eksperyment 1				eksperyment 1			
grupa G/sylaba 1				grupa G/sylaba 2			
	***	**	*		***	**	*
I	70	0	10	I	3	6	71
I _{II}	1	2	77	I _{II}	79	0	1
I _{IV}	6	13	61	I _{IV}	12	1	67

eksperyment 1			
grupa G/sylaba 3			
	***	**	*
I	6	1	73
I _{II}	0	2	78
I _{IV}	3	3	74

eksperyment 1			
grupa G/sylaba 4			
	***	**	*
I	3	6	69
I _{II}	4	11	65
I _{IV}	62	9	9

eksperyment 1
grupa G/sylaba 5

	***	**	*
I	0	13	67
I _{II}	1	16	63
I _{IV}	0	3	77

W powyższych tablicach I_I oznacza zresyntetyzowaną wypowiedź I bez zmian parametrów, I_{II} - transplantację przebiegu F₀ z wypowiedzi II do wypowiedzi I przy nie zmienionych wartościach pozostałych parametrów, I_{IV} - transplantację F₀ z wypowiedzi IV do wypowiedzi I przy nie zmienionych pozostałych parametrach (por. wyż.).

Wyniki zawarte w powyższych tablicach kontyngencji można opisowo przedstawić, jak następuje:

(1) Sylaba 1 jest silnie obciążona w kierunku wartości *** pod warunkiem I_I oraz silnie obciążona w kierunku wartości * pod warunkami I_{II} i I_{IV},

(2) Sylaba 2 jest silnie obciążona w kierunku wartości *** pod warunkiem I_{II} oraz silnie obciążona w kierunku wartości * pod warunkami I_I oraz I_{IV},

(3) Sylaba 3 jest silnie obciążona w kierunku wartości * pod wszystkimi trzema warunkami,

(4) Sylaba 4 jest silnie obciążona w kierunku wartości *** pod warunkiem I_{IV} oraz silnie obciążona w kierunku wartości * pod warunkami I_I oraz I_{II},

(5) Sylaba 5 jest silnie obciążona w kierunku wartości * pod wszystkimi trzema warunkami.

Z powyższych obserwacji wynika, że oznaczenie przez słuchaczy sylaby 1, 2 lub 4 jako posiadającej akcent główny zależy w decydującym stopniu od przebiegu parametru F₀, jako że

podstawienie F_0 z wypowiedzi II i IV do wypowiedzi I spowodowało taką lokalizację akcentu, jaka występowała pierwotnie w wypowiedziach II i IV mimo, że pozostałe parametry nie podlegały zmianie. Natomiast sylaby 3 i 5 były konsekwentnie percypowane jako nieakcentowane bez względu na to, który z trzech przebiegów F_0 zawiera wypowiedź syntetyczna.

Chociaż liczby w tablicach kontyngencji wskazują w sposób intuicyjnie nie budzący wątpliwości na prawidłowość wyciągniętych wyżej wniosków, jest rzeczą interesującą określenie w sposób statystyczny prawdopodobieństwa hipotezy zerowej o niezależności rozkładu wartości ~~xxx~~, ~~xx~~ i ~~x~~ od zmienności warunków (3). Do zbadania takiej niezależności nadaje się test χ^2 . Jednakże w klasycznym ujęciu zastosowanie testu χ^2 do pełnej tablicy kontyngencji jest uzależnione od tego, że poszczególne klatki nie zawierają wielkości mniejszych niż 8 (por. np. Greń 1974;131) lub - według innych autorów - mniejszych niż 5 (np. Volk 1973;95). Jeśli ten warunek nie jest spełniony, zachodzi konieczność łączenia klas w tablicy kontyngencji, co powoduje utratę uzyskanej informacji. Ponieważ żadna z powyższych tablic - podobnie jak prawie wszystkie pozostałe tablice w całości doświadczenia - nie spełnia tego warunku, zastosowano modyfikację opisaną w pracach Nassai (1959) oraz Potthoffa Wittinghilla (1966). Odpowiedni algorytm obliczeniowy podały Dobek i Kiełczewska (1977), a program na maszynę cyfrową systemu JS (Riad 32) zrealizowano w Ośrodku Obliczeniowym Uniwersytetu A. Mickiewicza w Poznaniu. Istotę zmodyfikowanego testu χ^2 przedstawiono w Dodatku B.

Z grupą G przeprowadzono eksperymenty 1...5, a grupą P eksperymenty 1...6. Skonstruowano 45 tablic kontyngencji dla grupy G i 48 dla grupy P, każdą tablicę poddając zmodyfikowanemu testowi χ^2 . Na podstawie obliczonych dla poszczególnych tablic wartości zmodyfikowanego χ^2 oraz odpowiednich wartości stopni swobody (zob. Dodatek B), określono z tablic statystycznych poziom istotności dla odrzucenia hipotezy zerowej.

