212422 c/1

Na prawach rękopisu

INSTYTUT TELEKOMUNIKACJI I AKUSTYKI
POLITECHNIKI WROCŁAWSKIEJ

Dział A

Akustyka

Raport nr 128/ I-28/PRE-080/79

REGUŁY GENERACJI POBUDZENIA KRTANIOWEGO W PROCESIE SYNTEZY FRAZ MOWY POLSKIEJ

Wojciech Myślecki

PRACA DOKTORSKA

Promotor: doc.dr inż. Janusz Zalewski

Słowa kluczowe: synteza mowy,
pobudzenie krtaniowe

Wrocław 1979

79 1009/1017

mgr inż. Wojciech Myślecki

Instytut Telekomunikacji i Akustyki
Politechniki Wrocławskiej

Wrocław, ul.Wybrzeże Wyspiańskiego 27

Raport wpłynął do redakcji 14.07.79.

Promotorowi
doc. dr inż. Januszowi Zalewskiemu
za pomoc i współpracę w realizacji
badań składam niniejszym
podziękowania

Wojciech Myślecki

SPIS TREŚCI

1. WSTĘP

1.1. Wprowadzenie

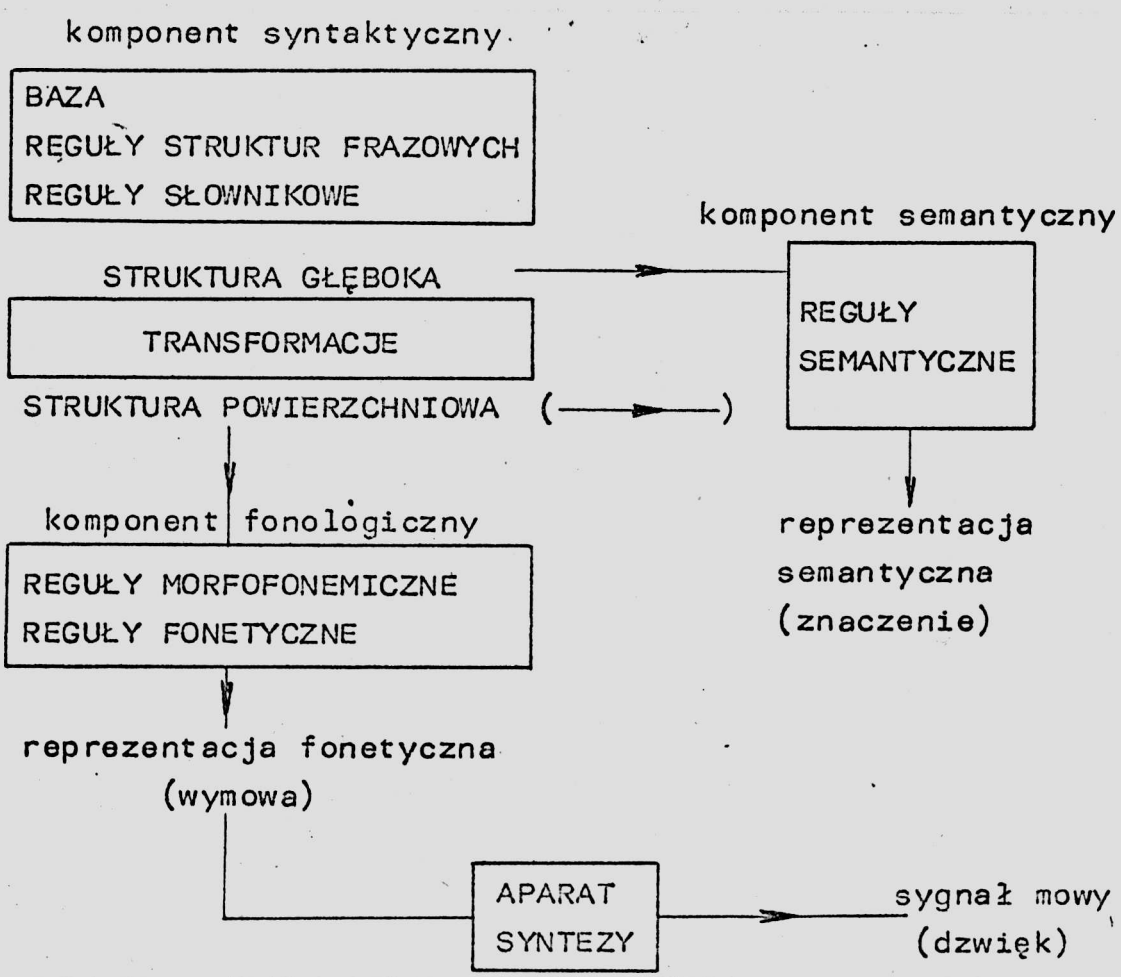
Mowa, według definicji de Saussure [1], stanowi realizację formy planu wyrażenia w substancji fonetycznej (dźwiękowej). Pod pojęciem formy Saussure rozumiał abstrakcyjną strukturę relacji, którą dany język z jednej strony odwzorowuje w substancji materialnej (substancji planu wyrażenia), a z drugiej narzuca substancji znaczeniowej (substancji planu treści), pojmowanej jako całokształt myśli, uczuć i idei wspólnych dla całej ludzkości, niezależnie od języka, jakim ludzie się posługują.

Chomsky tworząc strukturalną koncepcję opisu języka (gramatykę generatywną) [3, 4] przyjął saussurowską zasadę rozdziału formy i substancji, rezygnując jednak z jej uniwersalności na rzecz funkcjonalności, wprowadził nieco odmienną opozycję między uniwersaliami formalnymi (formą), a uniwersaliami substantywnymi (substancją). Uniwersalia formalne są warunkami funkcjonowania reguł opisu języka, natomiast uniwersalia substantywne stanowią elementy językowe[1)], których dotyczą te reguły [2]. Wprowadzenie powyższego rozróżnienia pozwoliło Chomskyému opracować sformalizowany opis gramatyki generatywnej, której poszczególne części składowe pokazano na rys.1.1.

Szczegółowy opis gramatyki generatywnej podano w szeregu pracach zarówno źródłowych [3, 4], jak i podręcznikowych [5, 6, 7], zatem w niniejszej pracy zrezygnowano z jego zamieszczenia. Celowe jest jednak podkreślenie dwóch istotnych aspektów zarówno założeń teoretycznych jak i modelu gramatyki Chomskyégo:

---

1) Pod pojęciem "elementy językowe" należy tu rozumieć wszystkie dające się wyodrębnić jednostki na różnych poziomach analizy języka - np. zdania, frazy, wyrazy, morfemy, fonemy, cechy dystynktywne itp.



Rys.1.1. Schemat blokowy gramatyki Chomsky'ego

1. rezygnacja z saussurowskiej nadrzędności formy nad substancją planu treści i wprowadzenie w jej miejsce relacji interpretowalności formy składniowej (zdań) przez komponent semantyczny. Interpretacja zachodzi w dziedzinie języka[2)] i jej wynikiem jest znaczenie (reprezentacja semantyczna) interpretowanej formy. To pozwoliło na ograniczenie pojęcia substancji wyłącznie do elementów języka (uniwersalia substantywne),
2. wprowadzenie, niespotykanego we wszystkich poprzednich teoriach językoznawczych odwzorowania formy języka w substan-

---

2) Dziedzina języka - zbiór informacji, którego elementy stanowią treść wyrażeń poprawnych danego języka [3].

cji dźwiękowej za pomocą aparatu syntezy, na wyjściu którego otrzymuje się sygnał mowy. Chomsky chciał tym podkreślić funkcjonalny charakter gramatyk generatywnych, które stanowią metodę wytwarzania poprawnych struktur składniowych i fonologicznych danego języka (a więc syntezy tych struktur), natomiast nie stanowią ścisłego modelu tworzenia tych struktur przez człowieka. W konsekwencji uznał, że ogniwem pomiędzy ciągiem abstrakcyjnych, dyskretnych symboli terminalnych stanowiących wyjście z komponentu fonologicznego, a dźwiękową postacią struktur językowych, opisanych tym ciągiem (akustyczny sygnał mowy), nie może się znajdować abstrakcyjna relacja odwzorowania formy w substancji, lecz urządzenie fizyczne (aparat syntezy sygnału mowy).

Model Chomsky'ego generacji struktur języka stanowił doniosły krok do przodu w rozwoju badań nad systemami syntezy mowy. Pozwolił na integrację rozproszonych do tego momentu cząstkowych dokonań w tej dziedzinie, ukazał zależności, związki i hierarchię pomiędzy różnymi płaszczyznami języka, usystematyzował i ustalił zakres pojęć i definicji stosowanych dotychczas niezbyt konsekwentnie przez przedstawicieli różnych dyscyplin zajmujących się omawianym problemem (lingwiści, fonetycy, akustycy, cybernetycy). Kompleksowość modelu Chomsky'ego ukazała wyraźnie kierunki dalszych szczegółowych badań nad syntezą mowy (ściślejszym określeniem byłoby "nad syntezą języka mówionego"), równocześnie stanowiąc dla nich czynnik integrujący oraz wskazujący ich umiejscowienie w hierarchicznym systemie syntezy mowy.

1.2. Stan badań nad syntezą mowy

Prowadzone w ostatnich latach w czołowych ośrodkach naukowych badania nad syntezą mowy[3] dotyczyły kompleksowych, hierarchicznych systemów opartych o przedstawiony w rozdz.1.1. model Chomsky'ego. Na podstawie rozwiązań przedstawionych w pracach [8-14] możliwe jest podanie uogólnionego systemu syn-

[3] Pod pojęciem syntezy mowy autor rozumie tutaj realizację formy języka w substancji fonetycznej, tzn. zarówno syntezę

tezy mowy, uwzględniającego wszystkie poziomy syntezy oraz alternatywy wejściowego ciągu sterującego systemem. Schemat blokowy takiego systemu pokazano na rys.1.2.

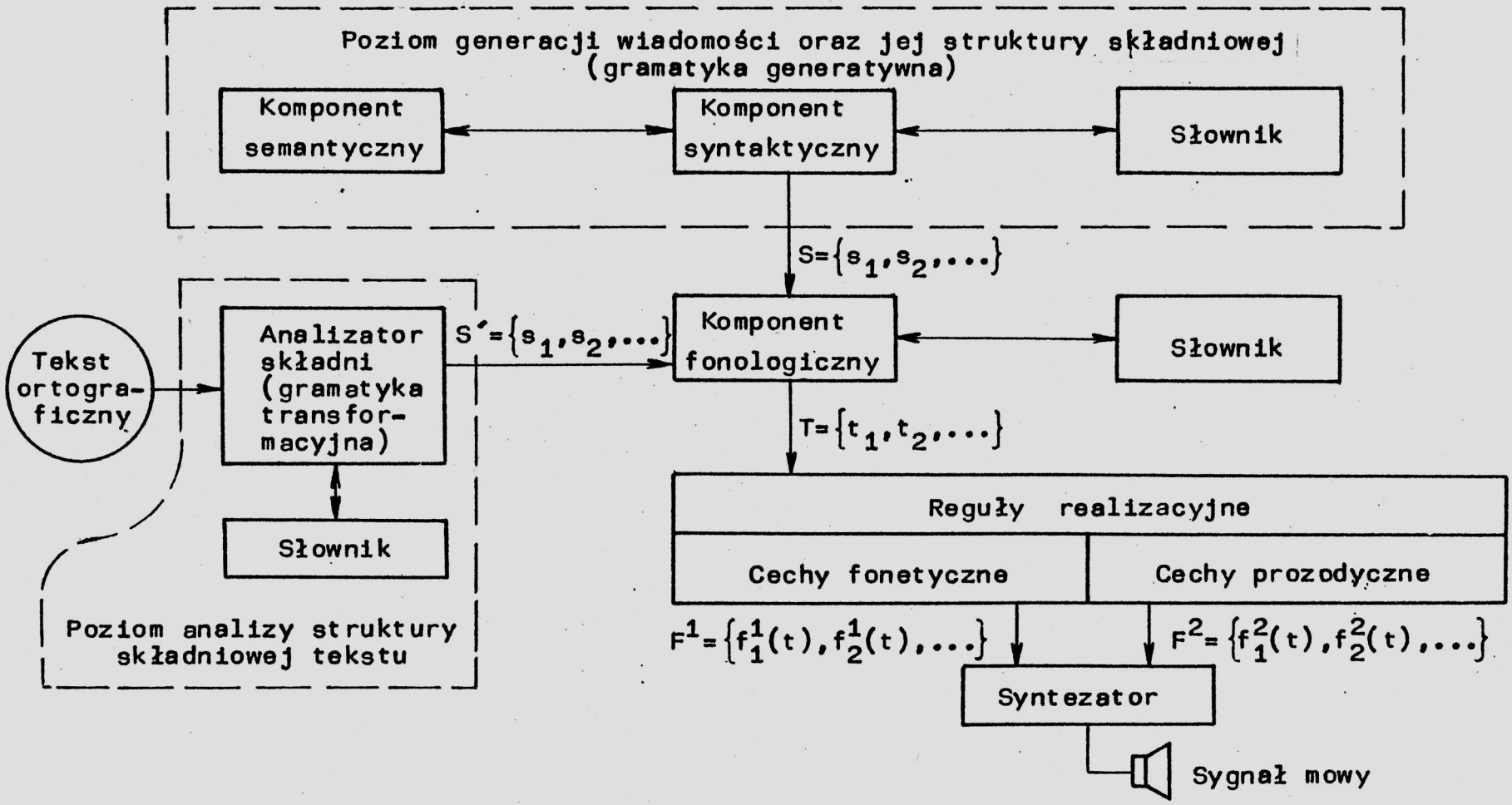
Wyjściowy sygnał mowy jest generowany z ciągu S lub S' abstrakcyjnych symboli reprezentujących zdania otrzymane bądź z poziomu generacji wiadomości (S), bądź w wyniku analizy syntaktycznej tekstu ortograficznego (S'). Ciągi S i S' zawierają informacje o strukturze powierzchniowej zdania, obejmujące kategorię zdania (oznajmiające, pytające itp.), umiejscowienie akcentu zdaniowego i wyrazowego oraz umiejscowienie granic fraz i zdań. Dodatkowo w tym ciągu mogą być zawarte informacje o uwydatnieniu pewnych fraz lub zdań (emfaza), tempie wymowy, elementach ekspresyjnych i innych zmiennych psychologicznych i semantycznych mających wpływ na akustyczną realizację zdania [10, 11, 13, 22].

Informacje zawarte w ciągu S lub S' stanowią wejście do komponentu fonologicznego, którego zadaniem jest ich zdekodowanie i wygenerowanie na wyjściu sekwencji symboli reprezentujących definitywną formę fonetyczną jednostek tworzących zdanie oraz pełny opis kształtu prozodycznego zdania (w postaci zbioru ustalonych wartości czasów trwania poszczególnych segmentów, częstotliwości podstawowej i intensywności) [11, 14]. Te informacje zawarte w ciągu T, stanowiącym wyjście z komponentu fonologicznego (rys.1.2.), zostają przetworzone przez reguły realizacyjne na zbiory funkcji czasowych F^1 i F^2 sterujących aparatem syntezy.

Reguły realizacyjne, zwane też regułami transformacji fonetyczno-akustycznych [11], mają charakter heurystyczny i przy ich ustalaniu konieczne jest dodatkowe uwzględnienie szeregu czynników nie mających interpretacji na płaszczyźnie językowej, takich jak budowa i funkcjonowanie kanału głosowego człowieka, zjawiska aerodynamiczne towarzyszące procesowi powstawania dźwięków mowy, transformacje koartykulacyjne, przyjęty model

---

struktury jak i dźwięków mowy. Często syntezę mowy pojmuje się w znacznie węższym sensie, to znaczy jako parametryczną syntezę dźwięków mowy [15-19].



Rys. 1.2. Schemat blokowy hierarchicznego systemu syntezy mowy

syntezy dźwięków mowy [20, 22]. Powyższy fakt jest w pewnym sensie oczywisty, jeżeli zgodnie z modelem strukturalnym przyjmiemy, że reguły realizacyjne stanowią odwzorowanie formy w substancji, z czego wynika ich dualizm, polegający na uwzględnieniu z jednej strony formalnych relacji językowych, a z drugiej prawideł fizycznych związanych z materialnym charakterem substancji dźwiękowej. Następstwa tego dualizmu reguł realizacyjnych,w odniesieniu do metodologii badań nad ich ustaleniem i praktycznym wdrożeniem w procesie syntezy,omówiono w dalszej części pracy.

Przedstawiony powyżej skrótowy opis strukturalnego modelu syntezy mowy wyraźnie ukazuje wielorakość, różnorodność i złożoność zagadnień teoretycznych i praktycznych związanych ze sztucznym wytwarzaniem mowy. W konsekwencji, mimo długoletnich badań i dużego postępu wiedzy w tym zakresie do chwili obecnej nie opracowano systemu zdolnego do wytworzenia (syntezy) mowy na wszystkich rozważanych poziomach [13, 25].

Aktualny stan badań nad modelem komponentu semantycznego i jego związkami z komponentem syntaktycznym znajduje się, z punktu widzenia zastosowań do syntezy mowy, w stadium początkowym, gdyż, jak stwierdza Jassem [21], "informacja semantyczna w sensie cybernetycznym jest stosunkowo mało opracowana". Rozwiązania proponowane przez Klatta [11] oraz Fallisade'a i in. [14] mają charakter czysto pragmatyczny i cząstkowy, odbiegający całkowicie od funkcji wyznaczonych komponentowi semantycznemu w strukturalnym modelu syntezy.

Znacznie bardziej są zaawansowane prace nad zagadnieniem generacji struktur składniowych (komponent syntaktyczny) lub analizy tych struktur (analizator składni). Omawiane w literaturze najnowsze rozwiązania modelu komponentu syntaktycznego [11, 14] lub analizatora syntaktycznego [8, 9, 10, 12, 22] wykorzystują koncepcję gramatyki generatywno-transformacyjnej Chomsky'ego [3, 4], z uwzględnieniem szeregu modyfikacji i uproszczeń, i w przypadku języka angielskiego znajdują już praktyczne zastosowanie w systemach syntezy [8-14].

Podobne lub nawet większe zaawansowanie prac można stwierdzić w odniesieniu do komponentu fonologicznego. Funkcjonalne modele tego komponentu dla języka angielskiego omówiono w cytowanych uprzednio pracach [8-14] , natomiast dla języka niemieckiego w pracy Mangolda i Stalla [22]. Podobnie jak w przypadku komponentu syntaktycznego, modele komponentu fonologicznego oparto o gramatykę generatywną systemu fonologicznego danego języka [11, 12, 14, 22, 26] . Formalny zapis reguł tej gramatyki w postaci akceptowalnej przez maszynę cyfrową, podał Carlson i Granström [24] a przykład programowej realizacji reguł fonologicznych dla języka angielskiego opisano w pracy Friedmana i Morina [25] . O stopniu zaawansowania badań nad modelem komponentu fonologicznego może świadczyć późniejsza praca Friedmana [26], gdzie podano program realizacji reguł fonologicznych dla szybkiej wymowy potocznej.

Stan badań nad regułami realizacyjnymi, ze względu na ich złożoność, wymaga obszerniejszego omówienia. Umeda [13] dokonując przeglądu badań nad tymi regułami syntezy wyróżnia w nich trzy etapy. Pierwszą połowę lat siedemdziesiątych, którą Umeda kwalifikuje do drugiego etapu[4)], poświęcono opracowaniu zbioru reguł uwzględniających znacznie większą liczbę czynników wpływających na ostateczną, dźwiękową postać mowy, co miało głównie na celu uzyskanie większej naturalności mowy syntetycznej. Przykładami charakterystycznymi dla tego etapu są prace Cokera, Umedy i Browmana [9] nad w pełni zautomatyzowanym systemem syntezy z tekstu, koncepcja strukturalnego modelu syntezy według reguł podana przez Klatta [11], czy też system syntezy z tekstu opracowany przez Allena [12] , gdzie wykorzystano rozwiniętą wersję zbioru minimalnych reguł syntezy podanych przez Mattingly'ego [27].

Równolegle do prac nad kompleksowymi systemami syntezy prowadzono w tym etapie szereg badań szczegółowych nad poprawą naturalności brzmienia mowy syntetycznej. Te badania, o charakterze fonetyczno-akustycznym, dotyczyły doboru właściwych pa-

---

[4)] Omówienie pierwszego, "historycznego" etapu pominięto.

rametrów występujących w regułach realizacyjnych oraz optymalizacji parametrów syntezy dźwięków mowy[5]. Jako przykład celowe jest wymienienie szeregu prac tego typu dotyczących syntezy mowy polskiej. Badania nad doborem optymalnych częstotliwości formantowych dla samogłoskowych fonemów mowy polskiej omówiono w pracach Majewskiego i Holliena [28] oraz Kudeli [29, 30]. Kacprowski i Mikiel podali uproszczone reguły syntezy sylab typu C-V, przy czym uwzględniono tu nie tylko zagadnienie doboru parametrów formantowych dla samogłoskowych i spółgłoskowych fonemów mowy polskiej, ale również reguły sterowania czasowymi przebiegami częstotliwości podstawowej (F_0) i amplitudy źródła pobudzającego (A_0) oraz zagadnienie charakterystyk szumowego i quasiperiodycznego źródła pobudzającego [17, 31, 32]. Badania wpływu różnorodnych czynników, takich jak charakterystyka funkcji źródła pobudzającego, przebiegi F_0 i A_0, sinusoidalna modulacja F_0, szybkie zmiany częstotliwości formantowych synchroniczne z okresem F_0, na naturalność brzmienia samogłosek oraz krótkich fraz syntetycznych było przedmiotem prac Myśleckiego, Zalewskiego i Gosa [33-37].

Wspólną cechą badań drugiego etapu była stosowana technika analizy przez syntezę, przy wykorzystaniu tylko fragmentarycznych danych uzyskanych z mowy naturalnej[6] oraz subiektywna weryfikacja eksperymentów syntezy [13]. Ta metoda badawcza posiada ograniczone możliwości w polepszeniu naturalności mowy syntetycznej, gdyż, jak stwierdza Umeda [13], "... human speech is governed by a far greater number of factors than one can think of at the conscious level". W konsekwencji, mimo istot-

---

[5] Termin "parametry syntezy", użyty przez J.Kacprowskiego w cytowanej pracy [18], jest zdaniem autora trafnym określeniem czynników, które mają wpływ na jakość syntezowanej mowy, a nie są związane z regułami interpretowalnymi na płaszczyźnie językowej. Przykładem jest charakterystyka funkcji źródła pobudzającego [12].

[6] Przykładem są omówione badania J.Kacprowskiego i W.Mikiela, gdzie wykorzystywano opis głosek polskich za pomocą parametrów fonetyczno-akustycznych opracowany przez W.Jassema [21, 38].

nego, w ostatnich latach, postępu badań nad syntezą mowy, uzyskiwana jakość mowy syntetycznej, szczególnie dłuższych fraz i zdań, ciągle odbiega od standardu mowy naturalnej [8, 10, 12, 13, 36] . Rosenberg [40] oraz Myślecki i Zalewski [36] stosując cyfrowe modele kanału głosowego, uzyskali brzmienie mowy syntetycznej zbliżone w opinii słuchaczy do mowy naturalnej, zatem przyczyn nienaturalności brzmienia mowy syntetycznej przy dłuższych komunikatach należy upatrywać w niedoskonałości reguł realizacyjnych i w nie do końca rozpoznanych związkach tych reguł z wyższymi poziomami syntezy mowy [10, 13] . Stąd w etapie trzecim konieczne jest podjęcie badań nad tymi zagadnieniami z zastosowaniem metodologii opartej na szczegółowych i kompleksowych badaniach mowy naturalnej. Umeda [13] przewiduje zakończenie prac etapu trzeciego, zwanego "badaniem reguł derywowanych z mowy naturalnej", na początku lat osiemdziesiątych.

W odniesieniu do technicznych środków realizacji procesu syntezy mowy, na wszystkich omawianych poziomach, ostatnie lata stanowią zdecydowany postęp nie tylko ilościowy, ale również jakościowy. Opracowana w latach sześćdziesiątych koncepcja syntezy z końcówką analogową sterowaną z komputera (terminal-analog speech synthesizer) [32, 39-42] jest w zasadzie nadal aktualna, przy czym zmienił się zakres procedur wykonywanych przez komputer sterujący oraz zasady konstrukcji aparatu syntezy (końcówki analogowej). W prezentowanych w literaturze najnowszych rozwiązaniach systemów syntezy w czasie rzeczywistym [14-16, 43-47] implementację reguł syntezy, aż do poziomu reguł realizacyjnych włącznie, dokonuje minikomputerowy system sterujący [14, 15, 43, 44] wyposażony najczęściej w pojemną pamięć dyskową [14, 15, 43] . Funkcje czasowe wygenerowano przez system minikomputerowy sterują wyspecjalizowanymi układami hardwarowymi pracującymi w czasie ciągłym [14-16, 43, 44] lub w czasie dyskretnym (dotyczy szczególnie nowych technik cyfrowej syntezy, jak liniowe kodowanie predykcyjne (LPC) lub synteza homomorficzna) [45-48] . W latach osiemdziesiątych należy przewidywać wprowadzenie zintegrowanych syste-

mów syntezy, opartych całkowicie na wyspecjalizowanych procesorach o dużych zdolnościach obliczeniowych, specjalnym oprogramowaniu "hardware'owym" i pojemnych pamięciach półprzewodnikowych, zapewniających wielokrotną aktualizację zawartego w nich zbioru danych[7]. Przykładem wstępnego, częściowego opracowania tego typu systemu syntezy jest model Allena i Steingarta [47], gdzie wyspecjalizowany procesor typu AMD. 2901 z pamięcią RAM realizuje nie tylko syntezę sygnału mowy, ale również uproszczony zestaw reguł fonologicznych.