Łącznie dla grupy G hipoteza zerowa o braku zależności rozkładu zmiennej losowej od warunków (3), tj. zmian F_0 , została odrzucona w odniesieniu do 2 tablic na poziomie istotności = 0.01, w odniesieniu do 19 tablic nie można było wiarygodnie podać poziomu istotności, gdyż zmodyfikowana liczba stopni swobody była mniejsza niż 1. W odniesieniu do pozostałych 24 tablic

hipotezę zerową odrzucono na poziomie $\alpha = 0.001$. Dla grupy P hipotezę zerową odrzucono na poziomie istotności $\alpha = 0.01$ w jednym przypadku. Ponownie dla 19 tablic liczba stopni swobody wynosiła mniej niż 1, wobec czego nie określono poziomu istotności. W odniesieniu do pozostałych 28 tablic hipotezę zerową odrzucono w tej grupie na poziomie $\alpha = 0.001$.

W obu grupach prawie wszystkie tablice dające liczbę stopni swobody mniejszą niż 1 odnosiły się do sylab 3 i 5. Są to sylaby nie posiadające akcentu potencjalnego i zostały wymówione we wszystkich trzech wersjach (tj. I, II, IV) bez akcentu realnego. W tych przypadkach test χ^2 czy jakkolwiek inny nie jest w rzeczywistości potrzebny gdyż mamy tu do czynienia z przypadkami oczywistymi poniższego typu:

Eksperyment 1			
grupa P/sylaba 5			
	KKK	KK	K
I _I	0	1	99
I _{II}	0	0	100
I _{IV}	0	0	100

Niezależność od warunku (3) jest tutaj oczywista. W kilku przypadkach sylab 3 i 5 tablica kontyngencji przedstawiała się jak w następującym przykładzie:

Eksperyment 2			
grupa P/sylaba 5			
	KKK	KK	K
I _I	0	0	100
I _{II}	0	0	100
I _{IV}	0	0	100

5.3. Wyniki doświadczeń odsłuchowych

Eksperyment 1 - rys. 3

Na rys. 3 oś pozioma na każdym wykresie odnosi się do kolejnych sylab. Na osi pionowej odcięto w procentach udział poszczególnych odpowiedzi (ocen) na ogólną liczbę 80. Linia grubą połączono oceny "akcent główny", cienką linią ciągłą oceny "akcent poboczny", a linią przerywaną - "brak akcentu". Trzy

górne wykresy odnoszą się do grupy G. Kolejne wykresy w obu rzędach dotyczą bodźców I_I , I_{II} , I_{IV} . Przesunięcie akcentu głównego z sylaby pierwszej (bodziec I_I) na sylabę drugą (bodziec I_{II}) oraz na czwartą (bodziec I_{IV}) jest widoczne bardzo wyraźnie. Wyniki są podobne dla obu grup słuchaczy.

Eksperyment 2 - rys. 4

Pod względem osi współrzędnych oraz stosowanych linii łączących rys. 4 i dalsze są analogiczne do rys. 3. W obu rzędach (odpowiednio grupa G i P) kolejne wykresy odnoszą się do wypowiedzi I_S , II_S i IV_S . Wyniki są różne dla każdego z trzech wariantów warunku (3). Zachodzi natomiast pewne podobieństwo między grupami G i P, mniejsze niż w eksperymencie 1, ale lokalizacja akcentu jest w tym eksperymencie w sposób wyraźny niejednoznaczna. W przypadku bodźca I_S maksimum dla akcentu głównego przypada na sylabach 1 i 4, dla bodźca II na sylabach 2 i 4. Bodziec IV daje znaczne maksimum na sylabie 4 oraz słabe na sylabie 2. Tak więc zastąpienie przebiegu F_0 przez szum (por. str. 6) w znacznym stopniu ogranicza możliwość lokalizacji akcentu. W odniesieniu do akcentu pobocznego grupy G i P różnią się między sobą tak w eksperymencie 1, jak i 2. Również w obu eksperymentach minima dla akcentu głównego (a zarazem maksima dla braku akcentu) przypadają na sylaby 3 i 5.

Eksperyment 3 - rys. 5

Rys. 5.1 odnosi się do bodźców I_I , I_{II} , I_{IV} , rys. 5.2 do bodźców II_I , II_{II} , II_{IV} a 5.3 do bodźców IV_I , IV_{II} , IV_{IV} (gdzie, jak poprzednio, indeks oznacza wypowiedź pierwotną, z której przeniesiono przebieg F_0). Wszystkie trzy części rys. 5 ukazują dobitnie, że lokalizacja akcentu jest tutaj wyłącznie zależna od przebiegu F_0 , bez względu na to, czy resyntezywane były oryginalne wypowiedzi (I, II, IV), czy też wypowiedzi z transplantowanymi przebiegami F_0 .

Eksperyment 4 - rys. 6

Na rys. 6 kolejne wykresy w poszczególnych rzędach odnoszą się do wypowiedzi skrajnych (np. I_I , I_{IV} , na pozycji pierwszej i piątej) oraz wypowiedzi z kolejnymi przebiegami pośred-

nimi. Grupy G i P różnią się nieco między sobą pod względem lokalizacji akcentu pobocznego, ale dają zbliżony obraz w odniesieniu do akcentu głównego, a mianowicie:

Wersja $(I_I; I_{II})_2$ (Rys. 6.1) nie różni się w sposób wyraźny od wersji I_I podczas gdy w wersji $(I_I; I_{II})_3$ następuje drastyczny spadek lokalizacji akcentu głównego na pierwszej sylabie z równoczesnym słabym wzrostem ocen lokalizujących akcent główny na sylabie 2. Dla $(I_I; I_{II})_4$ następuje dalszy spadek wskazań "akcent główny" dla sylaby 1 na korzyść sylaby 2, aby w wersji I_{II} dać obraz całkowicie jednoznaczny (por. eksperyment 1 i 3): akcent główny przesunął się z sylaby 1 na sylabę 2.

Rysunki 6.2 i 6.3 odnoszą się do układu bodźców analogicznego do przedstawionego na rys. 6.1. W miarę przechodzenia przebiegu F_0 od postaci I_I do I_{IV} poprzez przebiegi pośrednie udział oznaczeń akcentu głównego na sylabie 1 stopniowo i konsekwentnie maleje, a zarazem udział oznaczeń tego akcentu na sylabie 4 równie konsekwentnie rośnie.

W przypadku stopniowego przechodzenia od wersji I_{II} do I_{IV} zilustrowanego na rys. 6.3 lokalizacja akcentu wykazuje analogiczne zależności jak w poprzednich zestawach w tym eksperymencie: akcent poboczny ma słaby udział, a główny znika stopniowo z sylaby 2 pojawiając się coraz częściej na sylabie 4.

Eksperyment 5 - rys. 7

Porównanie rys. 7.1 i 7.2 z rys. 4 (eksperyment 2) świadczy o tym, że manewrowanie poziomem szumu zastępującego parametr F_0 nie ma istotnego wpływu na lokalizację akcentu.

Eksperymenty 1...5 przeprowadzono ze słuchaczami narodowości niemieckiej i polskiej (por. wyż., str. 8). Jest rzeczą znaną, że wyniki dla obu grup: G i P były, w każdym razie dla akcentu głównego i braku akcentu bardzo zbliżone. Świadczy to o tym, że w obu językach dominująca rolę w percepcji akcentu odgrywa parametr F_0 .

Eksperyment 6 - rys. 8

Dodatkowo przeprowadzono eksperyment 6 (ze względów technicznych) tylko ze słuchaczami polskimi. Na rys. 8, na którym kolejne wykresy odnoszą się odpowiednio do wypowiedzi I_{const} ,

II_{const} i IV_{const} widać, że zastąpienie zmiennego przebiegu F₀ wartością stałą niemal całkowicie uniemożliwia lokalizację akcentu, który dla wszystkich trzech wersji (I, II, IV) w ponad 70% percypowany był na sylabie 4.

Doświadczenie, w postaci opisanych wyżej sześciu eksperymentów, niezależnie od jego implikacji teoretycznolingwistycznych, ma istotne znaczenie dla elektronicznej syntezy mowy polskiej. Wykazało ono mianowicie, że akcentuację polskiej wypowiedzi uzyskuje się poprzez odpowiednią manipulację parametrem F₀.

DODATEK A

Wyniki doświadczeń odsłuchowych rozpatrzono najpierw uwzględniając wyniki poszczególnych osób w obrębie grupy. Dla oceny homogeniczności ocen wykorzystano eksperyment 1. Tablica A podsumowuje wyniki eksperymentu 1 uzyskane z grupą G. W każdym okienku lewa górna liczba odnosi się do akcentu głównego, liczba środkowa do akcentu pobocznego, a liczba w prawym dolnym rogu - do braku akcentu. Nagłówki kolumn oznaczają poszczególnych słuchaczy. Rzędy klatek odnoszą się do sylab w każdym z trzech bodźców. Numer kolejny sylaby widnieje na prawym marginesie, a rodzaj wypowiedzi (wariant warunku (3)) na lewym. W każdym okienku trzy liczby sumują się do 10, tj. liczby replikacji określonego bodźca - zawsze takiej samej. Tablica służy do oceny systematyczności (konsekwencji) słuchaczy w ocenie bodźców. Przypominamy, że w obrębie każdego eksperymentu replikacje były randomizowane. Zatem liczba zbliżona do 10 w określonej klatce oznacza, że dany słuchacz, mimo randomizacji, tak samo - lub prawie tak samo - oceniał dany bodziec w jego 10 replikacjach. To samo dotyczy liczby takich samych ocen zbliżonej do 0.