Podsumowując dokonany przegląd stanu badań nad syntezą mowy i przewidywanych perspektyw, można stwierdzić relatywne opóźnienie badań nad przejściem z formy języka do substancji dźwiękowej (reguły realizacyjne). Dlatego też jest celowe i niezbędne prowadzenie żmudnych i często mało spektakularnych badań nad tym problemem. Niniejszą pracę, dotyczącą optymalizacji prostych reguł generacji pobudzenia krtaniowego pod kątem uzyskania maksymalnej naturalności generowanych, syntetycznych fraz mowy polskiej, należy zaliczyć do tego kierunku badawczego, przy czym odnosząc się do podziału Umedy [13], metodologicznie i czasowo (prace realizowane w latach 1975-1978) jest ona umiejscowiona w drugim etapie badań nad syntezą mowy.

---

7) W 1978 roku firma Intel oraz Fairchild zaprezentowała 16-bitowe mikroprocesory Intel 8086 wykonane techniką HMOS. Zdaniem przedstawicieli firm zdolności operacyjne tego mikroprocesora są porównywalne z minikomputerem PDP-11/45, a jego poziom techniczny wyprzedza o 2 lata stan badań nad możliwościami jego zastosowania. Firma przewiduje wprowadzenie w roku 1980 specjalnego oprogramowania (m.in. assemblera).

1.3. Definicje[8)]

1.3.1. Definicje związane z procesem syntezy

W pracach dotyczących zagadnień przekształcania w procesie syntezy mowy ciągu abstrakcyjnych symboli otrzymywanych na wyjściu komponentu fonologicznego w zbiór funkcji czasowych stanowiących dziedzinę reguł realizacyjnych i wytwarzania w oparciu o te funkcje akustycznego sygnału mowy [8-14], istnieje rozbieżność w stosowanym nazewnictwie i określeniach. Celowe jest zatem podanie znaczenia następujących pojęć stosowanych w dalszych rozdziałach pracy:

Def.1. Synteza mowy : Wynik zastosowania hierarchicznie uporządkowanych gramatyk generujących formę składniową i fonologiczną komunikatów oraz przekształcających abstrakcyjny zapis formy w akustyczny sygnał mowy syntetycznej.

Def.2. Synteza dźwięków mowy : Zastosowanie gramatyki reguł realizacyjnych G_R z aksjomatem tożsamościowo-równym ciągowi T symboli terminalnych komponentu fonologicznego (rys.1.2.) i wykorzystanie produktu tego zastosowania, w postaci zbioru funkcji czasowych $\{ F^i \} \in DF$, gdzie DF dziedzina G_R, do sterowania urządzeniem syntezy dźwięków mowy (modelem kanału głosowego).

Def.3. Cechy prozodyczne mowy : elementy dziedziny funkcji prozodycznych [13] jak akcent, intonacja, cechy graniczne. Cechy prozodyczne należą do formy języka.

Def.4. Akustyczne składniki cech prozodycznych : subiektywnie odczuwane wielkości fizyczne sygnału mowy, lub czasowe zmiany tych wielkości, będące produktem odwzorowania cech prozodycznych (formy) w substancji dźwiękowej, np. poziom intensywności, wysokość tonu (ang. pitch), allo-

---

8) Opis symboli stosowanych w niniejszym rozdziale podano w "Wykazie ważniejszych oznaczeń i symboli" znajdującym się na w rozdziale 1.7.

foniczne transformacje dźwiękowej struktury fonemów w nagłosie lub wygłosie [11-13, 50, 51, 53, 56].

Def.5. Parametry realizacyjnych reguł syntezy w zbiorze argumentów: zbiór elementów alfabetu abstrakcyjnych symboli terminalnych komponentu fonologicznego po dokonaniu podstawienia:

$$T = V \qquad (1.1.)$$

gdzie: T - ciąg wyjściowy z komponentu fonologicznego
V - aksjomat gramatyki G_R należący do zbioru jej argumentów

Def.6. Parametry dziedziny realizacyjnych reguł syntezy: Zbiór liczbowych wartości stosowanych w regułach realizacyjnych przy generowaniu zbiorów funkcji $\{F^i\} \in DF$, np. czas narastania częstotliwości podstawowej na początku frazy, maksymalna wartość podbicia F_o w sylabie akcentowanej, ustalone wartości częstotliwości formantów, względne wydłużenie samogłoski akcentowanej itp.

Def.7. Parametry syntezy dźwięków mowy: czynniki wpływające na strukturę syntezowanego sygnału mowy, a nie mające formalnych związków z regułami syntezy. Te czynniki, jak typ i kształt funkcji źródła pobudzającego, czy charakterystyka częstotliwościowa korektora wyższych formantów, zależą najczęściej od przyjętego modelu kanału głosowego (por. odnośnik 5).

Podział na parametry dziedziny reguł realizacyjnych oraz na parametry syntezy nie jest jednoznaczny. Przykładowo można tu wymienić trudności w zakwalifikowaniu funkcji interpolacyjnej realizującej przejście między dwoma, ustalonymi wartościami częstotliwości formantowych. Przyjmując w systemie rozwiązanie Rabinera [57], gdzie stan przejściowy jest odpowiedzią układu liniowego drugiego rzędu na skokowe pobu-

dzenie generowane z poziomu reguł realizacyjnych, wartości pary biegunów transmitancji układu, które decydują o kształcie funkcji przejścia, należy zakwalifikować do zbioru parametrów syntezy dźwięków mowy. Z kolei w rozwiązaniu podanym przez Gosa i in. [37] reguły realizacyjne generują pełny przebieg czasowych zmian formantów i stosowany typ funkcji przejścia (funkcja trygonometryczna) jest parametrem dziedziny reguł realizacyjnych.

1.3.2. Definicje związane z jakościowymi aspektami mowy

Podstawowym kryterium oceny mowy syntetycznej jest zrozumiałość, oceniona najczęściej w oparciu o liczbowe miary jak wyrazistość głoskowa lub sylabowa [31, 58] . To kryterium oraz związane z nim miary są powszechnie stosowane również w ocenie mowy naturalnej i w pracach dotyczących syntezy mowy nie występują różnice w ich interpretacji.

Istnieje duża zgodność w opinii, że współczesne systemy syntezy zapewniają dobrą zrozumiałość lecz niewystarczającą naturalność [8, 10, 12, 16, 34] . To ostatnie pojęcie, stosowane często wymiennie z pojęciem jakości, jest interpretowane w dość dowolny sposób i w znanej autorowi literaturze nie podano jego ścisłego znaczenia. Przykładem rozbieżności jest porównanie interpretacji pojęcia naturalność przez Allena [8], który wiąże je ze wszystkimi poziomami syntezy, ze stanowiskiem zawartym w pracach Mongolda [22] oraz Wolfa [23] , gdzie stwierdza się możliwość wygenerowania mowy brzmiącej naturalnie, lecz zupełnie niezrozumiałej. Poniżej podano znaczenie stosowanych w pracy pojęć związanych z omówionym zagadnieniem.

Def.8. Jakość mowy syntetycznej: miara zbieżności mowy syntetycznej do mowy naturalnej z uwzględnieniem wszystkich poziomów wytwarzania mowy.

Def.9. Wierność odtworzenia formalnych elementów systemu języka: miara równowartościowości przekształcenia ciągu symboli wygenerowanych z wyższych poziomów syntezy (od

fonologicznego wzwyż).w zbiór odpowiadających im składników akustycznych.

Def.10.Naturalność brzmienia: subiektywna miara podobieństwa dźwiękowej struktury mowy syntetycznej do struktury mowy naturalnej.

Pierwsza z wprowadzonych miar - jakość - ma charakter uniwersalny (m.in. zawierają się w niej dwie następne miary) i na aktualnym etapie syntezy nie znajduje zastosowania.

Wierność odtwarzania formalnych elementów systemu języka, co należy wyraźnie podkreślić, jest miarą oceny wyłącznie reguł realizacyjnych i sterowanego przez nie procesu syntezy dźwięków mowy w aparacie syntezy. W przypadku otrzymania z komponentu fonologicznego ciągu symboli obarczonych błędami formalnymi (np. niewłaściwe umiejscowienie akcentu lub pauzy), oraz przy prawidłowo ustalonej gramatyce reguł realizacyjnych i prawidłowej procedurze syntezy dźwięków błąd ten zostanie odwzorowany w zbiorze składników akustycznych, jednak w sensie przyjętej miary procedura syntezy na poziomie reguł realizacyjnych oraz na poziomie syntezy dźwięku zostanie oceniona, zgodnie ze stanem faktycznym, jako prawidłowa(jest tu jednak wymagana znajomość ciągu T z komponentu fonologicznego). Wierność odtworzenia jest miarą subiektywną, (por. definicję składników akustycznych z rozdz.1.3.1.) i w pracy przyjęto ją jako podstawowe kryterium oceny prawidłowości doboru reguł realizacyjnych[9)].

Naturalność brzmienia jest miarą uzupełniającą w stosunku do wierności odtwarzania. W pracy stosowano ją w proce-

---

9)Warto zauważyć, że wyrazistość głoskowa jest jednym z elementów omawianej miary, gdyż jest to wierność odtwarzania symboli oznaczających podstawowe jednostki fonetyczne (fonemy), które są podzbiorem alfabetu symboli terminalnych komponentu fonologicznego.

durach optymalizacyjnych, gdy w oparciu o miarę wierności odtwarzania (lub innymi metodami) ustalono już reguły realizacyjne i dopuszczalny obszar zmienności parametrów w ich dziedzinie zapewniający uzyskanie żądanej cechy (np. wrażenia akcentu lub intonacji pytającej), a przedmiotem badań był dobór wartości lub zbioru wartości danego parametru zapewniających najbardziej naturalne brzmienie frazy syntetycznej. W przypadku badań nad optymalizacją parametrów syntezy dźwięków mowy naturalność brzmienia była kryterium podstawowym (i jedynym).

Def.11. Dobór reguł realizacyjnych: ustalenie typu reguły i obszaru zmienności parametrów jej dziedziny zapewniających jednoznaczne odwzorowanie danej cechy (jednostki) fonetycznej w zbiorze odpowiadających jej składników akustycznych.

Def.12. Optymalizacja parametrów dziedziny reguł realizacyjnych oraz parametrów syntezy dźwięków mowy: ustalenie, w zadanym obszarze zmienności, wartości lub zbioru wartości parametrów zapewniających najbardziej naturalne brzmienie syntezowanej frazy (por. wyżej zamieszczony komentarz do naturalności brzmienia).

1.4. Rola pobudzenia krtaniowego w syntezie mowy

1.4.1. Uwagi wstępne

Analiza strukturalnego modelu Chomsky'ego oraz hierarchicznych systemów syntezy (rozdz.1.1. i 1.2.) wskazuje na zasadniczą trudność w ustalaniu reguł syntezy mowy, jaką jest wpływ wszystkich poziomów syntezy na odpowiadające im składniki akustyczne. Często szereg różnych elementów formy języka "steruje" (za pośrednictwem reguł realizacyjnych) równocześnie tym samym składnikiem akustycznym, przy czym dodatkowo nakładają się na ten proces czynniki nie stanowiące elementów systemu języka (por. rozdz.1.2. s.8). Przy ustalaniu reguł sterowania daną wielkością fizyczną współtworzącą dźwiękową strukturę mowy (np. w niniejszej pracy jest nią pobudzenie krtanio-

we), niezbędne jest określenie:

1. Parametrów opisujących daną wielkość,
2. Związku parametrów z p-tu 1. z akustycznymi składnikami sygnału mowy,
3. Zbioru cech i elementów formy języka odwzorowywanych w składnikach akustycznych z p-tu 2.,
4. Zbioru czynników nieinterpretowalnych w systemie języka, a wpływających na parametry z p-tu 1.,
5. Podziału parametrów z p-tu 1. na parametry dziedziny realizacyjnych reguł syntezy (por. Def.6) i parametry syntezy dźwięków mowy (Def.7).

Powyższą analizę przeprowadzoną dla pobudzenia krtaniowego podano w następnych rozdziałach.

1.4.2. Związki parametrów pobudzenia krtaniowego[10] ze składnikami akustycznymi sygnału mowy

Pobudzenie krtaniowe stanowi w procesie syntezy, podobnie jak w procesie wytwarzania mowy naturalnej, quasiperiodyczny przebieg odpowiednio ukształtowanych impulsów. Częstotliwość powtarzania impulsów pobudzenia (F_0) jest bezpośrednio związana z wysokością tonu[11] [55, 56] natomiast amplituda impulsów pobudzenia, jak wykazały badania Browna i Mc Glone'a [49] jest silnie skorelowana z poziomem intensywności [11,12]. Zmiany w kształcie impulsów pobudzenia, które są trzecim, rozważanym

---

[10] Określenie "pobudzenie krtaniowe", ścisłe w odniesieniu do pobudzenia w mowie naturalnej, nie jest adekwatne w przypadku syntezy dźwięków mowy, gdzie powinno stosować się określenia: "sygnał symulujący pobudzenie krtaniowe" [59], "quasiperiodyczny przebieg pobudzający" lub podobne. W pracy przyjęto stosować określenie "pobudzenie krtaniowe" zarówno w przypadku mowy naturalnej jak i syntetycznej (por. również pracę J.Kacprowskiego i W.Mikiela [31].

[11] Wysokość tonu i poziom intensywności są cechami wrażenia słuchowego, a więc jakością subiektywną [56]. Równocześnie, zgodnie z Def.4, należą do zbioru składników akustycznych.

[12] Oprócz amplitudy pobudzenia krtaniowego, na poziom intensywności wpływają parametry transmisyjne kanału głosowego (lub jego modelu). Zmiany tych parametrów związane są z sekwencją

w syntezie parametrem pobudzenia krtaniowego, związane są z subtelnymi zmianami mikrostruktury zespolonego widma [59, 60] . W przestrzeni percepcji słuchowej tym zmianom odpowiada zmiana barwy dźwięku.

1.4.3. Podzbiór cech i elementów formy języka odwzorowywanych w składnikach akustycznych związanych z parametrami pobudzenia krtaniowego

1.4.3.1. Wstęp

Szczegółowe omówienie zagadnienia wyodrębnienia podzbioru podanego w tytule rozdziału wykracza znacznie poza ramy nakreślone tematem pracy[13]. Ograniczono się zatem do rozważenia związków zachodzących w obrębie krótkich fraz mowy polskiej.

Wprowadźmy następujące definicje:

Def.13. Fraza: element dziedziny reguł syntaktycznych, ograniczony dwustronnie początkową i końcową cechą graniczną, z zerowym prawdopodobieństwem wystąpienia cech granicznych lub pauz pomiędzy elementami tworzącymi frazę (wyrazy).

Def.14. Zestrój akcentowy[14]: grupy sylab podporządkowane prozodyjnie wspólnemu akcentowi o jednym, wspólnym akcencie głównym (Dłuska, [51]).

Def.15. Zestrój intonacyjny: jednostka prozodyjna nadrzędna w stosunku do zestroju akcentowego. Zależnie od treści i sensu wypowiedzi, zestrój intonacyjny może pokrywać się z jednym i to krótkim jednosylabowym zestrojem akcentowym lub może obejmować ich wiele (Dłuska, [51]).

Przyjmijmy ponadto założenia:

---

dyskretnych jednostek fonetycznych, zatem rozpatrując poziom intensywności jako składnik akustyczny cech o charakterze suprasegmentalnym, można przyjąć bezpośredni związek amplitudy pobudzenia z intensywnością.

13) Szersze naświetlenie problemu zawarto w pracach Rynowskiego [56], Majewskiego i Zalewskiego [55] , Allena [8, 12] oraz Umedy [13] .

14) Definicje przytoczone za innymi autorami zaopatrzono w odpowiednie cytowania. Definicje bez cytowania stanowią sformułowania podane przez autora.

Zał.1. W obrębie frazy, jako jednostki struktury suprasegmentalnej, realizowany jest obligatoryjnie zestrój akcentowy i możliwa jest fakultatywna realizacja zestroju intonacyjnego.

Zał.2. Fraza stanowi jeden, niepodzielny element formy języka[15].

Założenie 2 jest konsekwencją przyjęcia w pracy założenia 1 oraz podanej przez Dłuską [51] koncepcji zestroju akcentowego i intonacyjnego jako zorganizowanych, całościowych i niepodzielnych jednostek prozodyjnych (por. Def.13, 14). Celowe jest tu uwypuklenie niezwrotności relacji fraza - zestrój, gdyż możliwe jest wygenerowanie frazy nie obciążonej zestrojem intonacyjnym czy akcentowym, natomiast rozważanie sytuacji odwrotnej jest pozbawione sensu. Wynikają stąd następujące wnioski:

W.1. Fraza jest elementem prymarnym w stosunku do zestroju intonacyjnego i akcentowego.

W.2. Realizacja zestroju intonacyjnego i akcentowego w obrębie frazy nie wprowadza nowych składników w akustycznej substancji frazy, natomiast dokonuje przekształceń na akustycznych składnikach frazy.

1.4.3.2. Akustyczne składniki frazy związane z pobudzeniem krtaniowym

Przyjęcie definicji i założeń podanych w rozdz.1.4.3.1. oraz uwzględnienie zamieszczonych tam wniosków sprowadza zagadnienie rozważane w niniejszym rozdziale do określenia składników

---

15) Założenie 2 nie jest całkiem zgodne z ogólnie przyjętym stanowiskiem [5, 11-13, 51] , gdzie stwierdza się, że cechy graniczne stanowią samodzielny element formy języka. Na tym stanowisku zaważyło odrębne analizowanie w fonetyce opisowej cech nagłosowych oraz wygłosowych (np. kadencja - antykadencja [51] , typ nastawienia fonacyjnego [61] itp.). W pracach [11-13] zdaniem autora mechanicznie przeniesiono fakt generowania przez komponent syntaktyczny dwóch odrębnych symboli dla początkowej i końcowej

akustycznych frazy i wskazania, na których składnikach frazy są dokonywane przekształcenia związane z realizacją zestroju akcentowego i intonacyjnego. W pracy, kierując się założeniem 2, przyjęto konsekwentnie jeden składnik akustyczny frazy jakim jest grupa wydechowa BG .

Pojęcie grupy wydechowej wprowadził Lieberman [50] przyjmując za punkt wyjścia podporządkowanie procesu wytwarzania mowy w substancji dźwiękowej warunkom związanym z procesem oddychania (por. op.cit. s.26). Jak wykazały eksperymentalne badania przeprowadzone przez Liebermana (op.cit,) realizacji frazy w obrębie grupy wydechowej odpowiadają powtarzalne i charakterystyczne przebiegi czasowe częstotliwości podstawowej i ciśnienia podkrtaniowego. Ponieważ ciśnienie podkrtaniowe wykazuje silną korelację z amplitudą pobudzenia krtaniowego [50] , a więc i z poziomem intensywności [49] , zatem można sformułować wniosek:

W.3. Akustycznym składnikiem frazy jest grupa wydechowa BG . Elementarnym składnikiem grupy wydechowej są charakterystyczne przebiegi czasowe częstotliwości podstawowej i intensywności.

Koncepcja grupy wydechowej została również wykorzystana w wielu innych pracach dotyczących syntezy mowy. Można tu wymienić prace Ainswortha [62], Leufera [58], Allena [12] oraz Umedy i Teranishi [63] .

Zestrój akcentowy i intonacyjny jako jednostki prozodyczne mowy polskiej mają ścisły związek z przebiegiem częstotliwości podstawowej [54-56, 64] . Z kolei badania Jassema i in. [54] oraz Nowakowskiej [65] wykazały, że w przypadku akcentu istnieje również związek z poziomem intensywności. Zestrój akcentowy oraz intonacyjny dokonują zatem przekształceń na grupie wydechowej, tworząc akcentowe i intonacyjne warianty frazy.

---

cechy granicznej, nie zwracając uwagi na oczywistą implikację $S^i \Rightarrow S^t$

gdzie: S^i - symbol początkowej, S^t - symbol końcowej cechy granicznej.

1.4.4. Wpływ czynników nie mających interpretacji na płaszczyźnie języka na parametry pobudzenia krtaniowego

Do grupy omawianych czynników można zaliczyć:
- czynniki osobnicze
- czynniki artykulacyjne
- czynniki związane z budową i funkcjonowaniem krtani.

Pierwszą grupę czynników pomija się na aktualnym etapie syntezy.

Czynniki artykulacyjne jak pionowe ruchy krtani czy położenie języka przy artykulacji samogłosek mają charakter segmentalny i zależą od typu głoski. Ze względu na brak w literaturze dokładniejszych danych dotyczących wpływu tych czynników na parametry pobudzenia krtaniowego, pominięto je w dalszych rozważaniach.

Budowa i funkcjonowanie krtani w istotny sposób wpływa na uzyskiwany sygnał pobudzenia. Jednak dokładne rozróżnienie, które funkcje krtani są wykorzystywane w odwzorowaniu cech lingwistycznych, a które mają charakter uboczny, nie jest jeszcze możliwe, choćby ze względu na niewygasły spór między zwolennikami mioelastycznej teorii czynności fonacyjnej krtani, podanej przez van den Berga [70] , a zwolennikami teorii mięśniowo-nerwowej Hussona (por.omówienia tego problemu w pracach [66, 67]. Problemu nie rozwiązało opracowanie przez Flanagana i Ishizakę [68, 69] artykulacyjnego modelu źródła pobudzenia krtaniowego, gdzie uwzględniono szereg fizycznych i anatomicznych czynników wpływających na czynność fonacyjną krtani, jak sprężystość wiązadeł głosowych, wpływ powierzchni w neutralnej pozycji fonacyjnej, wpływ parametrów transmisyjnych kanału głosowego, wpływ ciśnienia podkrtaniowego, wpływ naprężenia wiązadeł głosowych oraz wpływ wzdłużnych przemieszczeń w drgających wiązadłach głosowych. Przyczynę, dlaczego tak szczegółowy i rozbudowany model nie pozwolił na rozwiązanie rozważanego problemu, lapidarnie ujęto w pracy Rothenberga i in. [60] gdzie stwierdzono: "The psychological approach presupposes a reasonably accurate model of glottal source ...,

however, introduces in a natural way parameters whose effects are not yet known...". Mimo tych trudności możliwe jest w oparciu o publikowane wyniki eksperymentów i analiz [50, 60, 68, 71-76] zestawienie następujących czynników należących do omawianej w rozdziale kategorii:

1. skorelowanie przebiegu $F_0(t)$ z ciśnieniem podkrtaniowym,
2. skorelowanie przebiegu $A_0(t)$ z ciśnieniem podkrtaniowym,
3. występowanie niskoczęstotliwościowej modulacji F_0,
4. występowanie losowych dewiacji F_0, szczególnie wyraźne na końcu grupy wydechowej,
5. skorelowanie zmian w kształcie impulsów pobudzenia krtaniowego z intensywnością (im wyższa intensywność , tym węższe impulsy pobudzenia),
6. rozpoczynanie wytwarzania pobudzenia od pewnej, niezerowej amplitudy (tzw. opóźnienie fonacyjne),
7. zmiany w kształcie impulsów pobudzenia oraz obniżenie ich częstotliwości powtarzania na początku i na końcu grupy wydechowej.

Z punktu widzenia reguł realizacyjnych czynniki 1-7 nie tworzą jednorodnej klasy. Przykładowo czynniki 1 i 2 można zastąpić jednym czynnikiem korelującym $F_0(t)$ z $A_0(t)$ pod warunkiem, że ciśnienie podkrtaniowe nie jest parametrem reguł. Dla fraz z dźwięcznymi elementami fonetycznymi w nagłosie i wygłosie czynniki 6 oraz 7 można traktować jako składniki cech granicznych.

W pracy uwzględniono czynniki 1-3 oraz 6. W odniesieniu do losowych dewiacji F_0 (czynnik 4) brak dokładniejszych danych dotyczących źródeł tego zjawiska oraz jego ilościowego opisu spowodował jego pominięcie. Dodatkowo, celem ograniczenia liczby rozpatrywanych parametrów reguł oraz parametrów syntezy, przyjęto stały kształt impulsów pobudzenia w obrębie syntezowanej frazy, eliminując tym samym czynniki 5 i 7. Nie założono jednak a priori kształtu impulsów.

1.4.5. Podsumowanie

W analizie przeprowadzonej w rozdziale 1.4. wykazano, że w obrębie fraz konieczne jest uwzględnienie szeregu czynników wpływających na generowane pobudzenie krtaniowe. Wyodrębniono dwie grupy czynników interpretowalnych i nieinterpretowalnych na płaszczyźnie językowej i stwierdzono, że przy przyjęciu pewnych założeń te czynniki oddziaływują na dwa parametry pobudzenia krtaniowego - częstotliwość podstawową F_0 i amplitudę A_0. Pozorną dysproporcję między licznym zbiorem czynników oddziaływujących a dwuelementowym zbiorem parametrów pobudzenia, na które te czynniki wpływają, wyjaśniają założenia wynikające z analizy podanej w rozdziale 1.4.:

Zał.3. W procesie generacji fraz pobudzenie krtaniowe związane jest wyłącznie ze strukturą suprasegmentalną.

Zał.4. Składnik akustyczny frazy - grupa wydechowa - stanowiący złożenie archetypowych wzorców czasowego przebiegu zmian wysokości tonu podstawowego i intensywności sygnału mowy, jest prymarny w stosunku do wszystkich pozostałych czynników wpływających w obrębie frazy na jej strukturę suprasegmentalną.