Zastosowanie w sposób racjonalny jakiegoś testu statystycznego do takiego rozkładu zmiennej jak w Tablicy A byłoby, z uwagi na bardzo częste zera i ogólnie niskie wartości, wysoce problematyczne, a w najlepszym razie niezwykle pracochłonne i mało precyzyjne. Postanowiliśmy zatem dokonać oceny tych wyników w sposób heurystyczny. Jako kryterium pierwsze przyjęliśmy liczby odnoszące się do akcentu głównego (górną lewą w każdym okienku) zakładając arbitralnie, że jeżeli ta liczba wynosi 8, 9 lub 10 albo 0, 1 lub 2, to słuchacze reagują konsekwentnie.

Z Tablicy A widać, że nasze kryterium zawodzi w 10 przypadkach na 120, z czego 5 przypada na słuchaczkę SK. W tym jedna klatka tej słuchaczki stanowi osobliwość: sylabę 1 w bodźcu I oznaczyła ona 8 razy jako nieakcentowaną, podczas gdy wszyscy inni słuchacze oznaczyli ją jako posiadającą główny akcent (z tego 6 słuchaczy wszystkie replikacje, a jedna słuchaczka 8 replikacji). Taka osobliwość stanowi wyjątek i została również uznana za błąd na równi z reakcjami niesystematycznymi (liczba

ocen 3...7). Tak więc osoba SK reagowała inaczej niż pozostałe osoby w tej grupie w 30% przypadków. Jako drugie, również arbitralne, kryterium homogeniczności przyjęliśmy 70% zgodności - takiej jak w przypadku słuchaczki SK - jako dolną granicę i na tej zasadzie uznaliśmy całą grupę G jako homogeniczną w sensie reakcji na bodźce w eksperymencie 1.

Ten sam, nieformalny test zastosowano do grupy II i stwierdzono, że na podstawie opisanego wyżej kryterium tylko 3 osoby zachowują się konsekwentnie, wobec czego pozostałych 6 wyeliminowano z dalszych obliczeń, jako nie należących do jednorodnej grupy.

W grupie III składającej się z 10 osób w analogiczny sposób wyeliminowano 3. Następnie połączono 3 osoby z grupy II z 7 osobami z grupy III otrzymując jednorodną grupę 10 osób reagujących konsekwentnie. Tą nową grupę oznaczono przez P. Podobnie jak w przypadku słuchaczy niemieckich (grupa I, następnie oznaczona przez G), test jednorodności oparto na wynikach eksperymentu 1.

Ponieważ żaden ze studentów polskich (grupa II) nie miał wykształcenia fonetycznego, ale z drugiej strony wszyscy słuchacze w grupie III mieli albo formalne wykształcenie fonetyczne, albo zawodowy kontakt z doświadczeniami fonetycznymi, nasze wyniki nie pozwalają na wysunięcie na tym etapie wniosków uogólniających co do wpływu treningu fonetycznego na wyniki badań w zakresie percepcji mowy. Porównanie wyników uzyskanych dla poszczególnych osób w całości doświadczenia sugeruje jednak, że istotny wpływ mogą mieć takie czynniki jak motywacja, skupienie uwagi oraz inne uwarunkowania psychiczne, których dotychczas w badaniach nad percepcją mowy nie uwzględniano.

DODATEK B

Procedura, z której korzystaliśmy w niniejszej pracy celem przeprowadzenia statystycznej analizy tablic kontyngencji, nie nakłada żadnych ograniczeń na liczebność w poszczególnych klatkach poza oczywistym przypadkiem takim, jak cytowany z eksperymentu 2 (P/sylaba 5) na str. 13.

Symbolicznie tablicę kontyngencji można przedstawić następująco:

x_{11}	x_{ij}	x_{1m}	z_1
.					
x_{i1}	x_{ij}	x_{im}	z_n
x_{n1}	x_{nj}	x_{nm}	z_n
y_1				y_m	N

W powyższej tablicy x_{ij} oznacza liczebność obserwacji w i -tym rzędzie i j -tej kolumnie, tj. liczebność podklasy (i, j) w podwójnej klasyfikacji z wartością z_i jako sumą brzegową dla i -tego rzędu oraz y_j jako sumą brzegową dla j -tej kolumny. Prawdopodobieństwo podklasy (i, j) jest wyznaczone przez odpowiednie sumy brzegowe oraz N .

Hipotezę zerową o niezależności cech można zapisać, jak następuje:

$$H_0 := p_i p_j \quad \text{dla } i=1,2,\dots,n \quad \text{oraz } j=1,2,\dots,m.$$

Klasyczny test tej hipotezy opiera się na statystyce

$$G = N \sum_{i=1}^n \sum_{j=1}^m x_{ij}^2 / (y_j z_i) - 1, \quad ,$$

gdzie

$$N = \sum_{i=1}^n \sum_{j=1}^m x_{ij}, \quad y_j = \sum_{i=1}^n x_{ij}, \quad z_i = \sum_{j=1}^m x_{ij}.$$

Jeżeli hipoteza zerowa jest prawdziwa, to statystyka G ma w przybliżeniu rozkład χ^2 z $\nu = (n-1)(m-1)$ stopniami swobody.

Jeżeli liczebności w poszczególnych klatkach są małe (nie spełniają kryteriów wymienionych na str. () , to test zostaje zmodyfikowany. Statystykę G zastępuje się statystyką G_1 :

$$G_1 = cG, \quad \text{gdzie}$$

$$c = 2E(G)/\text{Var}(G),$$

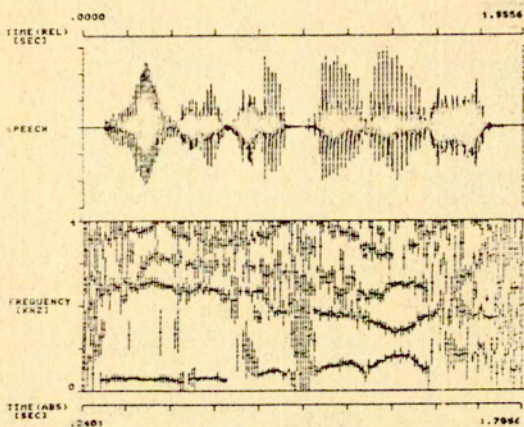
gdzie $E(G)$ oznacza wartość oczekiwaną, a $\text{Var}(G)$ wariancję statystyki G . Obliczenie wartości statystyki G_1 jest dość złożone. Odpowiednie metody numeryczne przedstawiły Dobek i Kiełczewska (1977). Ich procedura oblicza wartość zmodyfikowanego testu χ^2 oraz zmodyfikowanej liczby stopni swobody:

$$\nu_1 = c(m-1)(n-1)/(N-1).$$

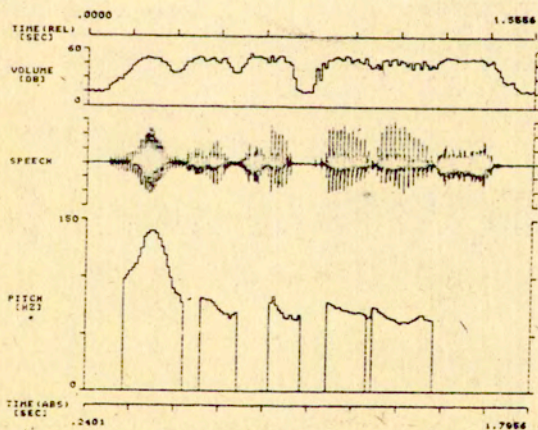
BIBLIOGRAFIA

- Barry, W.J., van Dommelen, W., Jan en, H., Kohler, K.J., Schäfer, K., Thon, W., Timmermann, G., 1982, Phonetic Data Processing at Kiel University, Arbeitsberichte Nr. 18, Inst. f. Phonetik.
- Dobek, A., Kiełczewska H., 1979, Testowanie niezależności i jednorodności prób dla małych liczebności oczekiwanych, Algorytmy biometryczne i statystyczne No. 6. 163-174.
- Dobrogowska, K., 1978, Badanie zależności między akcentem a intonacją w mowie nieemocjonalnej, *Lingua Posnaniensis* vol. 21, 65-75.
- Greń, J., 1974, Statystyka matematyczna, modele i zadania, wyd. 4, PWN, Warszawa.
- Jassem, W., 1954, The Phonology of Polish Stress, *Word* vol. 15, 252-269.
- Jassem, W., 1962, Akcent języka polskiego, *Prace Językoznawcze* No. 31, Komitet Językoznawstwa PAN.
- Jassem, W., Gibbon, D., 1980, Re-Defining English Accent and Stress, *Journ. Intern. Phon. Ass.* vol. 10, 2-16.
- Jassem, W., Morton, J., Steffen-Batogowa, M., 1968, The Perception of Stress in Speech-Like Stimuli by Polish Listeners, *Speech Analysis and Synthesis* vol. 1 (W. Jassem, ed.) PWN, Warszawa, 289-308.
- Ladefoged, P., 1975, *A Course in Phonetics*, Harcourt, Brace, Ivanovich Inc., New York.
- Lancaster, O., 1969, *The Chi-Squared Distribution*, Wiley and Sons, New York.
- Nass, C.A.G., 1959, The χ^2 Test for Small Expectations in Contingency Tables, with Special Reference to Accident and Absenteeism, *Biometrika* vol. 46, 365-385.
- Pothhoff, F., Whittinghall, M., 1966, Testing for Homogeneity. The Binomial and Multinomial, *Biometrika* vol. 53, 167-182.
- Rathjen, G., 1979, Numerische und experimentelle Methoden der linearen Prädiktion zur Datenreduktion von Sprachsignalen, *Arbeitsberichte* Nr. 11, Inst. f. Phonetik, Kiel, 3-133.