Zał.5. Wszystkie transformacje przebiegu $F_0(t)$ i $A_0(t)$ w obrębie frazy stanowią warianty grupy wydechowej w kontekście danej cechy.

Zał.6. Zbiorem parametrów pobudzenia krtaniowego w procesie syntezy suprasegmentalnej struktury frazowej jest zbiór wariantów grupy wydechowej (tzn. zbiór wariantów czasowych przebiegu $F_0(t)$ i $A_0(t)$.

Założenia 3-6 oraz dokonane w niniejszym rozdziale ustalenie zbioru czynników wpływających na generację pobudzenia krtaniowego stanowią punkt wyjścia do sformułowania celów i podstawowych założeń pracy i mają istotne następstwa metodologiczne przy ustaleniu struktury komponentu reguł realizacyjnych.

1.5. Cel i podstawowe założenia pracy

1.5.1. Uwagi wstępne

W cytowanych uprzednio pracach dotyczących syntezy mowy nie omówiono w sposób szczegółowy i wyczerpujący problemu metodyki oraz kryteriów doboru wartości parametrów dziedziny reguł realizacyjnych. Allen 8,[12] , Coker i in. [9] , Klatt [11], Umeda [13], Fallside i in. [14] , Łaufer [58] i Ainsworth [62] skupiając uwagę na formalnych aspektach syntezy podali tylko ogólny zarys strategii tworzenia reguł realizacyjnych i nie zamieścili konkretnych danych dotyczących wartości paramotrów, na których po stronie dziedziny (funkcje F^1 i F^2) te reguły operują.

Mattingly [27] , Umeda i Teranishi [63] oraz Coker i Umeda [71] w opisie proponowanego zbioru reguł realizacyjnych zamieścili szereg tabel z konkretnymi wartościami parametrów (np. F_0, intensywność, czasy trwania poszczególnych fonemów), lecz przyjęta przez autorów zasada tworzenia cech suprasegmentalnych w oparciu o konkatenację cech segmentalnych jest zasadniczo różna od koncepcji przyjętej w niniejszej pracy i wobec tego nie jest możliwe bezpośrednie wykorzystanie w niej wyników wymienionych prac.

Wartości parametrów podane przez Rabinera [57] , Kacprowskiego i Mikiela [31] czy Rao i Thosara [77] , dotyczące funkcji sterujących amplitudą i częstotliwością pobudzenia krtaniowego w syntezie fraz, nie stanowią, w podanym przez autorów ujęciu, parametrów dziedziny reguł realizacyjnych oraz nie są wynikiem systematycznych badań nad ich optymalnym wyborem.

Na opisany stan badań nad ustaleniem parametrów dziedziny reguł realizacyjnych złożyły się następujące czynniki:

1. Przytoczone prace należą do drugiego etapu badań nad syntezą mowy, gdzie główną uwagę skupiono na fundamentalnych zagadnieniach jak generacja formalnych struktur języka i ich związkach jakościowych ze strukturą dźwięków mowy oraz na odwzorowaniu w strukturze dźwiękowej podstawowych ele-

mentów informacyjnych, łączonych zazwyczaj z cechami segmentalnymi. W konsekwencji problem naturalności brzmienia oraz jakości odwzorowania cech prozodycznych (nie mówiąc już o cechach ekspresyjnych czy osobniczych) traktowano jako wtórny (por. rozdz. 1.2.).

2. Optymalizacja parametrów dziedziny reguł realizacyjnych lub parametrów syntezy dźwięków mowy (Def.12) jest procedurą typowo eksperymentalną, żmudną, praco- i czasochłonną, oraz kosztowną (wymaga przeprowadzenia szeregu badań odsłuchowych). Dodatkową komplikacją jest konieczność stosowania metody kolejnych prób i błędów (Kacprowski, Mikiel [31], s.362; Mangold, Stall [22] s.140), gdyż w tworzeniu dźwiękowej struktury mowy szereg czynników "nakłada się" wzajemnie na siebie i sekwencyjna optymalizacja jednoparametryczna nie zapewnia uzyskania optimum w całym zbiorze parametrów (tzn. optimum dla jednego parametru jest złożoną funkcją wartości innych parametrów i wymagane jest asymptotyczne zbliżanie się równocześnie do optimum dla wszystkich rozważanych parametrów - czyli do punktu optimum optimorum w wielowymiarowej przestrzeni parametrów)[16].

---

16) Ten czynnik zadecydował o stwierdzonym w rozdziale 1.2. relatywnym opóźnieniu prac nad regułami realizacyjnymi i o sugerowanym przez Umedę [13] zarzuceniu dalszych badań nad tym problemem techniką analizy przez syntezę na rzecz skrupulatnego przebadania relacji forma - substancja w mowie naturalnej (etap trzeci) i ustalenia na tej podstawie gramatyki reguł realizacyjnych oraz zbioru parametrów jej dziedziny.

Procedurę optymalizacji można zapisać:

$$(d_{1,k_1}, \ldots, d_{i,k_i}, \ldots, d_{n,k_n}) \in D_{DF} \Rightarrow$$

$$\Rightarrow Q(d_{1,k_1}, \ldots, d_{n,k_n}) \rightarrow \text{Max } Q\left(\bigtimes_{i=1}^{n} D_i\right) \qquad (1.2.)$$

gdzie:

$\bigvee_{i \in \{1,\ldots,n\}} D_i = [i,k_i \in \{1,\ldots,m_i\} : d_{i,k_i}]$ - zbiór rozważanych wartości i-tego parametru dziedziny reguł realizacyjnych,

Q - funkcja optymalizująca,

DF - dziedzina funkcji realizacyjnych,

D_{DF} - zbiór parametrów dziedziny reguł realizacyjnych

$\bigcup_{i=1}^{n} D_i \subset D_{DF}$.

Z zależności (1.2.) wynika, że przy większej liczbie parametrów, z których każdy opisano zbiorem m_i jego wartości, przeprowadzenie procedury optymalizacji zgodnie z definicją jest praktycznie niemożliwe, gdyż wymagałoby określenia metodami subiektywnymi $m = \prod_{i=1}^{n} m_i$ wartości funkcji Q. W konsekwencji procedury optymalizacji stosowane w badaniach mają charakter heurystyczny i przy ich realizacji wykorzystuje się znajomość zjawisk i zależności występujących w mowie naturalnej, założenia upraszczające, wyniki odpowiednio zaplanowanych eksperymentów wstępnych a nawet i intuicję. Często stosuje się też optymalizację wzdłuż osi jednego parametru, przyjmując wartości pozostałych parametrów jako stałe. Ta metoda istotnie upraszczająca eksperymenty jest obarczona jednak ryzykiem przyjęcia za optimum jednego z wielu możliwych optimów lokalnych. Dodatkowo, przy niewłaściwym doborze wartości parametrów stałych, zachodzi zjawisko "maskowania" ewentualnej poprawy jakości mowy syntetycznej w funkcji badanego parametru przez jej niską

jakość pochodzącą od źle dobranych parametrów [35, 78] . Ten czynnik, często pomijany w dotychczasowych badaniach z zastosowaniem optymalizacji jednoparametrycznej, może w istotny sposób zniekształcić otrzymane wyniki, gdyż jak stwierdzają Mangold i Stall "... A pitch contour which is only slightly wrong can produce very unnatural speech ... Correctly realizing formant transitions ... lends a natural quality to synthetic speech ..." itd.[17].

Przedstawione omówienie wybranych problemów metodologicznych związanych z optymalizacją parametrów dziedziny reguł realizacyjnych wykazuje, iż w tego typu badaniach planowanie eksperymentów odgrywa zasadniczą rolę i stanowi integralny element dokonań naukowo-badawczych.

1.5.2. Określenie celu pracy

Celem niniejszej pracy jest podanie podzbioru reguł realizacyjnych $RF \in R$ generujących w oparciu o pewną gramatykę $G_{RF} \in G_R$ funkcje czasowe sterujące pobudzeniem krtaniowym w procesie syntezy krótkich fraz mowy polskiej. Ustalono następujące szczegółowe cele pracy:

1. Przeprowadzenie, na podstawie analizy teoretycznej i publikowanych oraz własnych wyników eksperymentów, procedury wstępnego doboru podzbioru RF reguł realizacyjnych R (Def.11).

2. Opracowanie cyfrowego modelu syntezy dźwięcznych fraz mowy polskiej sterowanego wybranym w p-cie 1 podzbiorem reguł RF.

---

[17] Warto też zwrócić uwagę na dużą rozróżnialność przez ucho ludzkie zmian w strukturze dźwięków mowy. Por. rozdz.7.2. w [59] .

3. Zaplanowanie i wykonanie eksperymentów mających na celu:
   a) optymalizację parametrów dziedziny reguł RF (Def.6 i 12),
   b) optymalizację parametrów syntezy dźwięków mowy (Def. 7 i 12),
4. Ostateczne ustalenie reguł RF wraz z optymalnymi parametrami ich dziedziny oraz podanie gramatyki G_{RF} zarządzającej wykonaniem (implementacją) tych reguł.

1.5.3. Założenia przyjęte w pracy

1.5.3.1. Wstęp

Podstawowe założenia wynikające ze strukturalnej analizy problemu stanowiącego temat pracy podano w rozdziałach 1.4.3.1. (Zał.1 i 2) oraz 1.4.5. (Zał.3-6). Poniżej wyszczególniono dodatkowe założenia dotyczące metodologii i techniki realizacji eksperymentalnej części pracy:

Zał.7. Materiał eksperymentalny - dwu lub trzysylabowe frazy złożone z dźwięcznych fonemów mowy polskiej.

Zał.8. Technika syntezy - formantowa.

Zał.9. Model układu syntezy - cyfrowy syntezator formantowy w układzie kaskadowym.

Zał.10. Model źródła pobudzenia krtaniowego - cyfrowy generator przebiegów okresowych.

Zał.11. Typ funkcji źródła - funkcje analityczne podane przez Rosenberga [75].

Założenia 7-11 nie mają charakteru arbitralnego lecz wynikają z przyjętej koncepcji pracy, stanu dotychczasowych badań nad syntezą mowy polskiej oraz podanych celów pracy. Krótkie omówienie tych założeń zamieszczono w następnych rozdziałach.

1.5.3.2. Materiał eksperymentalny

Sformułowany temat pracy zawęża problem doboru reguł generacji pobudzenia krtaniowego do obrębu frazy, co niewątpliwie upraszcza zagadnienia, jednak celowe jest przytoczenie czynników, które na to wpłynęły:

1. W obrębie frazy mowy polskiej jest realizowany zestrój akcentowy i intonacyjny (por. Def.13-15 oraz Zał.1.), co umożliwia badanie podstawowych cech prozodycznych bez stosowania dłuższych jednostek strukturalnych mowy (np.zdania).
2. Grupa wydechowa BG jako składnik akustyczny frazy ma ograniczony czas trwania i jego maksymalna wartość wynosi 2400 ms [63] , przy czym badania eksperymentalne nad syntezą mowy angielskiej wykazały, że skrócenie czasu trwania BG poprawia wyrazistość (op. cit.).
3. Przyjęta metoda optymalizacji parametrów dziedziny reguł realizacyjnych opiera się o subiektywne badania odsłuchowe. Należy się więc liczyć z krótkoterminowym charakterem zapamiętywania przez słuchaczy struktury dźwiękowej prezentowanych bodźców syntetycznych. Dlatego w pracach dotyczących syntezy mowy, gdzie stosowano formalne lub nieformalne badania odsłuchowe, materiałem testowym były najczęściej krótkie frazy o rozciągłości od pojedynczej samogłoski do krótkiego zdania [28-31, 54, 64, 75, 76, 79-82, 87] .
4. W pracach nad syntezą mowy polskiej [17, 28-31, 54-56, 64] badano zjawiska wyłącznie na poziomie frazy.

Frazy syntetyczne generowane w pracy - stanowiące materiał eksperymentalny - były złożone wyłącznie z fonemów dźwięcznych. To upraszczało zagadnienie sterowania modelem kanału głosowego i nie wymagało rozpatrywania dodatkowych czynników jak przerwanie generacji pobudzenia krtaniowego w obrębie fonemów bezdźwięcznych lub generacja przebiegu szumowo-quasiperiodycznego dla fonemów o pobudzeniu mieszanym.

1.5.3.3.Technika i model układu syntezy

Zaletą przyjętej w pracy techniki syntezy widmowo-parametrycznej opartej o formantowy model kanału głosowego jest możliwość bezpośredniego wykorzystania w niej wyników badań fonetyczno-akustycznych [9, 12, 76] . Dodatkowo przy wyborze tej techniki syntezy uwzględniono:

1. Synteza formantowa była w momencie ustalania założeń pracy jedyną techniką stosowaną w badaniach nad mową polską (por. prace Majewskiego i in. [28, 55, 64] , Kacprowskiego [83], Kacprowskiego i Mikiela [17, 18, 31, 39] , Kudeli [29, 30].
2. Syntezator formantowy jest dokładnym modelem kanału głosowego i jak wykazał Holms [76] i Rosenberg [75] uzyskiwane z jego zastosowaniem frazy mowy syntetycznej są nie do rozróżnienia od fraz mowy naturalnej.

O wyborze cyfrowego syntezatora formantowego w układzie kaskadowym zadecydowało:

1. W przypadku syntezy samogłosek w układzie kaskadowym nie jest wymagane sterowanie amplitudą poszczególnych formantów [76, 83, 84] .
2. Syntezatory cyfrowe zapewniają stabilne oraz powtarzalne generowanie próbek mowy syntetycznej o dużej dokładności i rozdzielczości sterowania badanymi parametrami, i z tego względu są powszechnie stosowanym narzędziem badawczym w zagadnieniach związanych z optymalizacją parametrów reguł syntezy [57, 75-77, 82, 85, 86] .

1.5.3.4.Model i typ funkcji źródła pobudzenia krtaniowego

Badanie fizycznych charakterystyk pobudzenia krtaniowego w mowie naturalnej było przedmiotem szeregu prac eksperymentalnych i teoretycznych [59, 74, 88-92, 98] . Stwierdzono, że impulsy generowane przez krtań mają postać odkształconego przebiegu piłozębatego [59, 74, 88, 90] , z licznymi

lokalnymi nieregularnościami oraz z wyraźną nieciągłością pierwszej pochodnej na końcu zbocza opadającego (tzn. w momencie zwarcia wiązadeł głosowych) [75, 91] . Średni spadek obwiedni widma amplitudowego wynosi -12 dB/oktawę, widmo jest nieregularne, z licznymi zerami zespolonymi leżącymi zarówno w lewej jak i prawej półpłaszczyźnie płaszczyzny zmiennej zespolonej s (zatem widmo pobudzenia krtaniowego ma składową nieminimalnofazową) [59, 88, 90] . Wierność odtworzenia złożonej struktury widma pobudzenia krtaniowego jest jeszcze jednym czynnikiem, który należy uwzględnić przy wyborze modelu źródła krtaniowego w procesie syntezy dźwięków mowy (inne czynniki omówiono w rozdz.1.4.4.).

Modele źródła pobudzenia krtaniowego stosowane w syntezatorach można podzielić na dwie zasadniczo różne klasy [60] :

- modele artykulacyjne (fizjologiczne),
- modele akustyczne.

Modele artykulacyjne zarówno w wersji cyfrowej [68, 69, 94] jak i analogowej [92, 93], z przyczyn podanych w rozdz.1.4.4. (s.24) oraz w wyniku przyjęcia założeń 8 i 9, nie mogły znaleźć zastosowania w niniejszej pracy. Z tych samych powodów nie zastosowano trójparametrycznego modelu źródła pobudzenia krtaniowego opracowanego przez Rothenberga i in. [60].

W obszernej literaturze dotyczącej symulacji pobudzenia krtaniowego w oparciu o kształtowanie jego cech czasowo-widmowych (modele akustyczne) można wyróżnić trzy metody symulacji:

1. metodę kształtowania obwiedni widma amplitudowego [85, 95, 96] ,
2. metodę kształtowania impulsów pobudzenia funkcjami czasu [59, 75, 76, 82, 97] ,
3. metodę stanowiącą połączenie metody 1 i 2 [98, 99] .

W pracy założono stosowanie metody 2 (Zał.11), gdyż zapewnia ona nie tylko uzyskanie obwiedni widma amplitudowego generowanych impulsów zgodnej z obwiednią impulsów krtaniowych w mowie naturalnej [59, 75, 88, 98] , ale również odtwarza subtelną mikrostrukturę widma z minimalno i nieminimalnofazowym rozkła-

dem zer własnych w widmie [59, 88, 98] [18]. Dodatkowo ta metoda, co wykazały badania Rosenberga [75] i Holmesa [76], zapewnia uzyskanie bardzo dobrej naturalności brzmienia mowy syntetycznej oraz dzięki ścisłemu określeniu czasu trwania zerowych i niezerowych wartości impulsów pobudzających umożliwia badanie wpływu zmian parametrów modelu kanału głosowego, synchronicznych z interwałami zerowych wartości pobudzenia, na naturalność mowy syntetycznej [37].

Wybór cyfrowego modelu generacji pobudzenia krtaniowego jest konsekwencją przyjęcia Zał.9, natomiast zastosowanie podanych przez Rosenberga funkcji symulujących pobudzenie [75] wynika z ich prostego opisu analitycznego oraz z dobrych rezultatów eksperymentów przeprowadzonych z ich wykorzystaniem dla syntetycznych fraz mowy angielskiej [75, 76] [19].

---

[18] Obszerne omówienie zagadnienia symulacji fizycznej struktury impulsów krtaniowych odpowiednio dobranymi funkcjami czasowymi podano w pracach Flanagana [57], Millera [88] i Kacprowskiego [98].

[19] Funkcje podane przez Rosenberga zastosowano również w 1977 roku (a więc już po ustaleniu koncepcji niniejszej pracy i opublikowaniu wyników szeregu eksperymentów przeprowadzonych przez autora z zastosowaniem omawianych funkcji [33-36]) w badaniach nad optymalizacją parametrów pobudzenia krtaniowego w syntezie mowy metodą LPC [82].

1.6. Układ pracy

Praca składa się z siedmiu głównych rozdziałów, wykazu literatury oraz trzech dodatków.

W rozdziale wstępnym (rozdz.1.) podano podstawy strukturalnej teorii wytwarzania mowy i na tej podstawie dokonano przeglądu aktualnego stanu badań nad syntezą mowy, przeprowadzono analizę roli pobudzenia krtaniowego w procesie syntezy i ustalono związki między parametrami pobudzenia a składnikami akustycznymi sygnału mowy. W tym rozdziale określono również cel i założenia pracy oraz podano podstawowe określenia i definicje stosowane w pracy.

W rozdziale 2. opisano i podano wyniki eksperymentów wstępnych dotyczących optymalizacji parametrów pobudzenia krtaniowego w procesie syntezy samogłosek i zamieszczono podstawowe informacje dotyczące metody generacji samogłosek syntetycznych oraz techniki oceny uzyskanego materiału dźwiękowego.

W rozdziale 3. podano zbiór RF reguł realizacyjnych generujących funkcje sterujące pobudzeniem krtaniowym w procesie syntezy fraz oraz dokonano podziału rozważanego w pracy zbioru parametrów pobudzenia krtaniowego na parametry dziedziny reguł RF oraz paramatry syntezy dźwięków mowy.

Rozdział 4. poświęcono omówieniu zaprojektowanego w pracy cyfrowego modelu syntezy dźwięcznych fraz mowy ze szczególnym uwzględnieniem struktury komponentu reguł realizacyjnych.

W rozdziale 5. omówiono metody subiektywnej oceny fraz syntetycznych. Podano metody oceny wprowadzonego przez autora kryterium wierności odtwarzania formalnych elementów systemu języka oraz metody oceny naturalności brzmienia próbek mowy syntetycznej. Szczególną uwagę zwrócono na metody oparte o testy porównań w parach, dla których podano teoretyczny model dokonywania ocen preferencyjnych w przestrzeni percepcji słuchowej, sposób tworzenia ilorazowej skali ocen preferencyjnych oraz metody statystycznej oceny wyników.

Rozdział 6. poświęcono eksperymentom zasadniczym dotyczącym optymalizacji parametrów syntezy dźwięków mowy (rozdz.6.3.2.1. do 6.3.2.5.) oraz optymalizacji parametrów dziedziny reguł RF (rozdz.6.3.3.1. do 6.3.3.6.).

W podsumowaniu (rozdz.7.) podano zestawienie optymalnych wartości parametrów syntezy dźwięków mowy i parametrów dziedziny reguł RF, gramatykę reguł RF zarządzającą ich stosowaniem w procesie syntezy oraz zamieszczono uwagi końcowe dotyczące teoretycznych i eksperymentalnych aspektów pracy.

W dodatkach, stanowiących wyodrębnioną część pracy, podano algorytm i przykładowy wydruk z programu cyfrowej syntezy fraz (Dodatek 1.), program statystycznej analizy wyników testu ocen porównawczych (test A-B) (Dodatek 2.) oraz wykresy zbioru konturów intonacyjnych badanych w eksperymencie EZ II (2) (Dodatek 3.).

W związku z zalecaną tendencją do ograniczania objętości prac doktorskich w ostatecznej redakcji pracy zamieszczono opis tylko części eksperymentów o charakterze końcowym, rezygnując z przytaczania szeregu eksperymentów wstępnych. Pominięto również omówienie zagadnień szeroko znanych i cytowanych w literaturze (np. budowa i funkcjonowanie krtani, filtry cyfrowe) skupiając się na elementach stanowiących pośrednio lub bezpośrednio oryginalne dokonanie autora pracy.

1.7. Wykaz ważniejszych symboli i oznaczeń

(Podano w kolejności pojawiania się w tekście).

G_R - gramatyka reguł realizacyjnych

T - ciąg symboli terminalnych komponentu fonologicznego

$\{F^1, F^2\}$, $\{F^i\}$ - zbiór funkcji czasowych generowanych przez komponent reguł realizacyjnych

DF - dziedzina reguł realizacyjnych

F_o - częstotliwość podstawowa tonu krtaniowego

BG - oznaczenie grupy wydechowej

$F_o(t)$ - czasowy przebieg częstotliwości podstawowej tonu krtaniowego

$A_o(t)$ - czasowy przebieg amplitudy pobudzenia krtaniowego

$\overset{n}{\underset{i=1}{P}} D_i$ - uogólniony produkt kartezjański

$\forall$ - kwantyfikator ogólny

D_{DF} - zbiór parametrów dziedziny funkcji realizacyjnych

$\bigcup A_i$ - uogólniona suma zbiorów

$\subset$ - znak zawierania się (inkluzji) zbiorów

$\in$ - znak należenia do zbioru

G_{RF} - gramatyka reguł realizacyjnych generujących funkcje czasowe sterujące pobudzeniem krtaniowym w procesie syntezy mowy

R - zbiór reguł realizacyjnych

RF - podzbiór reguł ze zbioru R generujących funkcje czasowe sterujące pobudzeniem krtaniowym

$f_A, \ldots, f_E$ - funkcje opisujące kształt impulsów pobudzenia krtaniowego

T_F - czas trwania frazy syntetycznej

$A\{a_1, a_2, \ldots, a_n\}$ - zbiór A o elementach $a_1, a_2, \ldots, a_n$

$A(i \in \{1,\ldots,I\}, j \in \{1,\ldots,J\}, \ldots : a_{i,j,\ldots})$ - zbiór A o elementach $a_{i,j,\ldots}$ o indeksach w zbiorach $\{i\}$, $\{j\}$ $\ldots \in N$

$\cup$ - znak sumy zbiorów

$A \times B$ - produkt kartezjański zbiorów A i B

$A \xrightarrow{SUB} B$ - odwzorowanie zbioru A elementów fizycznych w zbiór B subiektywnych reakcji na elementy zbioru A

SbA - podzbiór zbioru A

$w_{i,j}^{k,l}$ - ocena w teście punktowym uzyskana przez próbkę $p_{i,j}$ dla k-tego typu próbki i l-tej oceny niezależnej

$p_{i,j}$ - próbka z i-tą i j-tą wartością badanego parametru

k - indeks w zbiorze typów próbek

l - indeks w zbiorze ocen niezależnych

i,j - indeksy w zbiorze wartości badanych parametrów

P_s - znacznik akcentu

α_r - znacznik intonacji

F_{k,α_r} - r-ta funkcja intonacyjna

$:=$ - znak podstawienia bezwarunkowego

$\overset{[\beta]}{:=}$ - znak podstawienia warunkowego. Podstawienie jest wykonywane jeżeli spełnione jest $[\beta]$.

R1 $\rightarrow$ R2 - znak kolejności stosowania reguły : R2 jest stosowana po R1

$\xrightarrow{OPT}$ - znak odwzorowania poprzez procedurę optymalizacji

$\cap$ - znak przecięcia (mnożenia) zbiorów

$e \leftarrow p_j$ - e pochodzi ze zbioru o elementach p_j

P - zbiór (alfabet) fonemów

$\emptyset$ - zbiór pusty

RK - podzbiór reguł ze zbioru R generujących funkcje czasowe sterujące modelem kanału głosowego

RT - podzbiór reguł czasowych ze zbioru R reguł realizacyjnych

$\wedge$ - funktor koniunkcji

$A \rightarrow B$ - odwzorowanie zbioru A w zbiór B

$\Longrightarrow$ - funktor implikacji

$\Longleftrightarrow$ - funktor równoważności

Q_i - wartość oczekiwania procesu dyskryminacyjnego dla i-tej próbki

$D_{i,j}$ - estymator różnic pomiędzy wartościami oczekiwanymi procesów dyskryminacyjnych dla i-tej i j-tej próbki

r_s - współczynnik korelacji pomiędzy skalami rangowymi (współczynnik r - Spearmana).