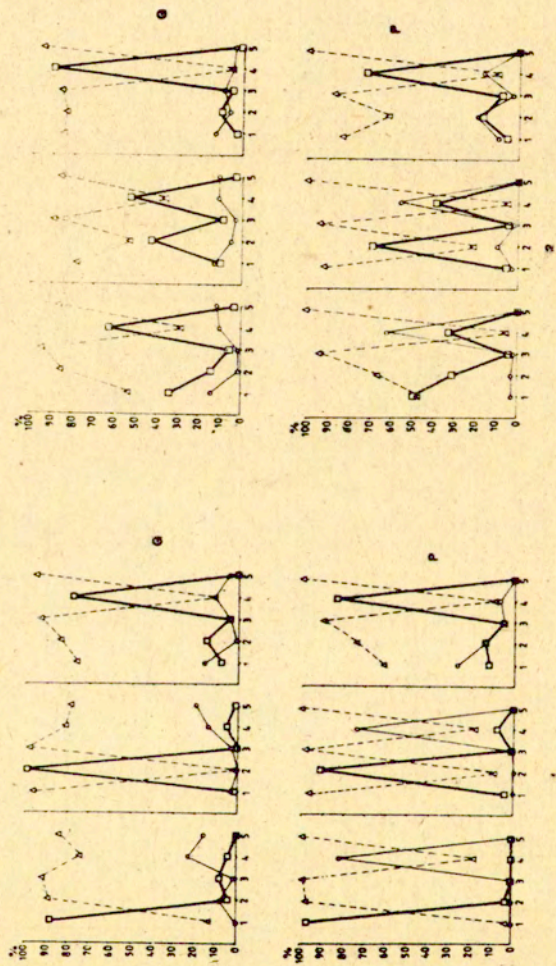
- Schäfer, K., 1982, Concepts in the SSP Programme, Arbeitsberichte Nr. 18, Inst. f. Phonetik, Kiel, 111-126.
- Volk, W., 1973, Statystyka matematyczna dla inżynierów, Wyd. Nauk.-Techn., Warszawa.
- Wierzchowska, B., 1965, Wymowa polska, Państw. Zakł. Wyd. Szkol., Warszawa.
- Wierzchowska, B., 1967, Opis fonetyczny języka polskiego, PWN, Warszawa.
- Wierzchowska, B., 1980, Fonetyka i fonologia języka polskiego, Zakł. Nar. im. Ossolińskich, Wrocław.



Rys. 1. Analiza LPC zdania I: "Widzi cie teraz". Oscylogram oraz parametry widmowe.



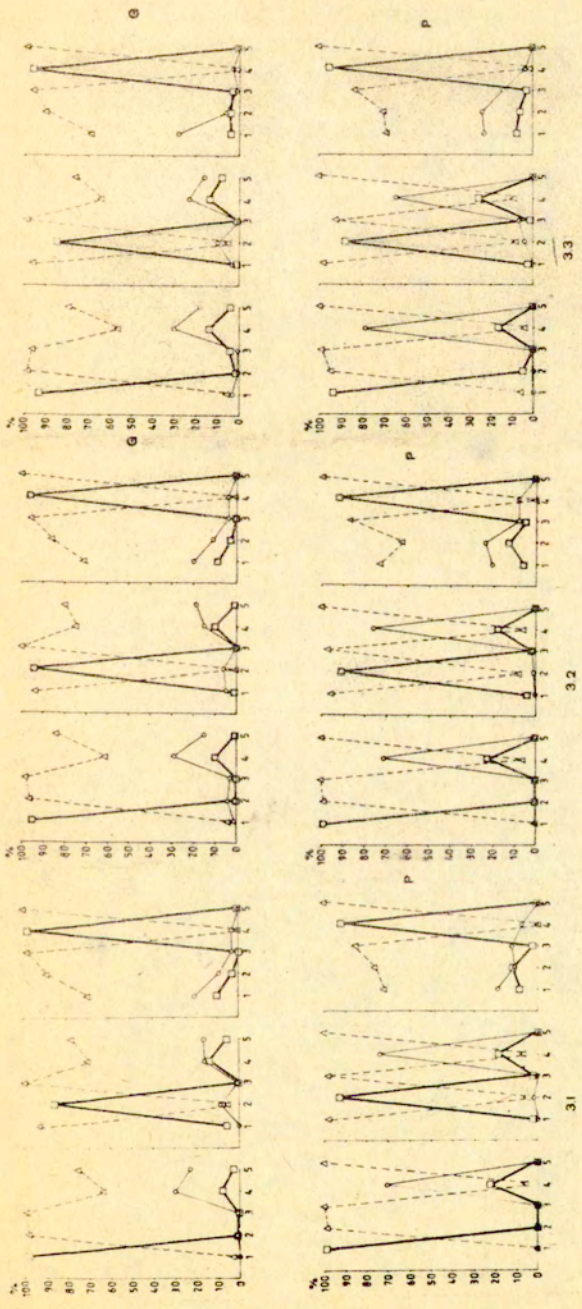
Rys. 2. Analiza LPC zdanie I: "Widzi cie teraz". Poziom syg-
nal, oscylogram oraz przebieg parametru F_0 .



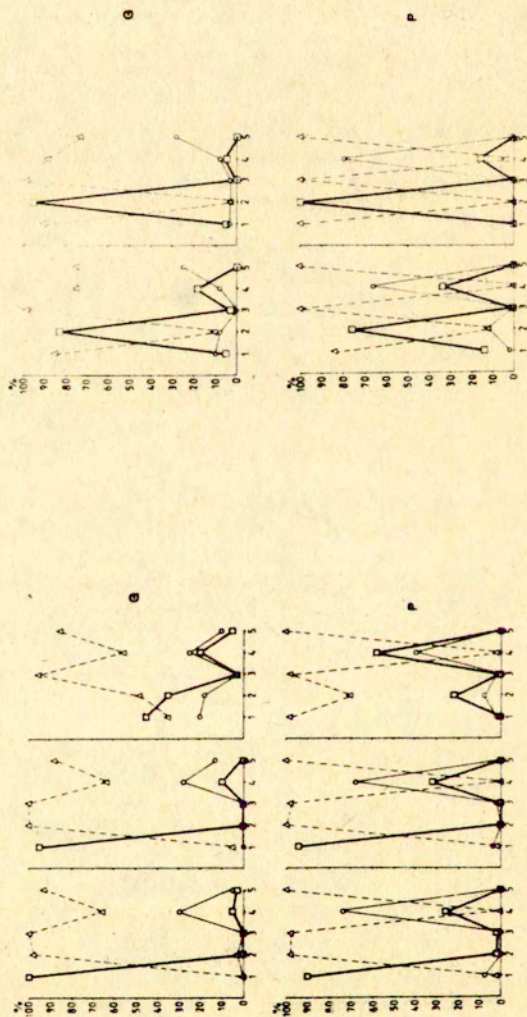
Rys. 3. Wyniki eksperymentu 1.

Rys. 4. Wyniki eksperymentu 2.

G - słuchacze niemieccy,
 P - słuchacze polscy.
 Na osi poziomej kolejne sylaby.
 Linia gruba - akcent główny,
 linia cienka - akcent poboczny,
 linia przerywana - brak akcentu.



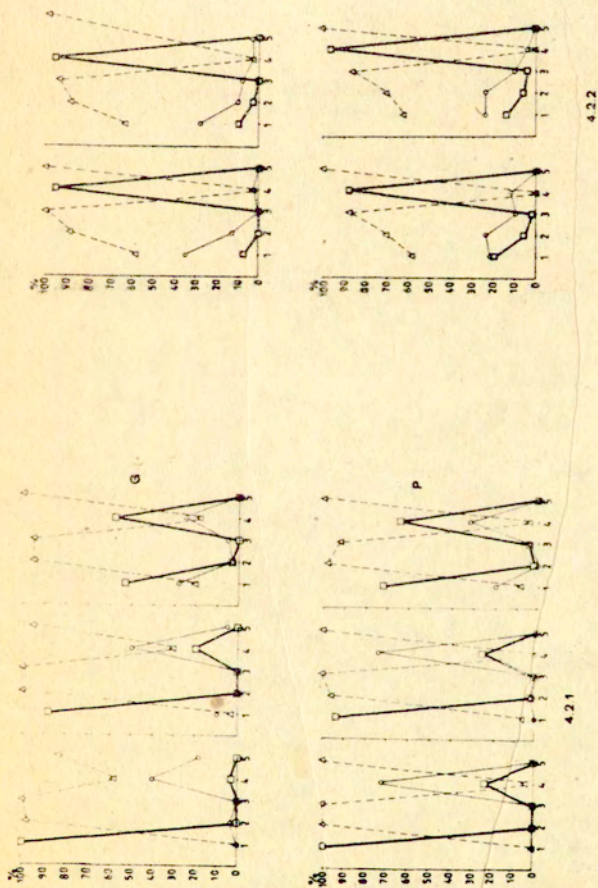
Rys. 5. Wyniki eksperymentu 3.