2. EKSPERYMENTY WSTĘPNE - OPTYMALIZACJA PARAMETRÓW POBUDZENIA KRTANIOWEGO W PROCESIE SYNTEZY SAMOGŁOSEK POLSKICH

2.1. Wprowadzenie

Celem wstępnych eksperymentów było ustalenie obszaru wartości parametrów pobudzenia krtaniowego[20], dla których w dalszej części pracy przeprowadzono procedurę doboru reguł realizacyjnych generujących pobudzenie (Def.11) oraz procedurę optymalizacji parametrów dziedziny tych reguł (Def.12). W szczególności celem tych eksperymentów było:

- ustalenie typu funkcji $A_0(t)$ i $F_0(t)$ sterujących amplitudą i częstotliwością pobudzenia krtaniowego zapewniających uzyskanie naturalnie brzmiących, nieintonowanych, syntetycznych samogłosek polskich,
- ustalenie jakościowych i ilościowych związków między typem funkcji opisującej impulsy pobudzenia krtaniowego oraz stosunkami czasowymi w obrębie tych impulsów a naturalnością brzmienia syntezowanych samogłosek,
- ustalenie wpływu niskoczęstotliwościowej, sinusoidalnej modulacji $F_0(t)$ oraz parametrów funkcji modulującej, dla których uzyskuje się poprawę naturalności brzmienia samogłosek syntetycznych.

Przeprowadzenie eksperymentów wstępnych było jednym ze szczegółowych celów pracy (rozdz.1.5.2. p-t 1).

2.2. Metoda generacji samogłosek syntetycznych

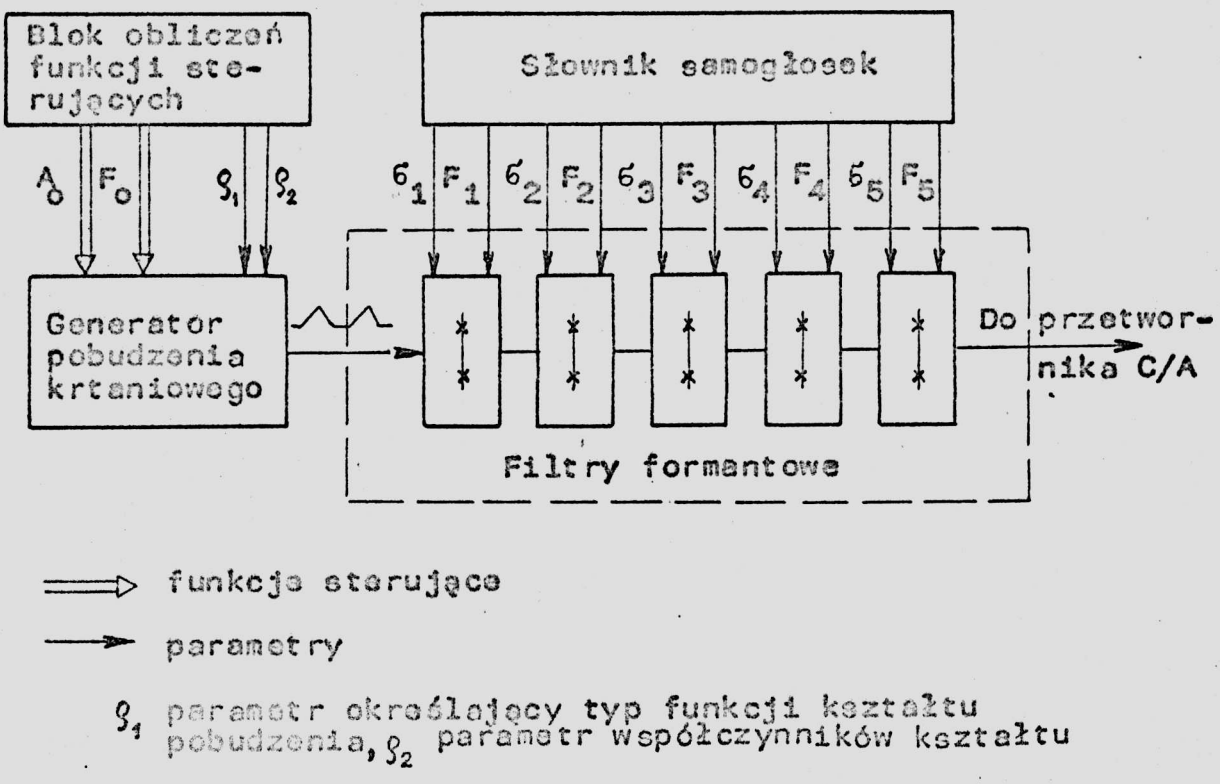
2.2.1. Model kanału głosowego

Samogłoski generowano z zastosowaniem cyfrowego, formantowego modelu kanału głosowego w układzie kaskadowym (rys.2.1.). Funkcję transmitancji kanału przybliżono za pomocą pięciu filtrów cyfrowych odpowiadających kolejnym formantom syntezowa-

---

[20] W eksperymentach wstępnych nie stosowano zbioru reguł realizacyjnych i stąd badane parametry określono jako "parametry pobudzenia krtaniowego" nie dokonując ich podziału na parametry dziedziny reguł i parametry syntezy (Def.6 i 7).

nych samogłosek. Zespolone wartości częstotliwości pary biegunów sprzężonych poszczególnych filtrów formantowych dobrano w oparciu o wyniki badań Majewskiego [28] i Kudeli [29,30] oraz na podstawie eksperymentów przeprowadzonych przez autora (tab.2.1.). Korekcję biegunów wyższego rzędu uzyskano wykorzystując zjawisko cyklicznego zwielokrotnienia widma w systemach z czasem dyskretnym [85] . W tym celu urojoną część zespolonej częstotliwości piątego filtru formantowego ustalono na 4500 Hz, (przy szęstotliwości Nyquista f_N = 5000 Hz). Procedurę syntezy przeprowadzono z zastosowaniem minikomputera Varian; Brüel-Kjaer 7504. Generowany sygnał przetworzono na postać analogową za pomocą rejestratora Brüel-Kjaer 7502 wyposażonego w przetwornik cyfrowo-analogowy. Interwał pomiędzy kolejnymi, obliczonymi wartościami samogłosek wynosił 10^{-4} s. ($f_p = 10^4$ Hz), a poziom kwantyzacji wartości chwilowych 8 bitów/próbkę.

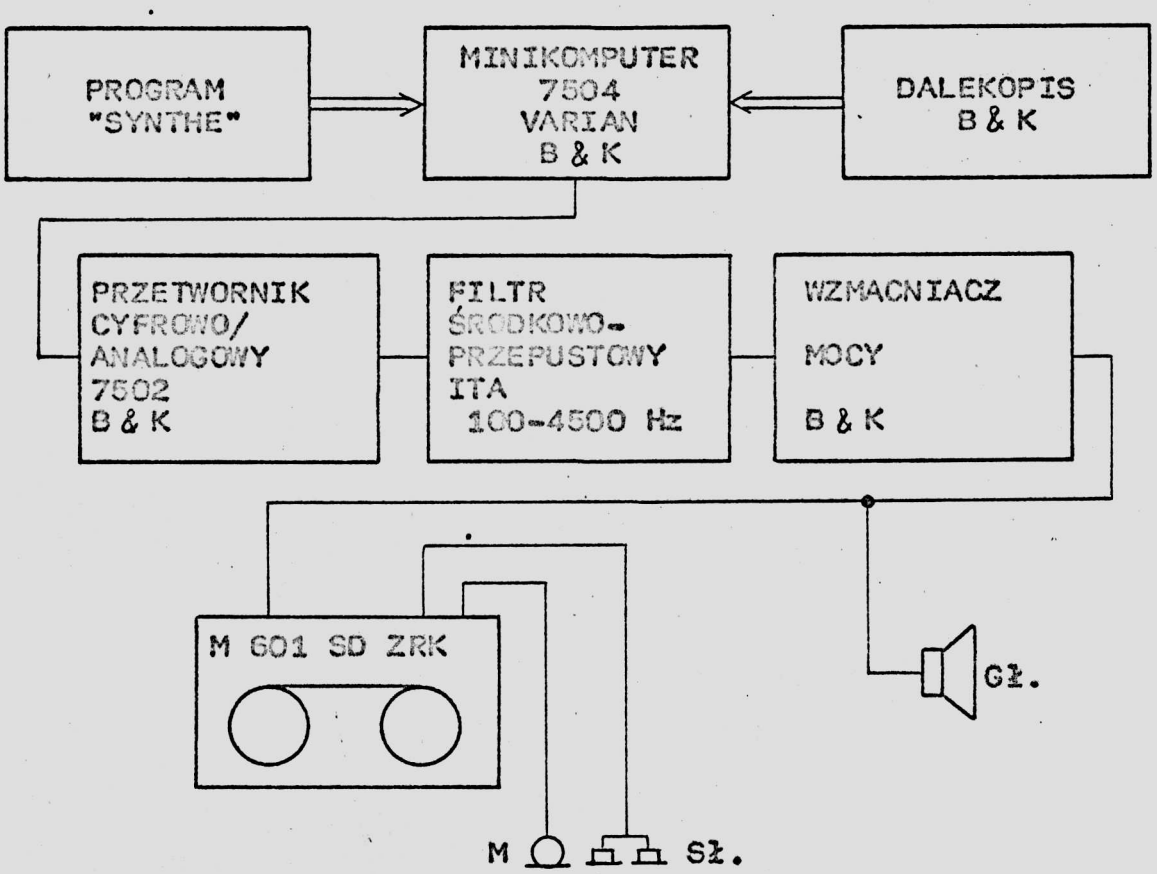


Rys.2.1.Schemat blokowy syntezatora samogłosek stosowanego w eksperymentach wstępnych.

Tabela 2.1. Zespolone częstotliwości formantowe samogłosek polskich.

Głoska	F1	F2	F3	F4	F5
i	30+ j230	38+ j3020	120+ j3250	115+ j3900	165+ j4500
y	22+ j380	63+ j2000	39+ j2800	67+ j3700	70+ j4500
o	19+ j525	48+ j1930	85+ j2560	162+ j3270	155+ j4500
a	29+ j770	36+ j1250	65+ j2460	88+ j3060	100+ j4500
o	27+ j570	33+ j 840	50+ j2800	68+ j3700	78+ j4500
u	35+ j320	25+ j 660	55+ j2660	58+ j3500	55+ j4500

Schemat układu syntezy samogłosek pokazano na rys.2.2.



Rys.2.2. Schemat blokowy układu syntezy samogłosek.

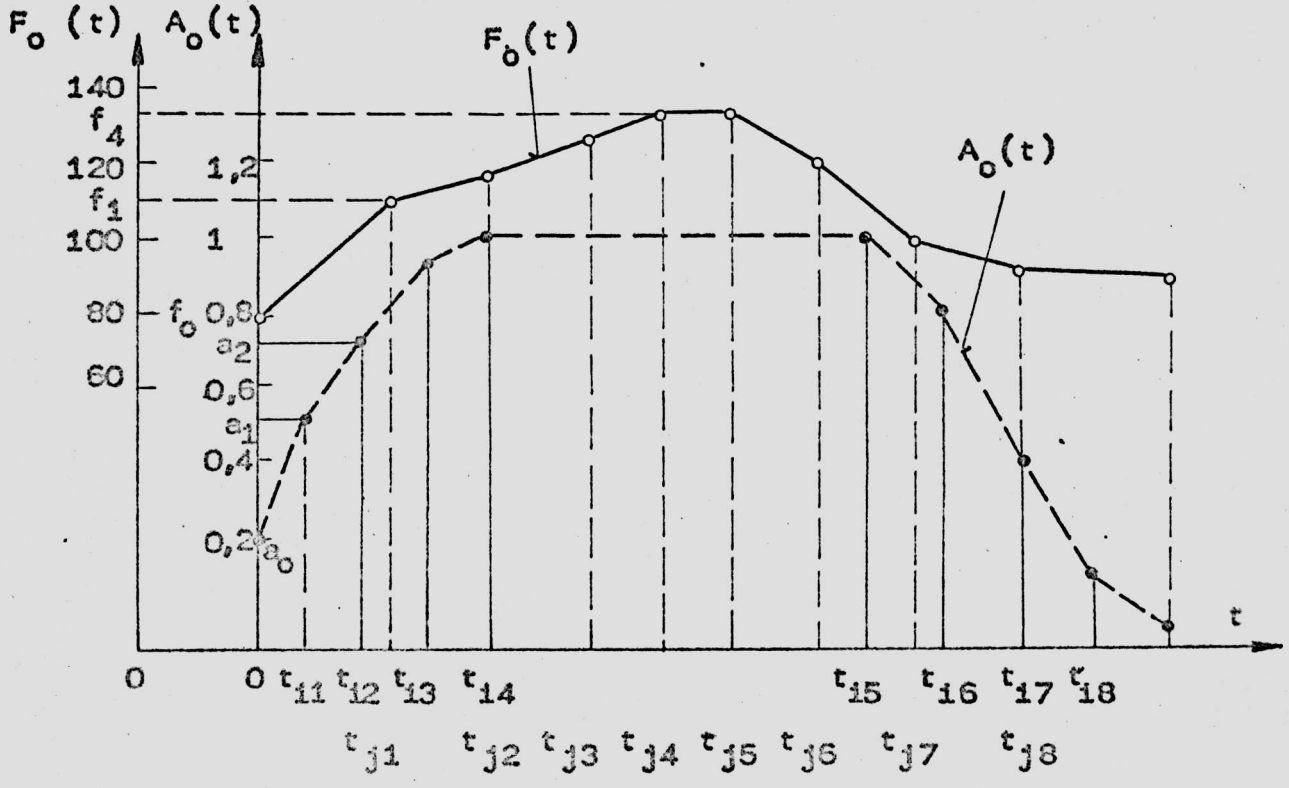
2.2.2. Funkcje sterujące parametrami pobudzenia

W eksperymentach wstępnych nie założono całościowego modelu generacji pobudzenia krtaniowego i zatem nie stosowano reguł tworzenia konturów $A_o(t)$ i $F_o(t)$ aproksymując je 10-węzłowymi funkcjami liniowo-odcinkowymi (rys.2.3):

$$A_o(t) = \frac{a_{i-1} - a_i}{t_{i-1} - t_i}(t - t_i) \qquad t_{i-1} < t < t_i \qquad i \in \{0,9\} \tag{2.1.}$$

$$F_o(t) = \frac{f_{j-1} - f_j}{t_{j-1} - t_j}(t - t_j) \qquad t_{j-1} < t < t_j \qquad j \in \{0,9\} \tag{2.2.}$$

W zależnościach (2.1.) i (2.2.) dla $i = j = 0$ $t_i = t_j = 0$, dla $i = j = 9$ $t_i = t_j = T_F$ gdzie T_F - czas trwania samogłoski syntetycznej.



Rys.2.3. Funkcje aproksymujące kontury $F_o(t)$ i $A_o(t)$.

Funkcję wolnozmiennej modulacji $F_0(t)$ opisano zależnością:

$$F_M(t) = A_D \sin (2\pi F_D t) \qquad t \in \{0, T_F\} \qquad (2.3.)$$

Po zastosowaniu (2.3.) w (2.2.) otrzymano regułę dewiacji:

$$F_0(t) := F_0(t) + A_D \sin (2\pi F_D t) \quad t \in \{0, T_F\} \qquad (2.4.)$$

gdzie:

A_D - amplituda dewiacji, F_D - częstotliwość dewiacji.

Badane typy funkcji opisujących impulsy pobudzenia krtaniowego (zwane dalej funkcjami kształtu) pokazano na rys.2.4.

2.3. Przygotowanie i ocena materiału eksperymentalnego

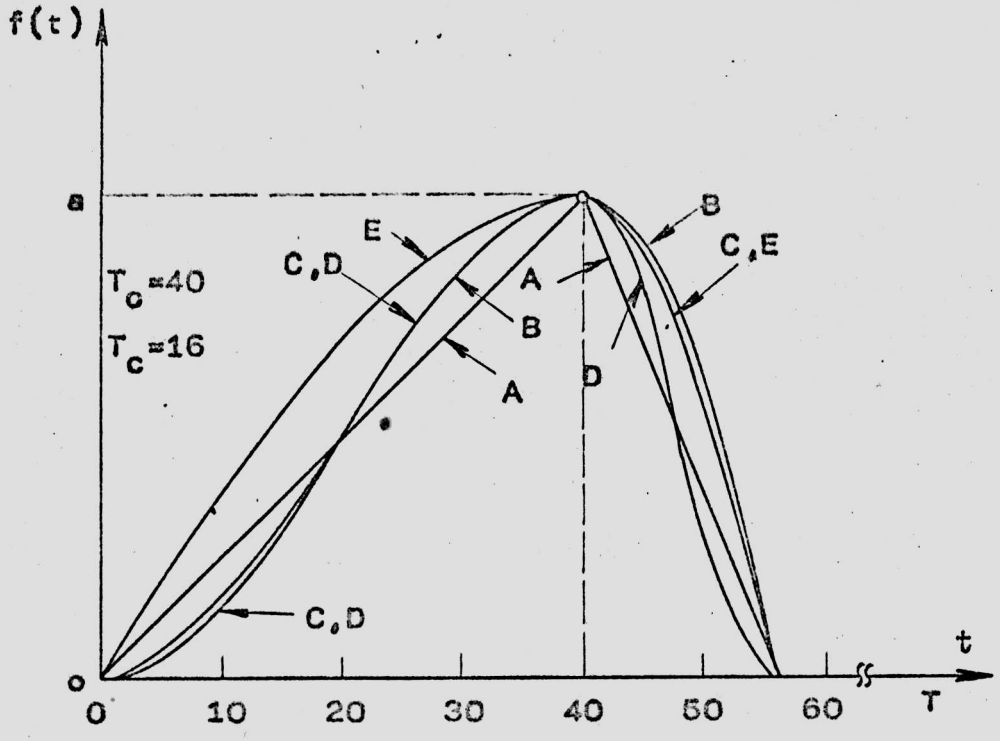
Samogłoski syntetyczne stanowiące w badaniach wstępnych materiał eksperymentalny rejestrowano na taśmie magnetofonowej (rys.2.2.), a następnie za pomocą profesjonalnego sprzętu nagłaśniającego prezentowano ekipie znajdującej się w studio odsłuchowym ITA[21]. Stosowany system nagłośnienia poprzez kolumnę głośnikową umożliwił pominięcie w badaniach czynników związanych ze stosunkami fazowymi w sygnale syntetycznym [76, 100, 101] . Kilkuosobowa ekipa odsłuchowa, wytrenowana w ocenie bodźców syntetycznych i znająca zagadnienie syntezy dźwięków mowy[22]., oceniała prezentowane w losowej kolejności samogłoski syntetyczne metodą testu punktowego w skali 0 - 10. Jako kryterium oceny stosowano naturalność brzmienia.

Każda głoska była prezentowana 3-krotnie w odstępie co 1 s, odstęp pomiędzy kolejnymi trójkami głosek wynosił 3 s. Ze względu na małą liczebność ekipy odsłuchowej ocenę każdego zbioru samogłosek powtarzano dwu lub trzykrotnie z kilkudniowym odstępem, przyjmując kryterium:

---

[21] Objętość studia: 88 m^3, czas pogłosu w paśmie 100 do 4500 Hz : 0,5 ± 10 % s.

[22] W skład ekipy wchodzili pracownicy oraz dyplomanci z Zespołu Akustyki Cybernetycznej ITA.



Rys. 2.4 Stosowane w eksperymentach funkcje opisujące kształt impulsów pobudzenia krtaniowego

Nr	Funkcja kształtu	$0 \leqslant t < T_o$	$T_o < t \leqslant T_o + T_c$
1	f_A	$a\frac{t}{T_o}$	$a\left[1-\frac{t-T_o}{T_c}\right]$
2	f_B	$a\left[3\left(\frac{t}{T_o}\right)^2 2\left(\frac{t}{T_o}\right)^3\right]$	$a\left[1-\left(\frac{t-T_o}{T_c}\right)^2\right]$
3	f_C	$\frac{a}{2}\left[1-\cos\frac{t}{T_o}\pi\right]$	$a\cos\left(\frac{t-T_o}{T_c}\right)\frac{\pi}{2}$
4	f_D	$\frac{a}{2}\left[1-\cos\frac{t}{T_o}\pi\right]$	$\frac{a}{2}\left[1+\cos\left(\frac{t-T_o}{T_c}\right)\frac{\pi}{2}\right]$
5	f_E	$a\sin\frac{t}{T_o}\frac{\pi}{2}$	$a\cos\left(\frac{t-T_o}{T_c}\right)\frac{\pi}{2}$

$$l = m \cdot n \geq 6 \qquad (2.5.)$$

l - liczba niezależnych eksperymentów (ocen), $l \in \{1, \ldots, L\}$
m - liczba słuchaczy, $m \in \{1, \ldots, M\}$
n - liczba powtórzeń, $n \in \{1, \ldots, N\}$

2.4. Program eksperymentów wstępnych

Badania wstępne obejmowały pięć eksperymentów (oznaczonych dalej jako EW1 do EW5):

EW1 - Dobór funkcji $A_0(t)$ sterującej amplitudą pobudzenia krtaniowego
EW2 - Dobór funkcji $F_0(t)$ sterującej częstotliwością pobudzenia krtaniowego
EW3 - Dobór parametrów kształtu (tzn. czasu narastania T_0 i opadania T_c) impulsu pobudzenia krtaniowego (rys.2.7.)
EW4 - Dobór funkcji kształtu impulsów pobudzenia krtaniowego
EW5 - Dodatkowe badania nad doborem parametrów kształtu dla funkcji uznanej na podstawie EW4 za optymalną.

Ze względu na rozpoznawczy charakter eksperymentów EW1 - 5 oraz ze względu na fakt opublikowania części wyników w pracach [33-35] eksperymenty opisano w sposób syntetyczny i zamieszczono tylko najistotniejsze wyniki[23].

2.5. Opis i wyniki eksperymentów wstępnych

2.5.1. Dobór funkcji $A_0(t)$ - (EW1)

W eksperymencie EW1 w oparciu o wyniki badań Rosenberga [75], Holmesa [76] oraz Zalewskiego i Myśleckiego [34] przyjęto:

---

[23] Por. również uwagi dotyczące założeń do redakcji pracy, zamieszczone w rozdz.1.6.

1. Funkcję kształtu impulsu: f_C (rys.2.4.)
2. Parametry kształtu: $t_o = 0.45$, $t_c = 0.16$ (rys.2.7.)

Dodatkowo założono:

3. $F_o(t) = \text{const} = 120$ Hz
4. Czas trwania samogłosek syntetycznych: $T_F = 300$ ms.

Dla podzbioru SbS trzech samogłosek:

$$SbS\ (k \in \{1,2,3\} : s_k) \tag{2.6.1.}$$

$$S\ (k \in \{1,\ldots,6\} : s_k) \tag{2.6.2.}$$

gdzie:

s_1 - samogłoska [a] s_4 - [e]
s_2 - samogłoska [o] s_5 - [i]
s_3 - samogłoska [u] s_6 - [ɨ]

badano zbiór A_o dwunastu funkcji sterujących amplitudą pobudzenia krtaniowego:

$$A_o\ [(i,j) \in \{1,2,3\} \times \{4,5,6\} \cup \{(1,1),(2,2),(3,3)\} : : A_{o_{i,j}}(t)] \tag{2.7.}$$

gdzie:

i - indeks w zbiorze T_{AO} czasów narastania $A_o(t)$
j - indeks w zbiorze T_{AC_c} czasów opadania $A_o(t)$

$$T_{AO}(t_{o_1}, t_{o_2}, t_{o_3}) = \{38, 48, 60\}\ [\text{ms}]$$

$$T_{AC}(t_{c_1}, \ldots, t_{c_6}) = \{38, 48, 60, 96, 128, 160\}\ [\text{ms}]$$

Otrzymany zbiór $\{p^k_{i,j}\}$ próbek syntetycznych odwzorowano poprzez pomiar subiektywny w zbiór $\{\bar{w}^k_{i,j}\}$ ocen punktowych z przedziału (0 ; 10)

$$p^k_{i,j} \xrightarrow{SUB} \bar{w}^k_{i,j} \tag{2.8.}$$

gdzie:

$$\bar{w}^k_{i,j} = \frac{1}{L} \sum_{l=1}^{L} w^{k,l}_{i,j} \tag{2.9.}$$

jest średnią oceną uzyskaną przez próbkę $p^k_{i,j}$ w zbiorze L niezależnych ocen.

Jako kryterium oceny funkcji $A_{o_{i,j}}(t)$ przyjęto:

$$\bar{w}_{i,j} = \frac{1}{K} \sum_{k=1}^{K} \bar{w}^{k}_{i,j} , \quad K = 3 \qquad (2.10.)$$

Uwzględniając (2.9.) w (2.10.) otrzymujemy:

$$\bar{w}_{i,j} = \frac{1}{KL} \sum_{k=1}^{K} \sum_{l=1}^{L} w^{k,l}_{i,j} , \quad K = 3, \; L = 6 \qquad (2.11.)$$

gdzie $\bar{w}_{i,j}$ - średnia ocena w zbiorze ocen niezależnych oraz w zbiorze próbek syntetycznych.

Badane funkcje $A_{o_{i,j}}(t) \subset A_o$ pokazano na rys.2.5., a uzyskane przez nie oceny podano w tab.2.2.