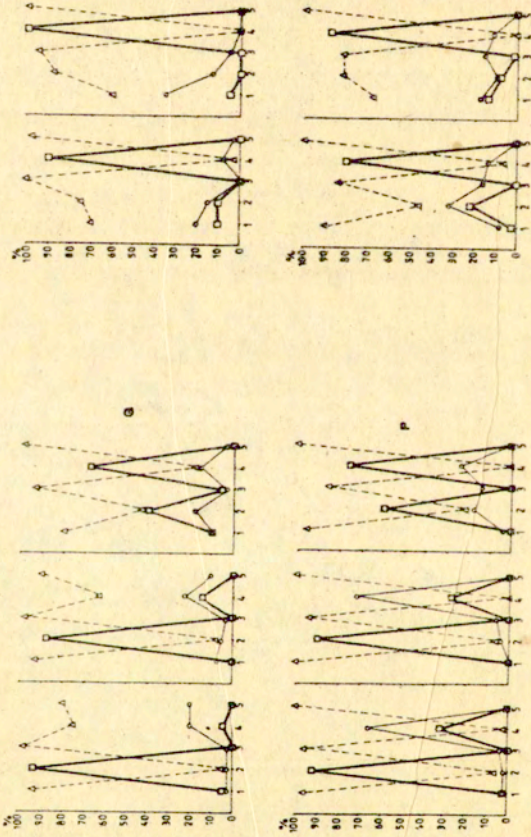
412

411

Rys. 6. Wyniki eksperymentu 4. Część I.



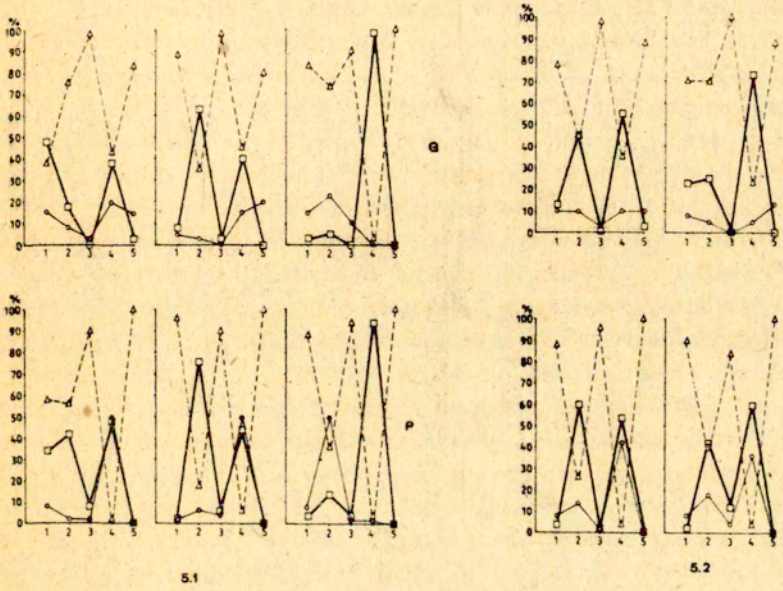
Rys. 6. Wyniki eksperymentu 4. Część II.



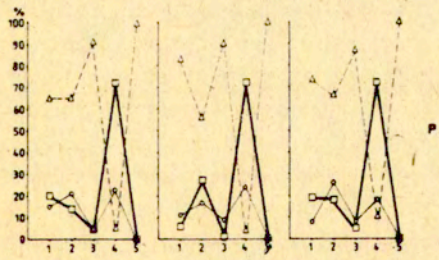
4.32

4.31

Rys. 6. Wyniki eksperymentu 4. Część III.



Rys. 7. Wyniki eksperymentu 5.



Rys. 8. Wyniki eksperymentu 6

	BK	CR	EB	FR	HW	MT	RD	SK	
A ₁	1
A ₂	
A ₃	
A ₄	2
A ₅	
A ₆	
A ₇	3
A ₈	
A ₉	
A ₁₀	4
A ₁₁	
A ₁₂	
A ₁₃	5
A ₁₄	
A ₁₅	

Tablica zgodności ocen.



56744