Tabela 2.2. Oceny $\bar{w}_{i,j}$ próbek syntetycznych dla różnych funkcji $A_o(t)$ sterujących amplitudą pobudzenia krtaniowego.

T_{A_c} [ms]

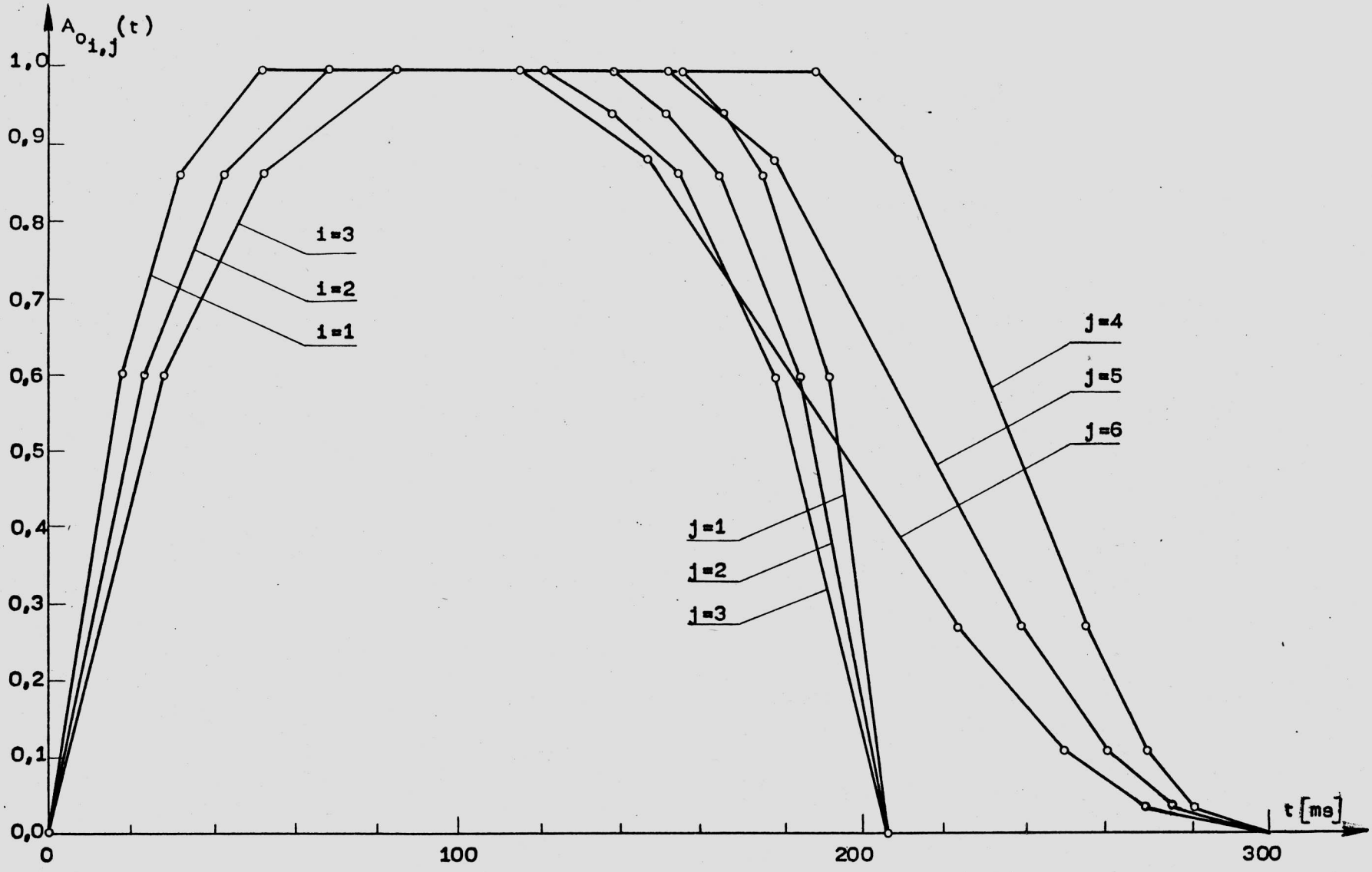
t_{o_i} \ t_{c_j}	38	48	60	96	128	160
38	3.7	-	-	9.1	5.4	3.9
48	-	5.5	-	7.1	5.5	5.2
60	-	-	4.2	5.5	4.6	3.7

T_{A_o} [ms]

2.5.2. Dobór funkcji $F_o(t)$ - (EW2)

W eksperymencie EW2 jako dane przyjęto:

1. Funkcję $A_{o_{1,4}}(t)$, która w eksperymencie EW1 uzyskała najwyższą ocenę.
2. Funkcję i parametry kształtu impulsu pobudzenia oraz czas trwania samogłosek jak w EW1 (p-ty 1,2,4).



Rys. 2.5 Funkcje $A_o(t)$ badane w eksperymencie EW1

Badano zbiór F_o trzynastu funkcji $F_{o_i}(t)$ skorelowanych z funkcją $A_{o_{1,4}}(t)$. Przyjęcie klasy funkcji $F_o(t)$ jednorodnej pod względem typu kształtu wynikało ze stwierdzonego w mowie naturalnej zjawiska silnego skorelowania $F_o(t)$ z $A_o(t)$ [50, 65, 71] oraz z negatywnego wyniku badań przeprowadzonych przez Zalewskiego i Myśleckiego [34] dla funkcji $F_o(t)$ o kształcie nie skorelowanym z funkcją $A_o(t)$.

Zbiór F_o określony zależnością:

$$F_o \quad [i \in \{1, \ldots, 13\} : F_{o_i}(t)] \tag{2.12.}$$

badano dla sześciu samogłosek sylabicznych tworzących zbiór S (2.6.2.):

$$S \ (k \in \{1, \ldots, 6\} : s_k) \tag{2.13.}$$

W tab.2.3. podano oceny $\{\bar{w}_i\}$ otrzymane przez poszczególne próbki syntetyczne ze zbioru $\{p_i^k\}$, gdzie:

$$p_i^k \xrightarrow{SUB} \bar{w}_i^k \tag{2.14.}$$

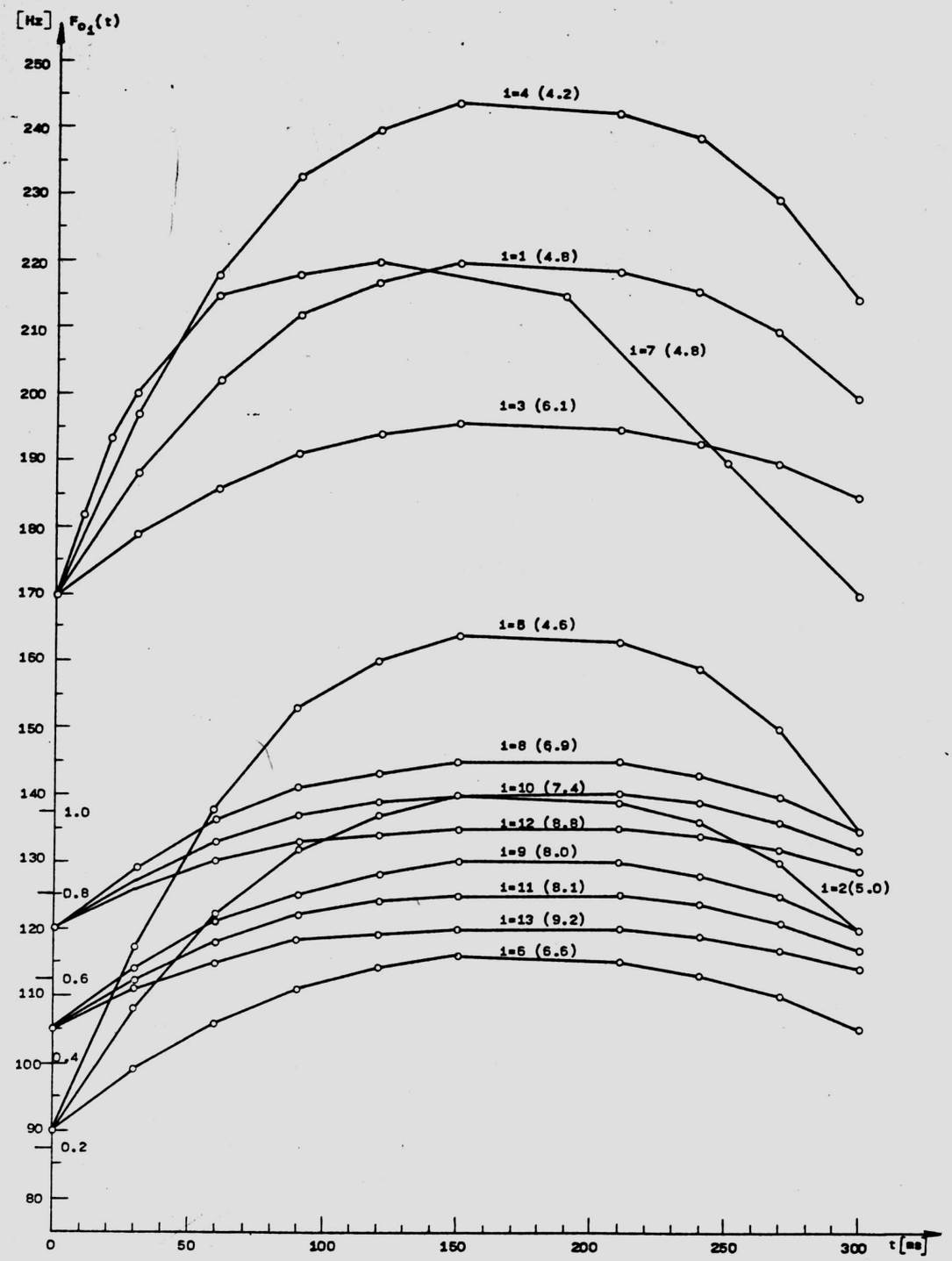
$$\bar{w}_i^k = \frac{1}{L} \sum_{l=1}^{L} \bar{w}_i^{k,l} \qquad L = 8 \tag{2.15.}$$

$$\bar{w}_i = \frac{1}{K} \sum_{k=1}^{K} \bar{w}_i^k \qquad K = 6 \tag{2.16.}$$

Zbiór F_o badanych funkcji $F_{o_i}(t)$, i=1, ...,13 pokazano na rys.2.6.

Tabela 2.3. Oceny $\bar{w}_i$ próbek syntetycznych generowanych z zastosowaniem różnych funkcji $F_o(t)$ sterujących częstotliwością pobudzenia krtaniowego.

$F_{o_i}(t)$	i=1	i=2	i=3	i=4	i=5	i=6	i=7	i=8	i=9	i=10	i=11	i=12	i=13
$\bar{w}_i$	4.8	5.0	6.1	4.2	4.6	6.6	4.8	6.9	8.0	7.4	8.1	8.8	9.2



Rys. 2.6 Funkcje $F_o(t)$ badane w eksperymencie EW2 (obok funkcji podano w nawiasach oceny W_i otrzymane dla danej funkcji

2.5.3. Dobór parametrów kształtu impulsów pobudzenia krtaniowego (EW3)

W eksperymencie EW3 jako dane przyjęto:

1. Funkcję $F_{o_{13}}(t)$, która w eksperymencie EW2 uzyskała najwyższą ocenę.
2. Funkcję $A_o(t)$, funkcję kształtu impulsu pobudzenia oraz czas trwania samogłosek jak w EW2.

Badano zbiór T_G dwunastu impulsów pobudzenia krtaniowego różniących się względnym czasem narastania $t_o = T_o/T$ oraz względnym czasem opadania $t_c = T_c/T$ (rys.2.7.). Zbiór T_G określa zależność:

$$T_G\left[\bigvee_{i=j} i \in \{1, \ldots, 12\} \;,\; j \in \{1,\ldots 12\} : f_c(t, t_{o_i}, t_{c_j})\right] \qquad (2.17.)$$

gdzie:

i - indeks w zbiorze T_o czasów narastania impulsu pobudzenia

j - indeks w zbiorze T_c czasów opadania impulsu pobudzenia

$T_o(t_{o_1}, \ldots, t_{o_{12}})$

$T_c(t_{c_1}, \ldots, t_{c_{12}})$

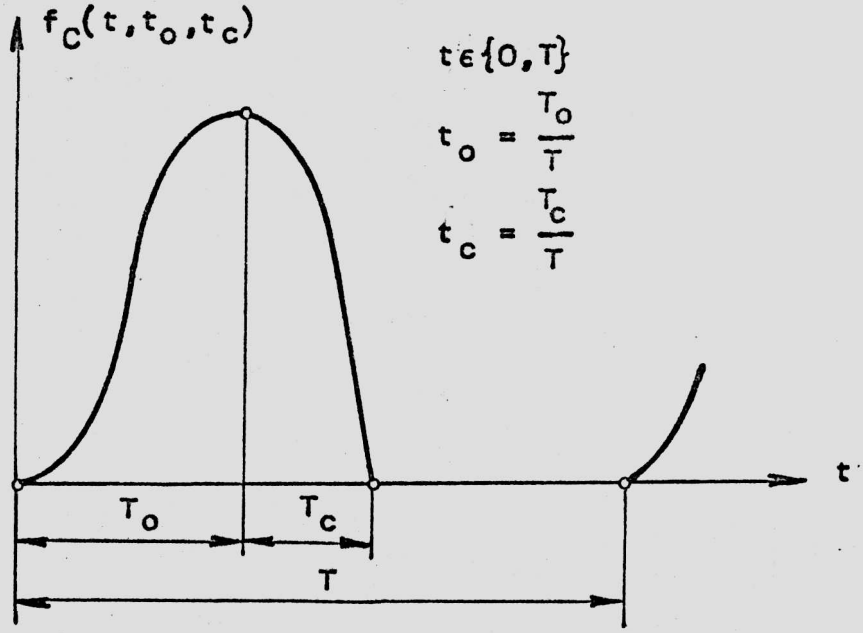
Zbiory T_o i T_c podano w tab.2.7.

Materiałem eksperymentalnym był podzbiór SbS trzech samogłosek syntetycznych:

$$SbS\ (k \in \{1,2,4\} : s_k) \qquad (2.18.)$$

Zgodnie z (2.6.2.) s_1 - samogłoska [a]
s_2 - samogłoska [o]
s_4 - samogłoska [e]

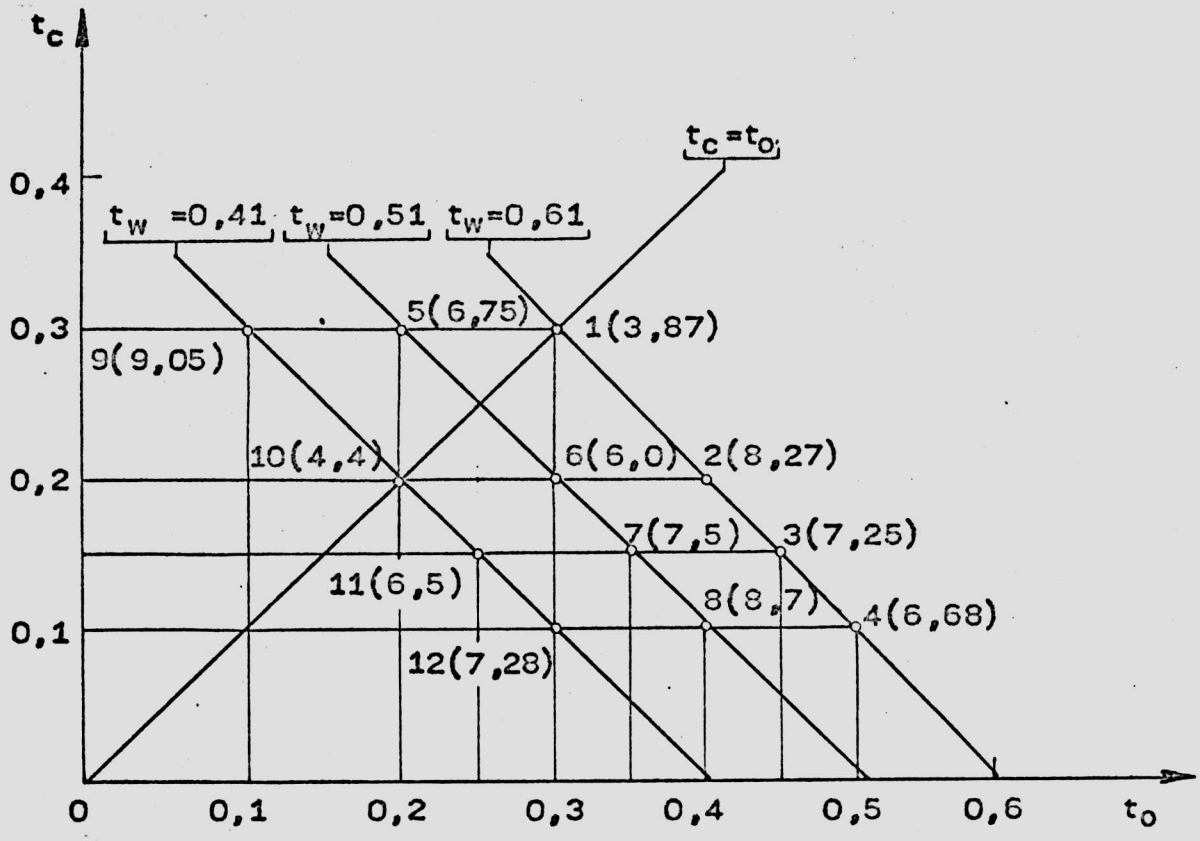
Oceny naturalności $\bar{w}_{i,j}$, (2.14. - 2.16.) otrzymane w wyniku testu badań subiektywnych przeprowadzonego dla zbioru $\{p^k_{i,j}\}$ próbek wygenerowanych w eksperymencie EW3 dla dwunastu par parametrów (t_o, t_c) podano w tab.2.7., natomiast rozkład badanych wartości parametrów t_o i t_c na płaszczyźnie t_o X t_c pokazano na rys.2.8.



Rys.2.7. Parametry kształtu impulsu pobudzenia krtaniowego.

i=j	$\bar{w}_{i,j}$	t_{o_i}	t_{c_j}
1	3.87	0.31	0.30
2	8.27	0.41	0.20
3	7.25	0.46	0.15
4	6.68	0.51	0.10
5	6.75	0.21	0.30
6	6.0	0.31	0.20
7	7.5	0.36	0.15
8	8.7	0.41	0.10
9	9.05	0.11	0.30
10	4.4	0.21	0.20
11	6.5	0.26	0.15
12	7.28	0.31	0.10

Tab.2.7. Oceny $\bar{w}_{i,j}$ naturalności brzmienia samogłosek syntetycznych dla 12 kombinacji (t_o, t_c) badanych w EW3.



Rys.2.8. Rozkład wartości 12 kombinacji parametrów (t_o, t_c) impulsu pobudzenia badanych w eksperymencie EW3. (W nawiasach obok numeru podano ocenę $\bar{w}_{i,j}$ uzyskaną dla danej kombinacji. Wkreślono również proste $t_o + t_c = t_w$, $t_w \in \{0,41, 0.51, 0.61\}$ oraz $t_o = t_c$).

2.5.4. Dobór funkcji kształtu impulsów pobudzenia krtaniowego (EW4)

W eksperymencie EW4 jako dane przyjęto:

1. Funkcję $A_o(t)$, $F_o(t)$ oraz czas trwania samogłosek jak w EW3.
2. Dwie kombinacje wartości parametrów (t_o, t_c), które w EW3 uzyskały najwyższe oceny $\bar{w}_{i,j}$ - (0.11 ; 0.30) oraz (0.41 ; 0.1).

Dla każdej z podanych w p-cie 2 kombinacji (t_o, t_c) badano 10 różnych funkcji kształtu impulsu pobudzenia krtaniowego. Zbiór badanych funkcji otrzymano ze zbioru funkcji pokazanych na rys.2.4. przez zastosowanie różnych kombinacji funkcji opisujących zbocze narastające i opadające impulsu pobudzenia. Materiałem eksperymentalnym był podzbiór SbS trzech samogłosek syntetycznych opisany zależnością (2.18.). Badane typy funkcji kształtu oraz otrzymane dla nich oceny naturalności brzmienia $\bar{w}_{i,j}$ (2.14. - 2.16.) podano w tab.2.8., gdzie $f_{A,A}$, $f_{B,C}$, $f_{C,E}$, ... oznaczają funkcję kształtu powstałe odpowiednio z kombinacji funkcji narastania f_A i opadania f_A, funkcji narastania f_B i opadania f_C, itd. (por.rys.2.4.).

Tabela 2.8. Oceny naturalności brzmienia $\bar{w}_{i,j}$ samogłosek syntetycznych dla 10-ciu funkcji kształtu impulsów pobudzenia badanych w EW4.

Lp.	Typ f-cji kształtu	$\bar{w}_{i,j}$ [*]	
		(t_o ; t_c) = (0.41 ; 0.1)	(t_o ; t_c) = (0.11 ; 0.30)
1	$f_{A,A}$	6.23 (9)	6.06 (8)
2	$f_{B,B}$	7.74 (5)	8.36 (2)
3	$f_{B,C}$	8.1 (2)	8.74 (1)
4	$f_{B,D}$	6.65 (7)	6.93 (6)
5	$f_{C,B}$	7.85 (3)	8.08 (4)
6	$f_{C,C}$	8.8 (1)	8.25 (3)
7	$f_{C,D}$	7.8 (4)	7.16 (5)
8	$f_{E,B}$	6.79 (6)	5.94 (9)
9	$f_{E,C}$	5.9 (10)	5.91 (10)
10	$f_{E,D}$	6.5 (8)	6.6 (7)

[*] Obok oceny $\bar{w}_{i,j}$ podano w nawiasie miejsce danej funkcji w szeregu rangowym naturalności, utworzonym według malejącej kolejności $\bar{w}_{i,j}$.

2.5.5. Dodatkowe badania nad doborem parametrów kształtu (t_o , t_c) impulsów pobudzenia krtaniowego (EW5)

W eksperymencie EW5 jako dane przyjęto:

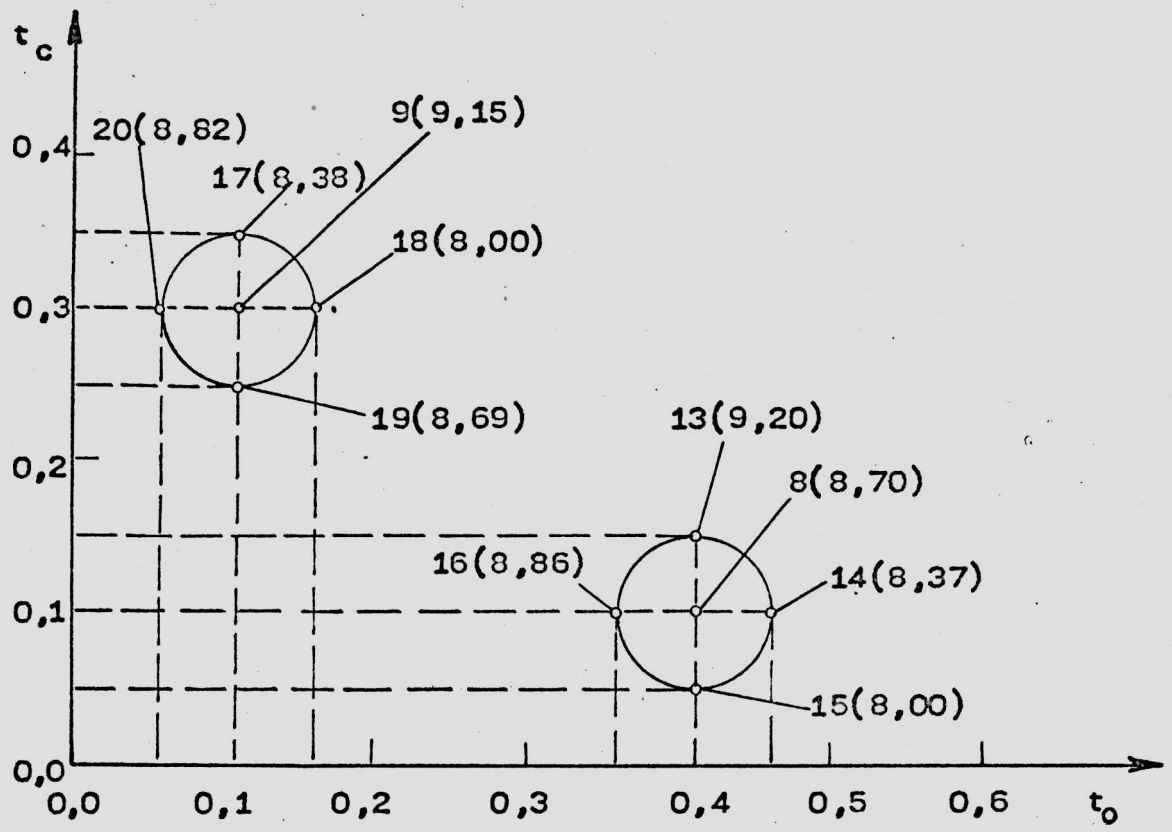
1. Funkcję $A_o(t)$ i $F_o(t)$ oraz czas trwania samogłosek syntetycznych jak w EW3.
2. Funkcję kształtu $f_{C,C} = f_C(t)$. Tę funkcję na podstawie wyników eksperymentu EW4 uznano za optymalną funkcję kształtu pobudzenia.

Badano 8-elementowy zbiór T_G (zależność 2.17. dla $i=j \in \{8,9,13, \ldots,20\}$) impulsów pobudzenia krtaniowego różniących się parametrami (t_o ; t_c). Wartości parametrów (t_o ; t_c) leżały w otoczeniu kombinacji (t_{o_8} ; t_{c_8}) oraz (t_{o_9} ; t_{c_9}), dla których w EW3 uzyskano najbardziej naturalnie brzmiące samogłoski syntetyczne. Rozkład badanych kombinacji (t_o ; t_c) na płaszczyźnie t_o X t_c pokazano na rys.2.9.

Materiał eksperymentalny stanowił podzbiór SbS trzech samogłosek syntetycznych (2.18.). Liczba niezależnych ocen wynosiła L = 8. Wyniki eksperymentu EW5 w postaci uśrednionych ocen $\bar{w}_{i,j}$ (2.14. - 2.16.) pokazano w tab.2.8.

i=j	$\bar{w}_{i,j}$	t_o	t_c
8	8.70	0.41	0.10
13	9.20	0.41	0.15
14	8.37	0.46	0.10
15	8.00	0.41	0.05
16	8.86	0.36	0.10
9	9.15	0.11	0.30
17	8.38	0.11	0.35
18	8.00	0.16	0.30
19	8.69	0.11	0.25
20	8.82	0.06	0.30

Tabela 2.8. Oceny $\bar{w}_{i,j}$ naturalności brzmienia dla 8 kombinacji (t_{o_i} ; t_{c_j}) badanych w eksperymencie EW5.



Rys.2.9. Rozkład wartości 8 kombinacji parametrów (t_o ; t_c) pobudzenia krtaniowego badanych w eksperymencie EW5 (obok numeru kombinacji podano w nawiasach ocenę $\bar{w}_{i,j}$).

2.5.6. Podsumowanie - optymalne funkcje sterujące oraz parametry pobudzenia krtaniowego w procesie syntezy samogłosek polskich

W oparciu o wyniki eksperymentów EW1 - 5 ustalono:

1. Optymalną[24)] funkcję $A_o(t)_{opt}$ sterującą amplitudą pobudzenia krtaniowego - tab.2.9.
2. Optymalną[24)] funkcję $F_o(t)_{opt}$ sterującą częstotliwością pobudzenia krtaniowego - tab.2.10.

[24)] W badanym zbiorze.

3. Optymalną[24)] funkcję kształtu oraz dwie optymalne dla niej kombinacje parametrów (t_o ; t_c) impulsów pobudzenia krtaniowego - tab.2.11.

Tabela 2.9. Optymalna funkcja $A_o(t)_{opt}$ - współrzędne 10-ciu węzłów funkcji aproksymującej.

Nr węzła	1	2	3	4	5	6	7	8	9	10
t [ms]	0	17	31	51	189	208	254	270	282	300
$A_o(t)$	0.000	0.600	0.864	1.000	1.000	0.884	0.276	0.112	0.040	0.000

Tabela 2.10. Optymalna funkcja $F_o(t)_{opt}$ - współrzędne 10-ciu węzłów liniowej funkcji aproksymującej.

Nr węzła	1	2	3	4	5	6	7	8	9	10
t [ms]	0	30	60	90	120	150	210	240	270	300
$F_o(t)_{opt}$ [Hz]	105	111	115	118	119	120	120	119	117	114

Funkcja kształtu	Parametry (t_o ; t_c)	
	1	2
funkcja f_c z rys.2.4.	t_o = 0.41 t_c = 0.15	t_o = 0.11 t_c = 0.30

Tabela 2.11. Optymalna funkcja kształtu oraz optymalne dla niej wartości parametrów (t_o ; t_c) impulsów pobudzenia krtaniowego.

[24)] W badanym zbiorze.

2.5.7. Uwagi końcowe

Wyników badań wstępnych nie analizowano pod kątem ich wiarygodności statystycznej, gdyż zarówno stosowana metoda (test punktowy), nieformalny charakter odsłuchów jak i założony cel badań implikowały jakościową a nie ilościową analizę wyników[25], które mimo tego stanowią podstawę do dokonania pewnych ustaleń i uogólnień:

1. Podane w rozdz.2.5.6. optymalne funkcje sterujące amplitudą i częstotliwością pobudzenia krtaniowego oraz funkcja i parametry kształtu impulsów pobudzenia zapewniały, w ocenie osób biorących udział w odsłuchach, uzyskanie naturalnie brzmiących samogłosek syntetycznych. (Warto zwrócić uwagę, że w końcowym eksperymencie EW5 średnie oceny były zawarte w przedziale (8; 9.2) , tzn. w górnym zakresie stosowanej skali ocen punktowych). Zatem podane funkcje i parametry mogą być wykorzystane w badaniach z zastosowaniem samogłosek syntetycznych, (np. badania zjawisk percepcyjnych).

2. Wyniki badań nad doborem funkcji $A_0(t)$ - eksperyment EW1 - wykazały, że są preferowane funkcje z nieciągłą pochodną na początku i z ciągłą pochodną na końcu (tego typu funkcje zajęły 2 pierwsze miejsca - funkcja $A_{o_{1,4}}(t)$ i $A_{o_{2,4}}(t)$, por.rys.2.5).

3. Preferowane są funkcje $F_0(t)$ skorelowane z $A_0(t)$ o stosunkowo niedużej różnicy między $F_{o_{max}}$ i $F_{o_{min}}$ wynoszącej 10 - 15 Hz ($F_{o_{13}}(t)$, $F_{o_{12}}(t)$, $F_{o_{11}}(t)$ - rys.2.6., tab.2.3.).

4. Preferowane są funkcje $F_0(t)$ o średniej, bezwzględnej wartości częstotliwości podstawowej wynoszącej ok. 120 Hz. Pierwsze cztery miejsca zajęły funkcje spełniające ten warunek: $F_{o_{13}}(t)$, $F_{o_{12}}(t)$, $F_{o_{11}}(t)$ i $F_{o_9}(t)$ - rys.2.6., tab. 2.3.

---

[25] Podobne podejście metodologiczne do analizy wyników badań wstępnych można znaleźć w licznych pracach dotyczących oceny naturalności (jakości) mowy syntetycznej [31, 60, 77, 82, 102]

5. Wyniki eksperymentów EW3 do EW5 wskazują, że w odniesieniu do kształtu impulsu pobudzenia krtaniowego:

a) Preferowane są funkcje z ciągłą pochodną na początku i z nieciągłą pochodną na końcu impulsu pobudzenia. W eksperymencie EW4 najwyżej oceniono próbki generowane z zastosowaniem tego typu funkcji: $f_{C,C}$, $f_{B,C}$, $f_{B,B}$, $f_{C,B}$ - rys.2.4., tab.2.8. Podobne wyniki otrzymał Rosenberg dla syntetycznych fraz mowy angielskiej [75] .

b) Na płaszczyźnie parametrów kształtu impulsów pobudzenia krtaniowego t_o X t_c występują dwa maksima subiektywnie ocenionej naturalności brzmienia samogłosek syntetycznych, umiejscowione w przybliżeniu symetrycznie względem prostej $t_c = t_c$ (rys.2.8., tab.2.8. oraz tab.2.11). Rosenberg w swoich badaniach [75] uzyskał tylko jedno maksimum ($t_o = 0.40$; $t_c = 0.16$) o prawie identycznym położeniu na płaszczyźnie t_o X t_c jak maksimum nr 1 uzyskane w niniejszej pracy - tab.2.11, co wynika z różnej techniki odsłuchów. Rosenberg stosował w odsłuchach słuchawki nagłowne, więc na ocenę naturalności miały wpływ stosunki fazowe. W niniejszej pracy stosowano odsłuch w polu rozproszonym, zatem stosunki fazowe nie odgrywały roli (por. rozdz.2.3.). Uwzględniając dodatkowo fakt podobieństwa widm amplitudowych impulsów o parach współczynników kształtu (t_o ; t_c) symetrycznie rozłożonych względem prostej $t_o = t_c$ (rys.2.7., rys.2.8.), uzyskanie w eksperymentach wstępnych dwóch wymienionych maksimum jest uzasadnione i świadczy o prawidłowości stosowanej metody oceny subiektywnej.

c) Najniższe oceny uzyskały próbki syntetyczne z symetrycznymi impulsami pobudzenia: $f_c(t, t_{o_1}, t_{c_1})$ i $f_c(t, t_{o_{10}}, t_{c_{10}})$ gdzie $t_{o_1} = t_{c_1} = 0.3$, $t_{o_{10}} = t_{c_{10}} = 0.2$. (tab.2.7.). Te wyniki są zgodne zarówno z wynikami eksperymentów Rosenberga [75] , jak i teoretyczną analizą przeprowadzoną przez Flanagana [59, rozdz.6.2.4] .

Dane zamieszczone w rozdz.2.5.6. oraz uwagi z niniejszego rozdziału, stanowiące podsumowanie i dyskusję wyników eksperymentów wstępnych, wskazują na osiągnięcie założonego celu tych badań. Jedynie w odniesieniu do sinusoidalnej dewiacji F_0 nie uzyskano konkretnych rezultatów, gdyż stosowana metoda oceny okazała się w tym przypadku za mało dokładna. W konsekwencji, w pracy nie zamieszczono wyników tych eksperymentów, natomiast celowe jest podanie dwóch stwierdzonych zjawisk:

1. Wpływ sinusoidalnej, addytywnej dewiacji F_0 (2.3.) jest zauważalny nawet dla bardzo małych wartości $A_D < 1$ Hz.
2. Zwiększenie częstotliwości dewiacji powyżej 15 Hz ($F_D > 15$ Hz) wyraźnie pogarsza naturalność samogłosek syntetycznych.

3. DOBÓR PODZBIORU REGUŁ REALIZACYJNYCH GENERUJACYCH FUNKCJE CZASOWE STERUJACE POBUDZENIEM KRTANIOWYM W PROCESIE SYNTEZY FRAZ MOWY POLSKIEJ

3.1. Wprowadzenie

W procedurze doboru omawianego podzbioru reguł realizacyjnych (Def.11.) uwzględniono:

1. Wnioski wynikające z teoretycznej analizy roli pobudzenia krtaniowego w syntezie mowy na poziomie fraz (rozdz.1.4.).
2. Założenia 1 - 6 (rozdz.1.4.).
3. Dokonany w rozdz.1.4.4. wybór czynników nieinterpretowalnych na płaszczyźnie językowej a mających wpływ na parametry pobudzenia krtaniowego.
4. Wyniki eksperymentów wstępnych (rozdz.2.).

W rozdziale 3. podano tylko reguły i dokonano podziału rozważanego zbioru parametrów na parametry dziedziny reguł realizacyjnych oraz na parametry syntezy. Ustalenia obszaru zmienności parametrów dziedziny reguł dokonano w rozdziale dotyczącym eksperymentów zasadniczych (rozdz.6).

3.2. Reguły

RF_1 - reguła generacji podstawowego konturu amplitudowego $A_o(t)$ - rys.3.1.

$$A_o(t) \cdot \begin{cases} H + (1-H)\sin\left(\frac{\pi}{2T_N}t\right), & 0 < t \leq T_N \\ 1, & T_N < t \leq T_F - T_o \\ \frac{h_{i-1} - h_i}{t_{i-1} - t_i}(t - t_i) + h_i, & t_{i-1} < t < t_i \\ & \text{dla } i = 1,2,3 \end{cases} \qquad (3.1.)$$

gdzie: $h_o = 1$, $h_1 = H_3$, $h_2 = H_2$, $h_3 = H$,

$t_o = T_F - T_o$, $t_1 = T_F - T_o + T_{o1}$,

$t_2 = T_F - T_o + T_{o1} + T_{o2}$, $t_3 = T_F$,

T_F - czas trwania frazy

T_o, T_F, T_{o2}, H, H_2, H_3 pokazano na rys.3.1.

RF_2 - reguła generacji podstawowego konturu częstotliwościowego $F_o(t)$ - rys.3.2.

$$F_o(t) = FO \frac{A_o(t) + a}{1 + a} \qquad (3.2.)$$

gdzie: FO - ustalona wartość częstotliwości podstawowej

a - współczynnik rzeczywisty, dodatni

RF_3 - reguła niskoczęstotliwościowej dewiacji F_o - rys.3.3.

$$F_o(t) := F_o(t) + A_D \sin (2\pi F_D t) \qquad (3.3.)$$

gdzie: A_D - amplituda dewiacji

F_D - częstotliwość dewiacji

RF_4 - reguła tworzenia akcentowego wariantu podstawowego konturu amplitudowego - rys.3.4.

RF_5 - reguła tworzenia akcentowego wariantu podstawowego konturu częstotliwościowego

$$A_o(t) := A_o(t) + 0.5\, V(P_s) \cdot A_a (1 - \cos 2\pi F_p t), \quad 0 \leq t \leq T_F \qquad (3.4.)$$

$$F_o(t) := F_o(t) + 0.5\, V(P_s) \cdot A_f (1 - \cos 2\pi F_p t), \quad 0 \leq t \leq T_F \qquad (3.5.)$$

gdzie: P_s - znacznik akcentu

$$P_s = \begin{cases} 1 & \text{dla sylaby akcentowanej } [+P_s] \\ 0 & \text{dla sylaby nieakcentowanej } [-P_s] \end{cases}$$

$$V(P_s = 1) = \begin{cases} 0, & 0 \leq t \leq T_s - 0.5\, T_p \text{ i } T_s + 0.5\, T_p \leq t \leq T_F \\ 1, & T_s - 0.5\, T_p < t < T_s + 0.5\, T_p \end{cases}$$

$$V(P_s = 0) = 0, \quad 0 \leq t \leq T_F$$

W zależnościach (3.4.) i (3.5.):

A_a, A_f - maksymalne podbicie konturu odpowiednio amplitudowego i częstotliwościowego w obrębie segmentu akcentowanego ($T_s - 0.5\,T_p \leqslant t \leqslant T_s + 0.5\,T_p$),

T_s - moment czasowy wystąpienia maksimum podbicia w $A_o(t)$ i $F_o(t)$ - środek samogłoski w sylabie akcentowanej,

$T_p = 1/F_p$ - czas trwania segmentu akcentowanego

$T_p = q \cdot t(v_i)$, q - współczynnik rzeczywisty, dodatni, $t(v_i)$ - czas trwania samogłoski v_i w sylabie akcentowanej.

RF_6 - <u>reguła tworzenia intonacyjnego wariantu podstawowego konturu częstotliwościowego - rys.3.6.</u>

$$F_{oc}(t) \overset{[+\alpha_r]}{:=} \begin{cases} F_o(t) & , 0 \leqslant t \leqslant T_F - T_r \\ F_o(t) \cdot F_{k,\alpha_r}(t) & , T_F - T_{\alpha_r} < t \leqslant T_F \end{cases} \qquad (3.6.)$$

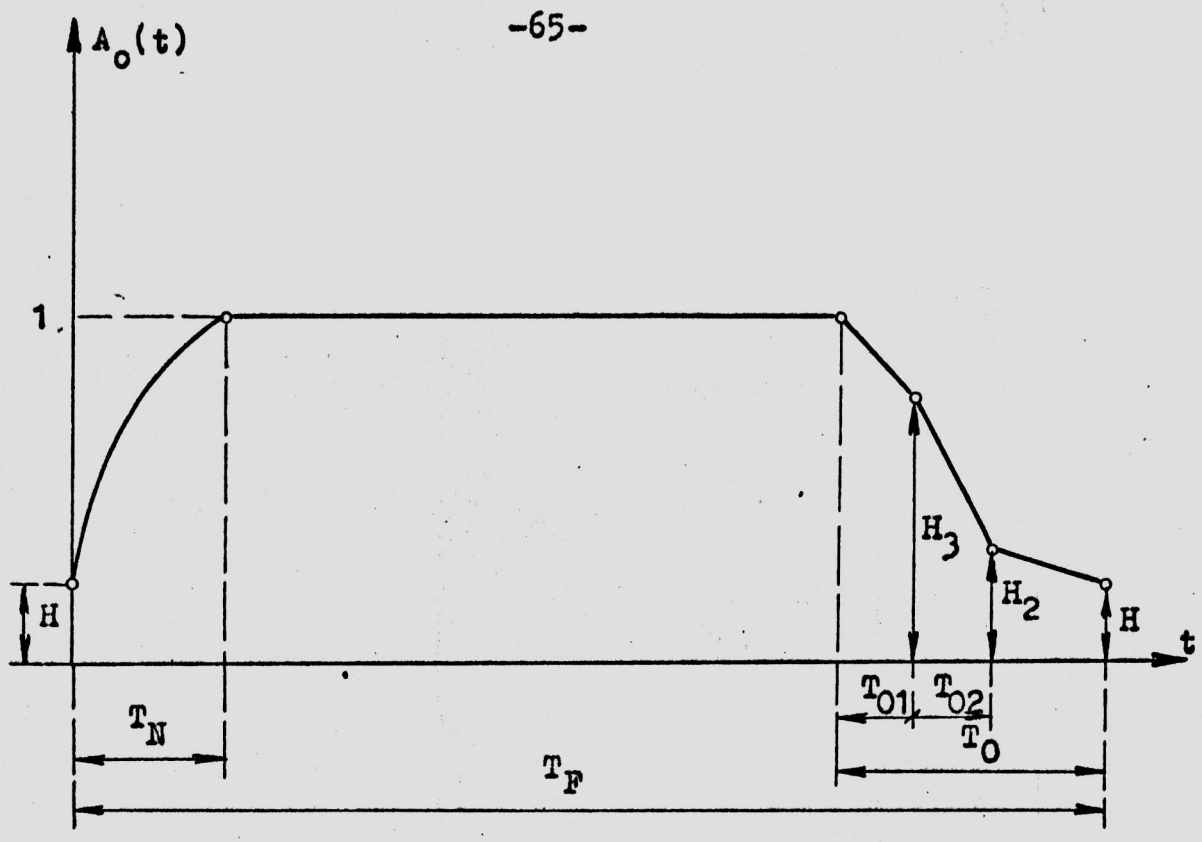
gdzie: $F_{oc}(t)$ - intonacyjny wariant podstawowego konturu częstotliwościowego

$F_{k,\alpha_r}(t)$ - r-ta funkcja intonacyjna aproksymowana 9-cio węzłową funkcją liniowo-odcinkową, (rys.3.5.)

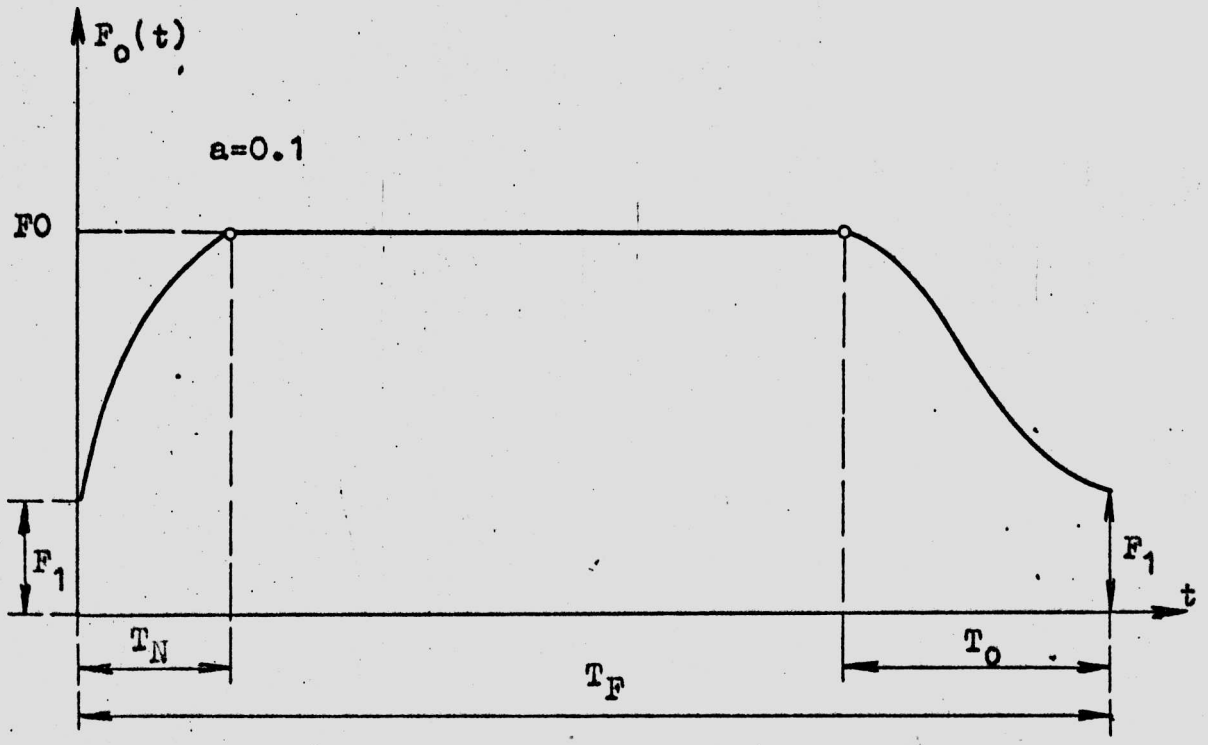
T_{α_r} - czas trwania intonowanego segmentu dla r-tej funkcji intonacyjnej

α_r - r-ty znacznik intonacyjny

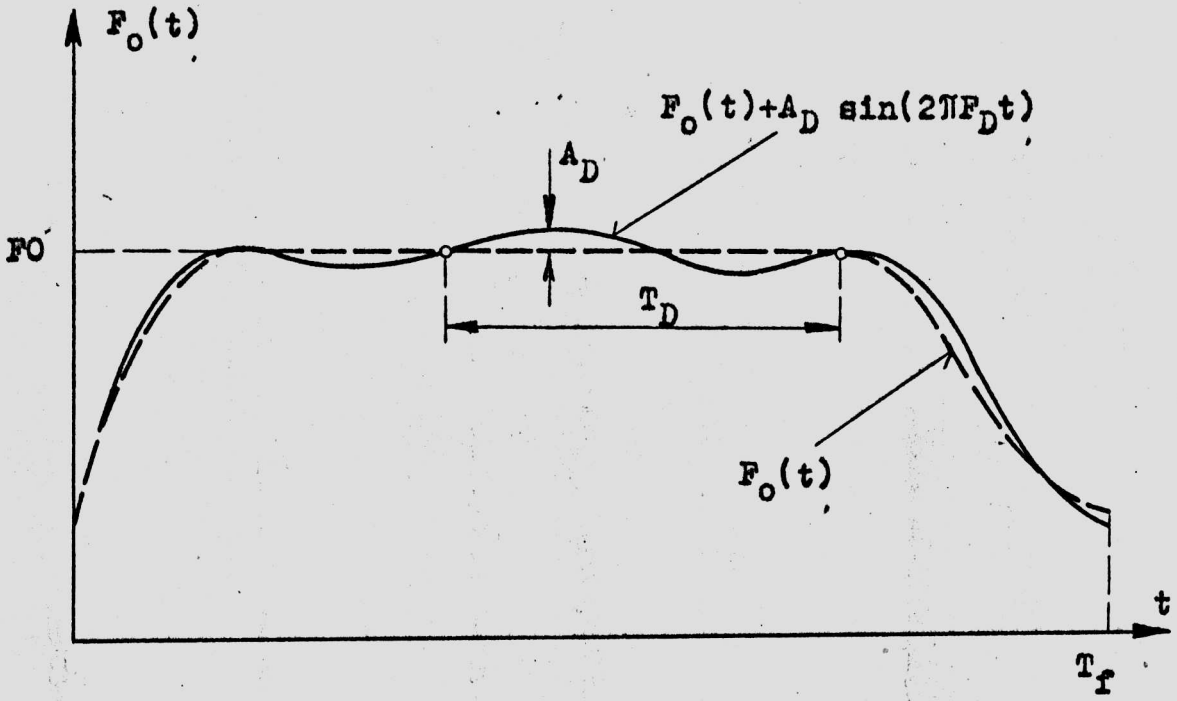
$[+\alpha_r]$ $:=$ - podstawienie warunkowe wykonywane przy wystąpieniu znacznika α_r.



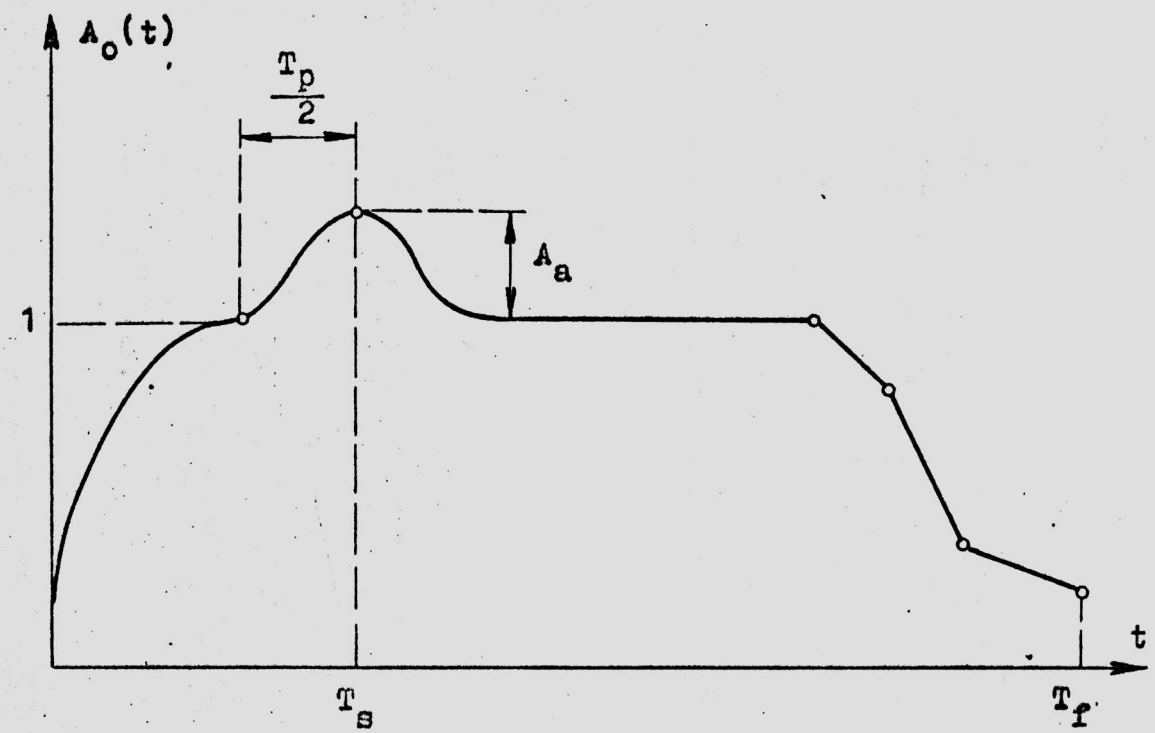
Rys. 3.1. Podstawowy kontur amplitudowy $A_o(t)$ otrzymany przez zastosowanie (RF_1)



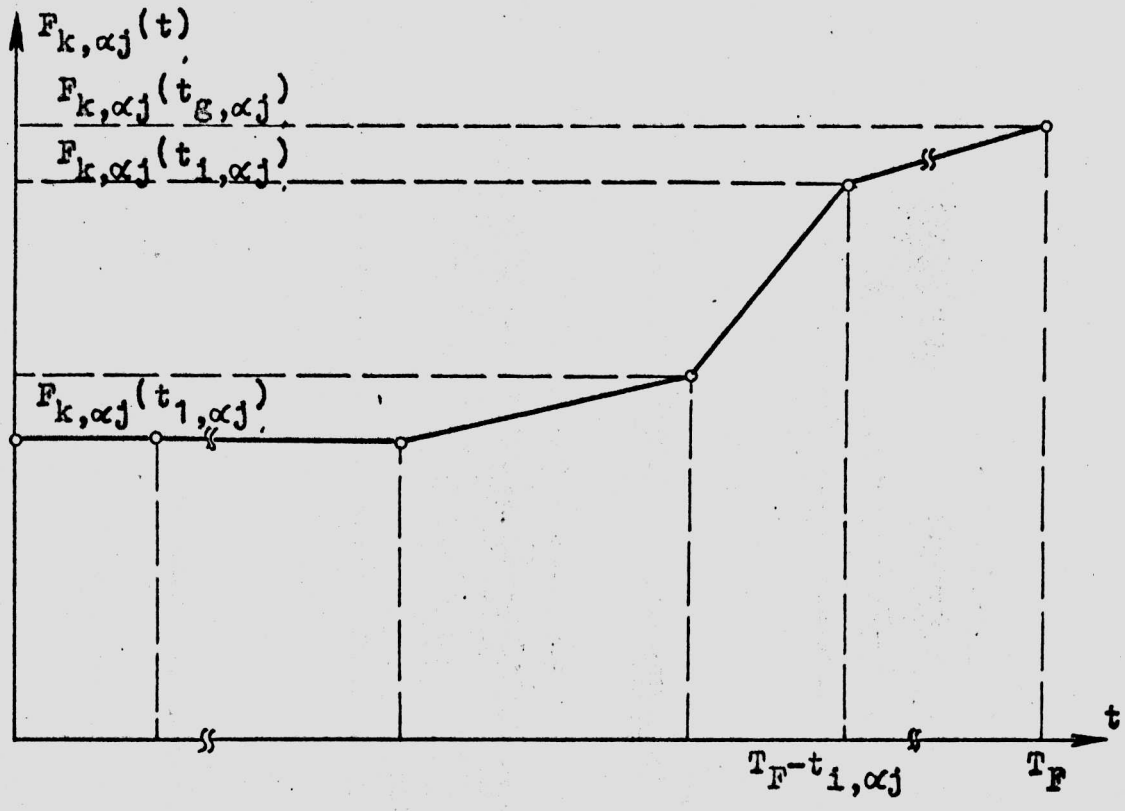
Rys. 3.2. Podstawowy kontur częstotliwościowy $F_o(t)$ otrzymany przez zastosowanie (RF_2), a=0.1



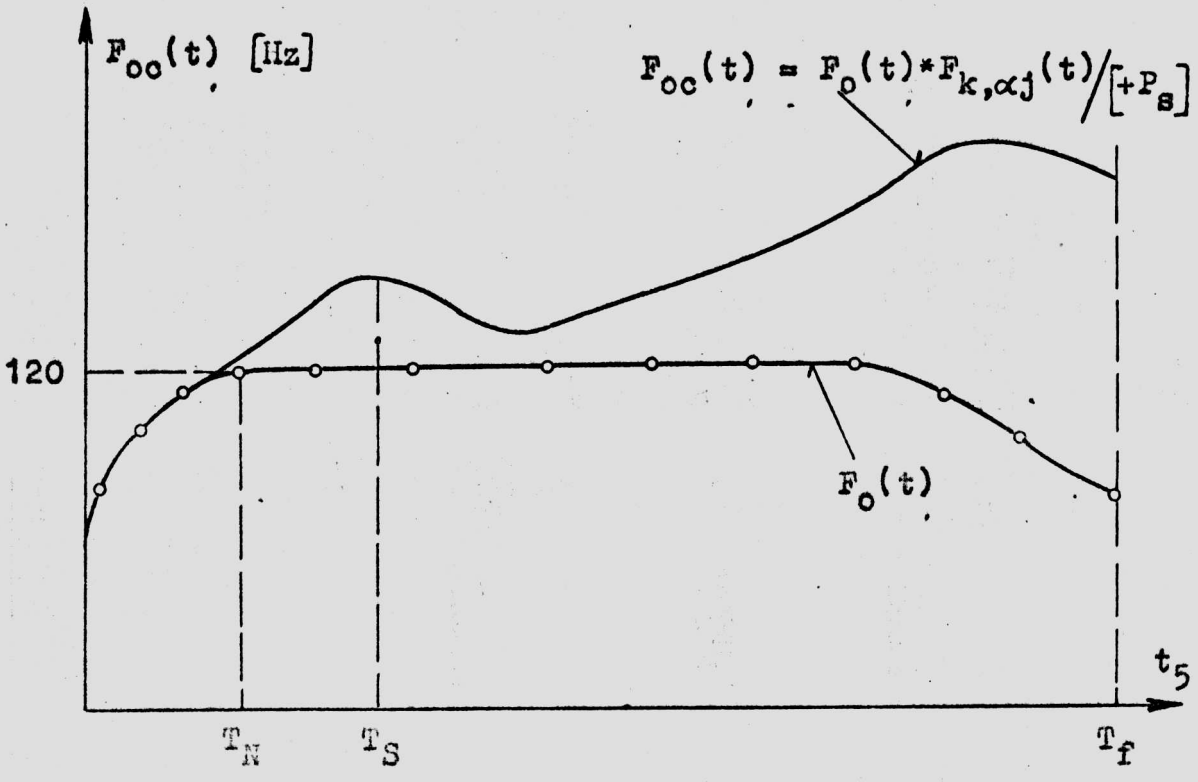
Rys. 3.3. Podstawowy kontur częstotliwościowy $F_o(t)$ po zastosowaniu (RF_3)



Rys. 3.4. Akcentowy wariant podstawowego konturu amplitudowego $A_o(t)$ otrzymany przez zastosowanie (RF_4)



Rys. 3.5. Przykład funkcji intonacyjnej $F_{k,\alpha j}(t)$



Rys. 3.6. Przykład intonacyjnego wariantu podstawowego konturu częstotliwościowego $F_{oc}(t)$ otrzymany przez zastosowanie (RF_6)

3.3. Uwagi do rozdziału 3.2.

1. Podstawowy kontur amplitudowy $A_0(t)$ (reguła RF_1) oraz podstawowy kontur częstotliwościowy $F_0(t)$ (RF_2) stanowią w zbiorze D_{DF} parametrów dziedziny reguł RF odpowiednik archetypowych wzorców czasowego przebiegu zmian częstotliwości podstawowej i intensywności w obrębie grupy wydechowej BG (por. Zał.5 i 6 z rozdz.1.4.5.).
2. Zastosowanie reguły RF_3 oraz reguł RF_4 i RF_5 jest obligatoryjne (tzn. zachodzi przez podstawienie bezwarunkowe), zatem w zapisie reguł (3.1. - 3.5.) nie stosowano odrębnych symboli na oznaczenie wynikowych konturów $A_0(t)$ lub $F_0(t)$.
3. Obligatoryjny charakter stosowania RF_3 założono arbitralnie[25] natomiast w przypadku reguł RF_4 i RF_5 wynika on z przyjęcia Zał.1 (rozdz.1.4.3.1., s.22).
4. Kolejność stosowania reguł nie zawsze spełnia postulat przechodności, np.:

$$G_{RF}\{RF_1 \rightarrow RF_4 \rightarrow RF_2\} \neq G_{RF}\{RF_1 \rightarrow RF_2 \rightarrow RF_4\} \qquad (3.7.)$$

3.4. Parametry dziedziny podzbioru RF reguł realizacyjnych

Ustalenie podzbioru RF reguł realizacyjnych pozwala obecnie na wyodrębnienie zbioru D_{DF} parametrów dziedziny tych reguł. Zgodnie z Zał.6 (rozdz.1.4.5.) elementami D_{DF} są podstawowe kontury amplitudowy $A_0(t)$ i częstotliwościowy $F_0(t)$ oraz ich warianty. W analitycznym opisie $A_0(t)$ i $F_0(t)$ oraz ich wariantów - zależności 3.1. do 3.6. - występuje szereg elementarnych parametrów składowych, istnieje więc konieczność wprowadzenie w zbiorze D_{DF} elementów wielowymiarowych, opisanych pewnymi różnolicznymi podzbiorami parametrów elementarnych:

---

[25] Bez tego założenia reguła RF_3, nie związana z żadnym elementem ciągu T symboli terminalnych komponentu fonologicznego, nie podlegała by zastosowaniu w sensie gramatyki G_{RF}.

$$D_{DF}\left[i \in \{1,\ldots, n\}\ ,\ j(i) \in \{1,\ldots, n(i)\}\ ,\ k[j(i)] \in \{1,\ldots, n[j(i)]\}\ :\ d_{i,j,k}\right] \qquad (3.8.)$$

gdzie: i - indeks w zbiorze parametrów głównych ($A_0(t)$, $F_0(t)$ oraz ich warianty)

$j = j(i)$ - indeks w zbiorze elementarnych parametrów składowych i-tego parametru głównego

$k = k[j(i)]$ - indeks w zbiorze wartości j-tego parametru elementarnego dla i-tego parametru głównego.

Po przeprowadzeniu w D_{DF} procedury optymalizacji (1.2.) zachodzi odwzorowanie:

$$\{d_{i,j,k}\} \xrightarrow{OPT} \{d_{i,j,k_{opt}}\} \qquad (3.9.)$$

co oznacza, że każdemu parametrowi elementarnemu odpowiada tylko jego jedna wartość optymalna. Można zatem zapisać:

$$\{d_{i,j,k_{opt}}\} = \{d_{i,j}\}\ opt = \{d_{i,j}\} \qquad (3.10.)$$

gdzie: $j = j(i)$ - indeks w zbiorze parametrów elementarnych (lub w zbiorze odpowiadających im wartości optymalnych) i-tego parametru głównego.

Po uwzględnieniu (3.9.) i (3.10.) w zależności (3.8.) otrzymujemy:

$$D_{DF_{opt}}\left[i \in \{1,\ldots, n\}\ ,\ j(i) \in \{1,\ldots, n(i)\}\ :\ d_{i,j}\right] \qquad (3.11.)$$

$j = j(i)$ - indeks w zbiorze wartości optymalnych

lub

$$D_{DF}\left[i \in \{1,\ldots, n\}\ ,\ j(i) \in \{1,\ldots, n(i)\}\ :\ d_{i,j}\right] \qquad (3.12.)$$

$j = j(i)$ - indeks w zbiorze parametrów elementarnych

Dla ustalonego w rozdz.3.2. podzbioru RF reguł realizacyjnych:

$$RF\ (l \in \{1,\ldots, 6\}\ :\ RF_l) \qquad (3.13.)$$

Jednowymiarowy zbiór D_{DF} parametrów głównych opisuje zależność:

$$D_{DF}\ \left(i \in \{1,\ldots,\ 6+m\}\ :\ D_i\right) \qquad (3.13.)$$

gdzie:

D_1 - podstawowy kontur amplitudowy $A_o(t)$

D_2 - podstawowy kontur częstotliwościowy $F_o(t)$

D_3 - wariant $F_o(t)$ z dewiacją niskoczęstotliwościową

D_4 - akcentowy wariant $A_o(t)$

D_5 - akcentowy wariant $F_o(t)$

$\forall$ D_{6+r} - (m+1) wariantów intonacyjnych $F_o(t)$

$r \in \{0,\ldots,\ m\}$

Po uwzględnieniu w (3.13.) zależności opisujących reguły RF (3.1. do 3.6.) oraz występujących w nich elementarnych parametrów składowych otrzymujemy:

$$D_1\ (i = 1,\ \ j \in \{1,\ldots,\ 7\}\ :\ d_{i,j}) \qquad (3.14.1.)$$

gdzie:

$d_{1,1} = T_N$ $\qquad$ $d_{1,5} = H$

$d_{1,2} = T_o$ $\qquad$ $d_{1,6} = H_2$

$d_{1,3} = T_{o_1}$ $\qquad$ $d_{1,7} = H_3$

$d_{1,4} = T_{o_2}$

$$D_2\ (i = 2,\ j \in \{1,2\}\ :\ d_{i,j}) \qquad (3.14.2.)$$

gdzie: $d_{2,1} = F_0$, $d_{2,2} = a$

$$D_3\ (i = 3,\ j \in \{1,2\}\ :\ d_{i,j}) \qquad (3.14.3.)$$

gdzie:

$d_{3,1} = A_D$, $d_{3,2} = F_D$

$$D_4\ (i = 4,\ j \in \{1,2\}\ :\ d_{i,j}) \qquad (3.14.4.)$$

gdzie:

$d_{4,1} = A_a$, $d_{4,2} = q$

$$D_5\ (i = 5,\ \ j = 1 : d_{i,j}) \qquad (3.14.5.)$$

gdzie: $d_{5,1} = A_f$

w zależności 3.14.5. wykorzystano $D_4 \cap D_5 = q$

$$D_6(i=1,\ j\ \ 1,\ldots,9\ : d_{i,j}) \qquad (3.14.6.)$$

.
.
.

$$D_{6+m}\ (i = 6 + m,\ j\in \{1,\ldots,\ 9\} : d_{i,j})$$

gdzie:

$d_{6+r,1}$; ... ; $d_{6+r,9}$ - dziewięć par współrzędnych węzłów r-tej funkcji intonacyjnej $F_{k,\alpha_r}(t)$

3.5. Parametry syntezy dźwięków mowy

W pracy przyjęto następujące parametry syntezy dźwięków mowy:

1. Typ funkcji opisującej kształt impulsów pobudzenia krtaniowego:

$$F\ (i \in \{1,\ldots,\ 5\} :\ f_i) \qquad (3.15.)$$

gdzie: F - zbiór typów funkcji

$f_1 = f_A$, $f_2 = f_B$, $f_3 = f_C$, $f_4 = f_D$, $f_5 = f_E$

(por. rys.2.4.)

2. Parametry kształtu impulsów pobudzenia krtaniowego $(t_o\ ;\ t_c)$ - por, rozdz.2.5.3. oraz rys.2.7. :

$$T_G\left[i = j \in \{1,\ldots,\ n\} : (t_{o_i}\ ;\ t_{c_j})\right] \qquad (3.16.)$$

4. CYFROWY MODEL SYNTEZY DŹWIĘCZNYCH FRAZ MOWY POLSKIEJ

4.1. Wstęp

Frazy syntetyczne, stanowiące materiał badawczy w eksperymentach zasadniczych (rozdz.6.), generowano za pomocą zaprojektowanego w ramach pracy cyfrowego modelu syntezy dźwięcznych fraz mowy polskiej. Stosowany model syntezy obejmował zarówno komponent reguł realizacyjnych, jak i modele kanału głosowego oraz źródła pobudzenia krtaniowego. Przy projektowaniu modelu syntezy uwzględniono Zał.7 - 11 (rozdz.1.5.3.).

Schemat blokowy modelu pokazano na rys.4.1., a sieć działań, listę instrukcji i przykładowy wydruk danych z programu SYNTHE, realizującego model syntezy, podano w dodatku D.1. Program SYNTHE napisany w języku FORTRAN IV uwzględniał wymogi procedur INPUT - OUTPUT stosowanego systemu minikomputerowego (rys.2.2.).

Poniżej omówiono ważniejsze bloki funkcjonalne modelu syntezy.

4.2. Komponent reguł realizacyjnych (KRR)

4.2.1. Ciąg T sterujący komponentem KRR

Ciąg sterujący, wprowadzany z klawiatury alfanumerycznej, stanowił symulację ciągu symboli terminalnych otrzymywanych w hierarchicznych systemach syntezy z komponentu fonologicznego (rys.1.2.). W niniejszej pracy ciąg T składał się wyłącznie z elementów podzbioru alfabetu symboli terminalnych komponentu fonologicznego rozpatrywanych na poziomie syntezy fraz. Ogólną postać ciągu T opisuje zależność:

$$T\left[i \in \{1,\ldots,n\} \;:\; e_i(P_s,j)/\alpha_{k_r}\right] \qquad (4.1.)$$

gdzie:

e_i - i-ty element w ciągu T należący do alfabetu P elementarnych jednostek fonetycznych (fonemów), co zapisujemy: $e \leftarrow \{p_j\} \quad \underset{j}{\forall} \; p_j \in P$

$j = j(i)$ - indeks w zbiorze (alfabecie) fonemów, przy czym zachodzi: $e(j) = p_j$

P_s - znacznik akcentu - por. (3.4.) i (3.5.)

αk_r - znacznik typu intonacji frazy

4.2.2. Reguły realizacyjne R

4.2.2.1. Wstęp

Reguły realizacyjne komponentu KRR tworzą trzy podzbiory reguł:

RF = Sb R - reguły generujące funkcje sterujące pobudzeniem krtaniowym

RK = Sb R - reguły generujące funkcje czasowe sterujące modelem kanału głosowego

RT = Sb R - reguły czasowe

Dla podanych reguł zachodzi:

$$RF \cup RK \cup RT = R \qquad (4.2.1.)$$

$$RF \cap RK \cap RT = \emptyset \qquad (4.2.2.)$$

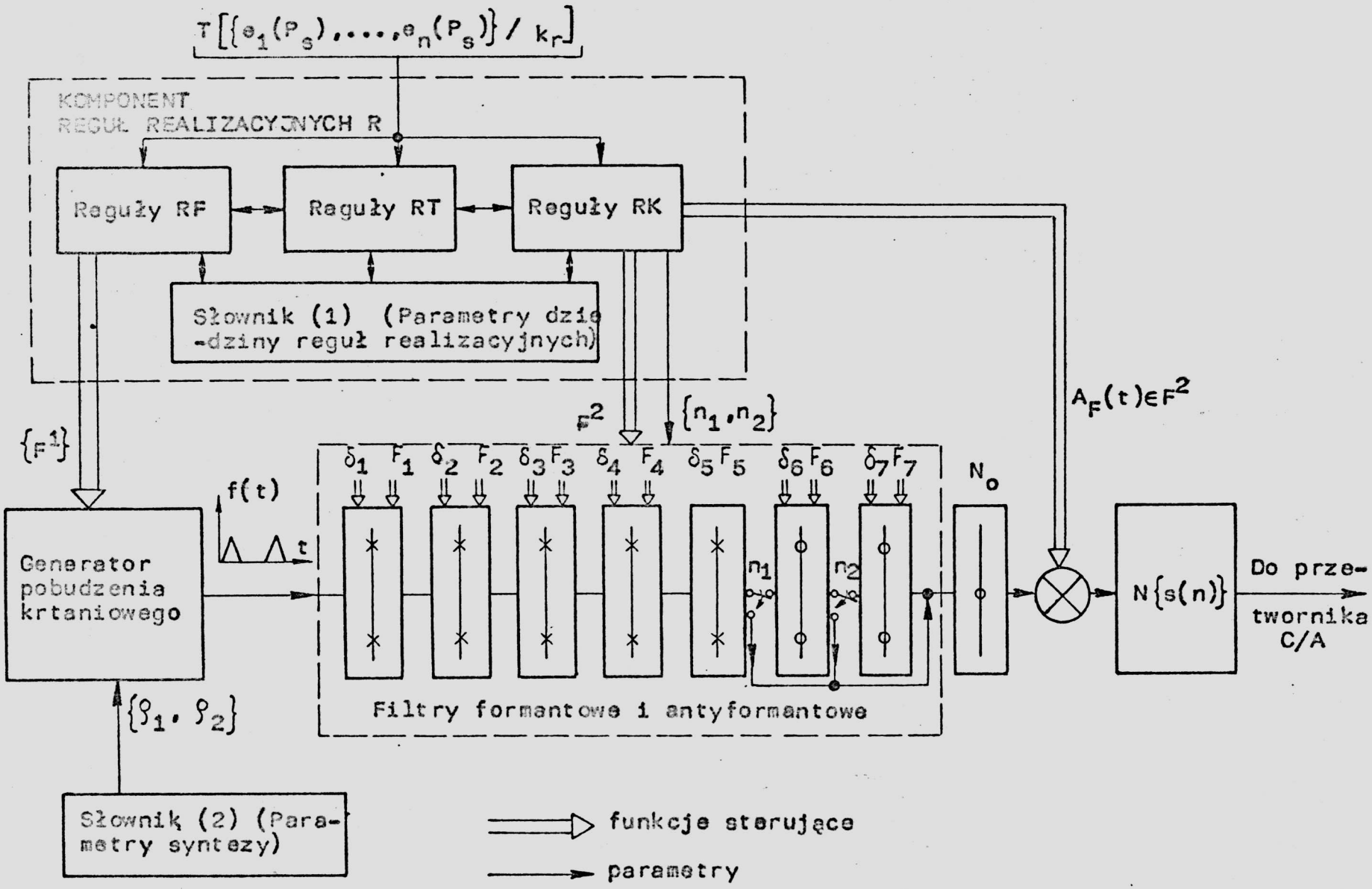
4.2.2.2. Reguły RF - omówiono w rozdz.3.

4.2.2.3. Reguły RK[26)]

RK_1 - reguła realizacji transientu pomiędzy parametrami formantowymi sąsiednich fonemów (rys.4.2.)

$$F_k(t) = \begin{cases} F_{k,i-1} & T'_{i-1} < t < T''_{i-1} \\ F_{k,i-1} + 0.5\,(F_{k,i} - F_{k,i-1}) \cdot & \\ \cdot\,[1 - \cos(\pi F_p t)] & T''_{i-1} \leq t \leq T_i \\ F_{k,i} & T'_i < t < T''_i \end{cases} \qquad (4.3.)$$

26) Podano tylko reguły podstawowe. Badania nad doborem reguł RK i optymalizacją parametrów ich dziedziny są przedmiotem odrębnej pracy doktorskiej, przewidzianej do zakończenia w 1980 r.



Rys. 4.1. Schemat blokowy cyfrowego modelu syntezy fraz

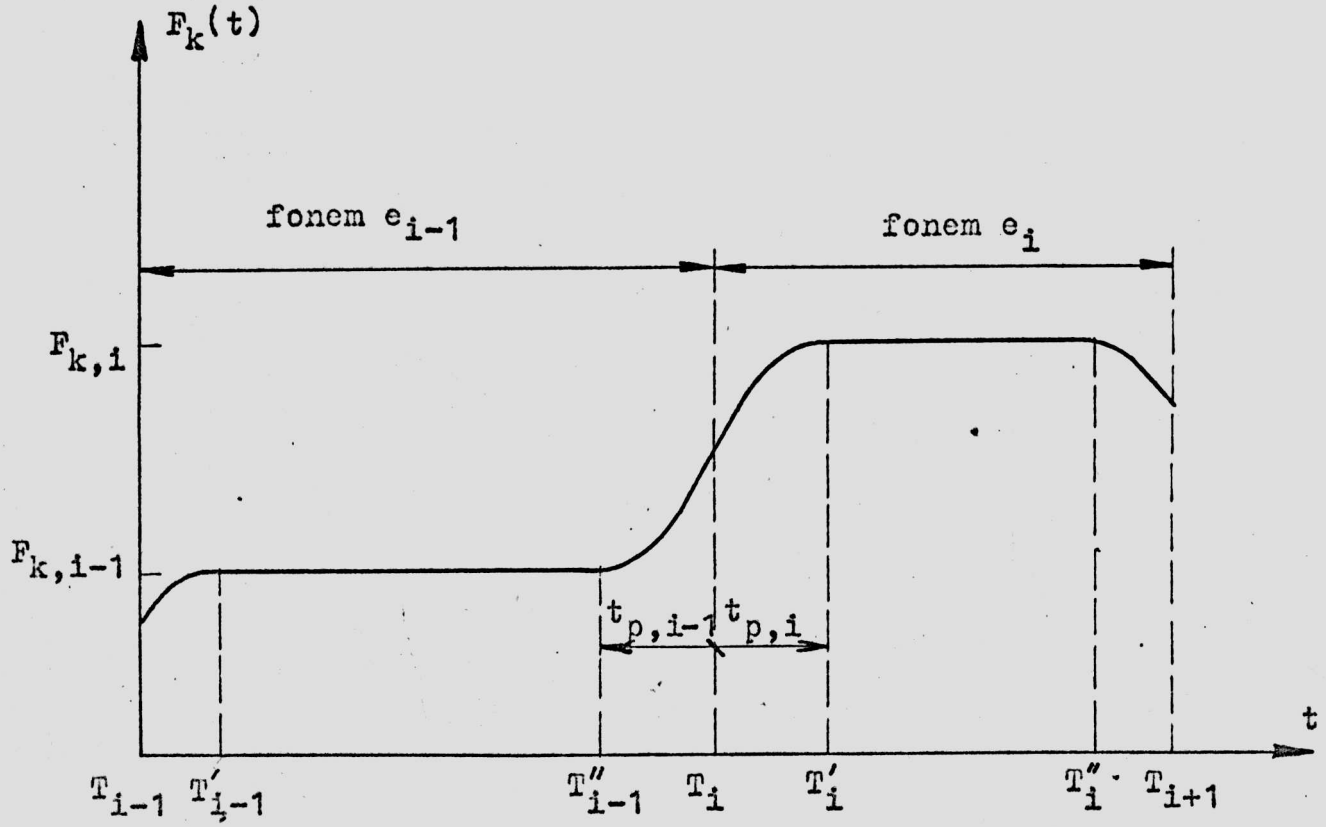
gdzie:

$F_k(t)$ - funkcja sterująca k-tym parametrem formantowym,

$F_{k,i-1}$, $F_{k,i}$ - ustalone wartości k-tego parametru formantowego odpowiednio dla fonemu e_{i-1} oraz e_i,

$F_p = \frac{1}{T_p}$, $T_p = T'_i - T''_{i-1} = t_{p,i-1} + t_{p,i}$ - czas trwania transientu (por.4.5.),

T'_{i-1} , T''_{i-1} , T_i , T'_i , T''_i , $t_{p,i-1}$, $t_{p,i}$ - por.(4.5.), (4.6.1.), (4.6.2.) oraz rys.4.2.



Rys.4.2. Realizacja transientu pomiędzy parametrami formantowymi.

Za pomocą reguły RK_1 generowano stany przejściowe dla wszystkich sterowanych parametrów formantowych, tzn. dla δ_1 do δ_4 oraz δ_6 i δ_7 i F_1 do F_4 oraz N_6 i N_7, gdzie:

$2\pi F_1 - 2\pi F_4$, $\delta_1 - \delta_4$ - część odpowiednio urojona i rzeczywista pary zespolonych biegunów sprzężonych filtrów formantowych,

$2\pi N_6$, $2\pi N_7$, δ_6, δ_7 - część odpowiednio urojona i rzeczywista pary zespolonych zer sprzężonych filtrów antyformantowych N_6 i N_7.

RK_2 - <u>reguła realizacji transientu pomiędzy amplitudami sąsiednich fonemów</u>

$$A_f(t) = \begin{cases} A_{f,i-1} & T'_{i-1} \leq t \leq T''_{i-1} \\ A_{f,i-1} + 0.5\,(A_{f,i-1} - A_{f,i})\,[1 - \\ - \cos(\pi F_p t)] & T''_{i-1} < t \leq T'_i \\ A_{f,i} & T'_i < t \leq T''_i \end{cases} \qquad (4.4.)$$

$A_f(t)$ - funkcja sterująca amplitudą fonemów

$A_{f,i-1}$, $A_{f,i}$ - ustalone wartości amplitud fonemów e_{i-1} i e_i

pozostałe parametry jak w (4.3.).

4.2.2.4. <u>Reguły RT</u>

RT_1 - <u>reguła obliczania czasu trwania transientu</u>

$$T_p = w_1\,(\tau_{i-1} + \tau_i) \qquad (4.5.)$$

w_1 - współczynnik długości transientu,

$0 < w_1 < 1$

τ_{i-1}, τ_i - czasy trwania fonemów e_{i-1} oraz e_i.

RT_2 - <u>reguła proporcjonalnego podziału czasu trwania transientu</u>

$$\begin{cases} T''_{i-1} = T_i - t_{p,i-1} = T_i - w_1 \tau_{i-1} & (4.6.1.) \\ T'_i = T_i + t_{p,i} = T_i + w_1 \tau_i & (4.6.2.) \end{cases}$$

gdzie:

T_i - granica na osi czasu pomiędzy fonemami e_{i-1} oraz e_i ,

$t_{p,i-1}$ - czas trwania transientu w obrębie fonemu e_{i-1} ,

$t_{p,i}$ - czas trwania transientu w obrębie fonomu e_i

RT_3 - <u>reguła wydłużania czasu trwania samogłoski akcentowanej</u>

$$\tau_i \overset{[+P_s]}{:=} w_2 \tau_i \qquad (4.7.)$$

w_2 - modyfikator , $w_2 > 1$

RT_4 - <u>reguła wydłużania czasu trwania ostatniej samogłoski we frazie</u>

$$\tau_i \overset{i=n \wedge e_i \in V}{:=} w_3 \tau_i \qquad (4.8.)$$

w_3 - modyfikator , $w_3 > 1$

$\overset{i=n \wedge e_i \in V}{:=}$ - znak podstawienia warunkowego: "pod warunkiem, że e_i jest ostatnim fonemem we frazie ($i = n$), oraz że e_i należy do zbioru V samogłosek.

RT_5 - <u>reguła obliczania czasu trwania podbicia amplitudowego i częstotliwościowego w akcentowym wariancie konturu $A_o(t)$ i $F_o(t)$</u>

Regułę RT_5 podano w rozdz.3., zależność (3.4.) i (3.5.).

RT_6 - reguła obliczania czasu trwania frazy

$$T_F = \sum_{i=1}^{n} \tau_i$$

Regułę RT_6 stosuje się po RT_3 i RT_4.

4.2.3. Dziedzina D reguł realizacyjnych R

Zgodnie z zależnością (4.2.1.):

$$RF \cup RK \cup RT = R$$

zachodzi zatem:

$$DF \cup DK \cup DT = D \qquad (4.9.)$$

gdzie: DF, DK, DT - dziedziny reguł odpowiednio RF, RK, RT.

Ponieważ zachodzi również[27]:

$$DT \in DF \cup DK \qquad (4.10.)$$

skąd:

$$DF \cup DK \cup DT = DF \cup DK \qquad (4.11.)$$

więc zależność (4.9.) upraszcza się do postaci:

$$D = DF \cup DK \qquad (4.12.)$$

Dziedzinę reguł RF tworzą funkcje sterujące pobudzeniem krtaniowym, co zapisujemy:

$$DF = \{F^1\} = \{A_0(t), F_0(t)\} \qquad (4.13.)$$

Dziedzinę reguł RK tworzą funkcje sterujące modelem kanału głosowego:

$$DK = \{F^2\} \qquad (4.14.)$$

gdzie:

$$F^2(i \in \{1,\ldots 4\} : \sigma_i(t), F_i(t) \cup k \in \{6,7\} : : \sigma_k(t), N_k(t) \cup A_F(t) \qquad (4.15.)$$

Oznaczenia występujące w (4.15.) objaśniono w rozdz.4.2.2.3.

---

27) Zależność (4.10.) wynika ze specyfiki reguł czasowych, które wpływają na przebieg funkcji generowanych przez reguły RK i RF, ale same nie generują żadnych funkcji. Można je zatem określić jako reguły z obcą dziedziną.

4.2.4. Zbiór D_D parametrów dziedziny reguł realizacyjnych R (Słownik (1))

Dla zbioru D_D zachodzi:

$$D_D = D_{DF} \cup D_{DK} \cup D_{DT} \tag{4.16.}$$

gdzie:

D_{DF} , D_{DK} , D_{DT} - zbiory parametrów dziedziny reguł odpowiednio RF, RK, RT.

Ponieważ zachodzi równocześnie:

$$D_{DF} \cap D_{DT} \cap D_{DK} \neq \emptyset \tag{4.17.}$$

zatem w modelu syntezy utworzono dla reguł RK, RF i RT jeden wspólny Słownik (1) zawierający zbiór parametrów D_D (rys.4.1.) opisany zależnością (4.16.).

Zbiór D_{DF} : podano w rozdz.3.

Zbiór D_{DK} :

$$D_{DK} \; (j \in \{1,\ldots,J\} : p_j) \tag{4.18.}$$

gdzie:

$$p_j \; (i \in \{1,\ldots,4\} : \mathcal{G}_i^j \, , \, F_i^j \cup k \in \{6,7\} :$$

$$: \mathcal{G}_k^j \, , \, N_k^j \cup A_F^j \cup \tau^j \tag{4.19.}$$

Objaśnienia oznaczeń podano w rozdz.4.2.2.3.

Zbiór D_{DT} :

$$D_{DT} \; (i \in \{1,\ldots 3\} : w_i \cup q) \tag{4.20.}$$

w_i - por. rozdz.4.2.2.3.

q - por. (3.4.) i (3.5.) w rozdz.3.

4.2.5. Uwagi do rozdziału 4.2.2.

1. W zbiorze reguł RK znajduje się pewna liczba reguł dodatkowych jak reguła generująca binarne parametry n_1 i n_2 sterujące włączaniem (wyłączaniem) filtrów antyformantowych.

2. W Słowniku (2) - rys.4.1. - zamieszczono parametry syntezy podane w rozdz.3.5.

4.3. Model kanału głosowego

Stosowano model w zasadzie identyczny, jak w eksperymentach wstępnych (rozdz.2.2.1.) uzupełniając go:

1. dwoma filtrami antyformantowymi N_6 i N_7 (rys.4.1.) realizującymi pary zespolonych zer sprzężonych,
2. filtrem NO realizującym zero rzeczywiste (rys.4.1.), którego zadaniem było wprowadzenie korekcji wynikającej z charakterystyki promieniowania ust (+6 dB/oktawę),
3. mnożnikiem na wyjściu kaskady filtrów (rys.4.1.) zapewniającym możliwość sterowania amplitudą sygnału wyjściowego,
4. blokiem normalizacji wartości chwilowych syntezowanego sygnału (rys.4.1.). W bloku normalizacji dokonywano liniowego przekształcenia:

$$N\ \{s(n)\} \xrightarrow[n]{\forall}\ s(n) \in \{-128\ ;\ 127\} \qquad (4.21.)$$

gdzie:

s(n) - ciąg chwilowych wartości sygnału.

Dzięki przekształceniu (4.21.) możliwe było optymalne wykorzystanie dynamiki stosowanego przetwornika C/A pracującego w 8-bitowym kodzie binarnym, uzupełnieniowym.

4.4. Układ syntezy fraz

Stosowano układ identyczny jak w eksperymentach wstępnych (por. rozdział 2.2.1., rys.2.2.).

5. METODY SUBIEKTYWNEJ OCENY FRAZ SYNTETYCZNYCH

5.1. Wstęp

W badaniach nad mową syntetyczną podstawową i najczęściej stosowaną metodą oceny uzyskanego materiału eksperymentalnego są subiektywne testy odsłuchowe [28-31, 33-37, 54, 58, 60, 64, 72, 75, 76, 79, 81, 82, 103-109] . Złożoność procesu syntezy mowy (por.rozdz.1.1., 1.2., 1.4.) oraz różnorodność czynników wpływających na generowany sygnał syntetyczny powodują, że dotychczas nie opracowano spójnej i powszechnie przyjętej metodyki subiektywnej oceny mowy syntetycznej. Trudne jest nawet podanie ogólnej systematyki stosowanych pomiarów subiektywnych, gdyż przyjmując podstawowe kryteria podziału np.:

- co mierzymy ?
- jak mierzymy i z jaką dokładnością ?
- w jakim celu mierzymy ?

otrzymujemy nierozłączne zbiory metod, przy czym w obrębie zbioru utworzonego przy przyjęciu jednego kryterium mogą się znaleźć metody, które przy zastosowaniu innego kryterium mają ze sobą niewiele wspólnego. W konsekwencji,o przyjęciu w badaniach danej metody oceny subiektywnej decyduje ogólna analiza struktury eksperymentu oraz założonych celów badawczych. Nie bez wpływu są tu też czynniki pragmatyczne (możliwości aparaturowe, finansowe i czasowe) oraz metodologiczne nawyki osób realizujących badania[28].

Te problemy są dobrze znane osobom zajmującym się szeroko pojętymi badaniami mowy i ich szczegółowe omówienie jest w niniejszej pracy zbędne. Celowe jest natomiast podanie przesłanek, którymi się kierowano w doborze stosowanych w pracy technik i metod ocen subiektywnych (por. następny rozdz.5.2.)

---

[28] Badania w tej dziedzinie prowadzone są przez przedstawicieli różnych dyscyplin naukowych - por. rozdz. 1.1., s.7.

5.2. Stosowane w pracy metody ocen subiektywnych

5.2.1. Kryteria wyboru

Zasadniczym kryterium wyboru metod oceny subiektywnej było przyjęcie podanych w rozdz.1.3.2. miar jakości mowy syntetycznej (Def.9. i 10.) oraz podstawowych procedur wykonywanych w zbiorze reguł generacji pobudzenia krtaniowego (Def.11. i 12.). Czynniki pragmatyczne były kryterium dodatkowym.

5.2.2. Wybór metod

5.2.2.1. Metody oceny wierności odtwarzania formalnych elementów systemu języka

W niniejszej pracy przyjęto (rozdz.1.3.2., s.18) wierność odtwarzania jako kryterium podstawowe w procedurze doboru reguł. Dopiero po spełnieniu tego kryterium przystępowano do optymalizacji parametrów reguł, gdzie stosowano kryterium naturalności brzmienia.

Przyjęta definicja wierności odtwarzania formalnych elementów systemu języka (Def.9.) narzuca jednoznacznie metodę subiektywnej oceny tej miary, polegającą na binarnym wyborze: M_1:{ element e jest odtwarzany ∨ element e nie jest odtwarzany }[29]. Przykładem są metody stosowane w badaniach Majewskiego i Blasdell'a [64] oraz Studdert-Kennedy i Hadding [79] dotyczących oceny konturów intonacyjnych, gdzie zadaniem słuchaczy było zakwalifikowanie badanej próbki do zbioru twierdzeń lub pytań. Podobną metodę stosował Jassem i in. [54] oraz Fiodorowa [103] w badaniach nad percpecją akcentu. W badaniach Jassema słuchacze podejmowali trójdzielną decyzję: a) akcent przypada na pierwszą sylabę, b) akcent przypada na drugą sylabę, c) brak akcentowego zróżnicowania sylab. Fiodorowa zastosowała klasyfikację wymuszoną, polegającą na dwudzielnej decyzji a) lub b).

---

29) Dla uproszczenia przyjęto tożsamość decyzji "element e nie jest odtwarzany" oraz "element e jest odtwarzany jako element d".

Przytoczone przykłady dotyczą wprawdzie wariantów metody M_1 podanej we wstępie rozdziału, zawierają jednak główną cechę metody, tzn. wybór w oparciu o decyzję binarną[30].

Do klasy M_1 należą również ilościowe miary jak wyrazistość głoskowa lub wyrazowa (por.rozdz.1.3.2. oraz odsyłacz [9]).

---

[30] Np., metoda stosowana przez Majewskiego [64] stanowi wariant M_1^{WM} metody M_1, gdzie przyjęto założenia:

1) $\overset{\forall}{i}\ e_i \in E_I \cup E_S$ (1)

2) $\overset{\forall}{i}\ e_i \xrightarrow{SUB} s_i \in S_I \cup S_S$

Natomiast w metodzie M_1 obowiązują założenia:

1) $\overset{\forall}{i}\ e_i \xrightarrow{SUB} s_i \in S(+E) \cup S(-E)$ (2)

2) $s_i \in S(+E) \Rightarrow e_i \in E$, $s_i \in S(-E) \Rightarrow e_i \notin E$

Reguły decyzyjne w M_i^{WM} i M_1 zapisujemy:

(M_1^{WM}) : $s_i \in S_I \Longleftrightarrow \tilde{P}(s_i \in S_I) > \tilde{P}(s_i \in S_S)$ (3)

w pozostałych przypadkach $s_i \in S_S$

(M_1) : $s_i \in S(+E) \Longleftrightarrow \tilde{P}[s_i \in S(+E)] > \varepsilon$ (4)

w pozostałych przypadkach $s_i \in S(-E)$

oznaczenia:
- e_i - badane elementy (typy konturów)
- s_i - próbki syntetyczne wygenerowane z zastosowaniem e_i
- E_I, E_S - zbiory konturów, odpowiednio, pytających i twierdzących
- S_I, S_S - zbiory próbek, odpowiednio, o intonacji pytającej i twierdzącej
- $S(+E)$, $S(-E)$ - zbiory próbek, odpowiednio, mających zadaną intonację i nie mających tej intonacji
- $\tilde{P}$ - subiektywnie określony estymator prawdopodobieństwa (procent głosów przypadających na daną decyzję)
- ε - zadana, progowa wartość prawdopodobieństwa, $\varepsilon \geqslant 0.5$.

Porównując (3) z (4) oraz (1) z (2) stwierdzamy, że M_1 oraz

W pracy, ze względu na ograniczenie problemu syntezy do poziomu frazy, w ciągu T symboli sterujących komponentem KRR (4.1.) można wyróżnić trzy zbiory formalnych elementów systemu języka, które podlegają postulatowi różnowartościowego odwzorowania w substancji dźwiękowej frazy, są to:

- zbiór elementarnych jednostek fonetycznych $\{e(j) = p_j\}$
- zbiór typów intonacji $\{\alpha k_r\}$
- jednoelementowy zbiór cechy akcentu $\{P_s\}$

Przy doborze parametrów reguł RK, dokonujących odwzorowania wykorzystywanego w pracy zbioru fonemów $\{p_j\}$ (tab.2.1. oraz tab.6.1.) w ich dźwiękową postać (głoski), stosowano kryterium wyrazistości głoskowej[31].

Przy ustalaniu obszaru wartości parametrów reguł akcentowych RF_4 i RF_5 (3.4.), (3.5.), zapewniających jednoznaczne uzyskanie cechy akcentu we frazie, stosowano metodę M_1 z decyzją binarną (występuje akcent $\vee$ brak akcentu).

W odniesieniu do typów intonacji (w pracy rozważano tylko intonację pytającą i twierdzącą - por.rozdz.6.) oparto się o wyniki badań podane w pracach [56, 64, 79] i nie przeprowadzono własnych badań nad doborem reguły intonacyjnej oraz jej parametrów.

5.2.2.2. Metody oceny naturalności brzmienia

Naturalność brzmienia (Def.10.) była podstawowym kryterium w procedurze optymalizacji parametrów dziedziny reguł oraz parametrów syntezy dźwięków mowy (Def.12.). Zgodnie z wprowadzoną przez autora definicją naturalności brzmienia (Def.10.) obej-

---

M_1^{WM} różnią się założeniami, natomiast reguły decyzyjne są zbliżone (identyczne, gdy w (4) przyjmiemy $\varepsilon = P\,[s_i \in S\ (-E)]$). W praktyce, to oznacza podobną technikę realizacji pomiaru subiektywnego przy różnicy w sformułowaniu zadania dla ekipy odsłuchowej.

31) W pracy nie zamieszczono opisu tych badań - por. odsyłacz[26] - podano tylko końcowe wyniki w postaci ustalonych wartości parametrów formantowych.

muje ona m.in. takie cechy fraz syntetycznych jak naturalność akcentu lub naturalność intonacji pytającej, przy czym kryterium jest tu podobieństwo subiektywnie odebranej cechy do archetypowego wzorca tej cechy dla mowy naturalnej (w pracy założono, że wzorce archetypowe są niezmiennikiem w procesie odbioru mowy przez użytkownika danego języka). Subiektywna ocena tak pojmowanej naturalności brzmienia wymaga zastosowania znacznie precyzyjniejszej techniki niż ocena ciągu próbek w zadanej skali punktowej. W pracy przyjęto technikę opartą o test porównań w parach (test A-B) [110-112]. Test A-B stosowano również w szeregu innych badaniach nad jakością lub naturalnością mowy syntetycznej [75, 82, 107, 108] lub w badaniach nad zdolnością rozróżniania zmian w wartościach parametrów reguł syntezy [105, 106]. Istotną niedogodnością testu A-B jest wzrost liczby ocenianych par z kwadratem liczby próbek, co w praktyce ogranicza liczność badanego zbioru próbek do kilkunastu. W przypadku zbioru próbek o większej liczności stosowano test eliminacyjny metodą ocen punktowych.

W badaniach nad doborem parametrów reguły niskoczęstotliwościowej dewiacji F_0 (reguła RF_3, (3.3.)), gdzie badania wstępne nie dały jednoznacznej odpowiedzi, czy jej wprowadzenie poprawia naturalbość brzmienia (por.rozdz.2.5.7.), zastosowano test A-C. W tym teście próbka C była we wszystkich parach identyczna i generowano ją bez zastosowania reguły RF_3. Zbiór próbek A generowano z zastosowaniem RF_3 dla różnych wartości parametrów A_D i F_D (3.3.).

Metodę analizy wyników testu A-B i A-C oraz metodę oceny ich statystycznej wiarygodności podano w rozdziale 5.2.3.1. i 5.2.3.2.

5.2.3. Testy porównań w parach

5.2.3.1. Test A-C

W teście A-C dla zbioru S ($i \in \{1,\ldots,n\} : s_i$) próbek o liczności n dokonywano n porównań z próbką odniesienia C, przy czym S = A.

Kryterium oceny próbek $s_i \in A$ był wynik testowania hipotezy:

$$H_0: \bar{x}(s_i) > \bar{x}(c) \qquad (5.1.)$$

gdzie:

$$\bar{x}(s_i) = \frac{1}{LK} \sum_{l=1}^{L} \sum_{k=1}^{K} r_i^{k,l}$$

$$\bar{x}(C) = 1 - \bar{x}(s_i)$$

$l \in \{1,\ldots, L\}$ - indeks w zbiorze ocen niezależnych

$k \in \{1,\ldots, K\}$ - indeks w zbiorze fraz syntetycznych

$r_i^{k,l}$ - ocena uzyskana w l-tym eksperymencie, dla k-tej frazy z i-tą wartością badanego parametru

$r_i^{k,l} \in \{0;1\}$

Test istotności różnic pomiędzy średnimi przeprowadzono za pomocą standardowego programu "T TEST FOR GIVEN MEAN" z biblioteki programów statystycznych mikrokomputera TEK-31.

Do dalszych badań, realizowanych metodą testu, A-B brano próbki z wartościami parametru, dla których na poziomie $\alpha = 0.03$ nie było podstaw do odzrucenia hipotezy H_0.

5.2.3.2. Test A-B

5.2.3.2.1. Procedury analityczne

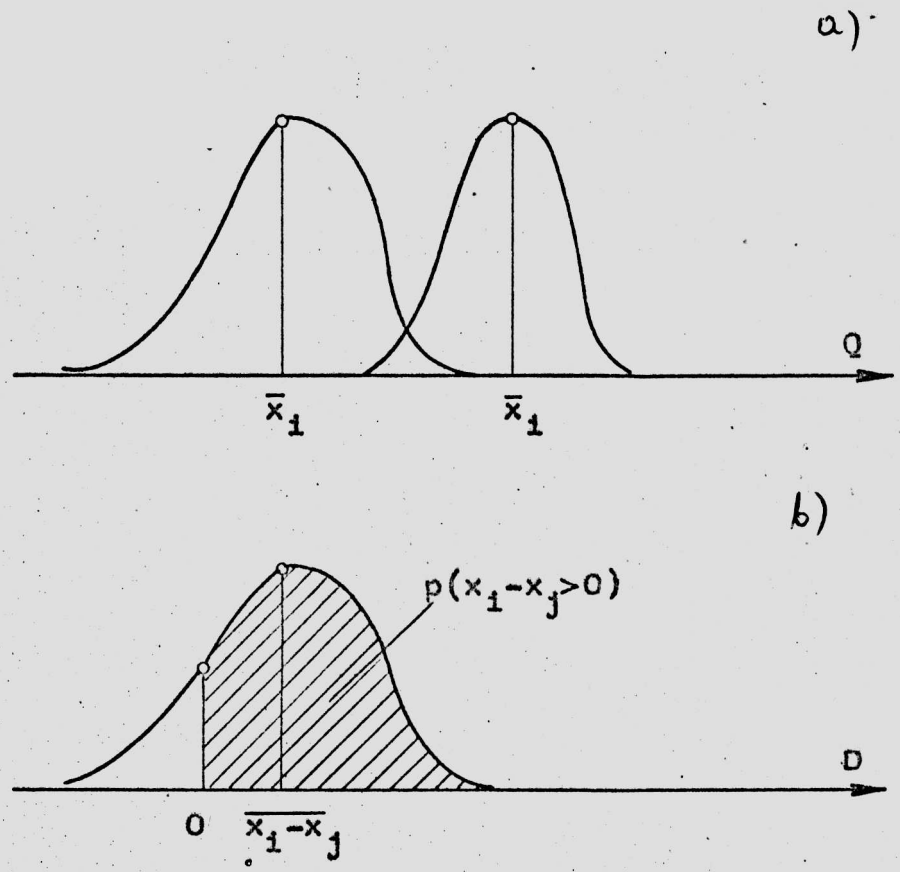
Podstawową procedurą w analizie wyników odsłuchów n próbek metodą testu A-B jest przekształcenie n(n-1)/2 oszacowań względnych różnic między ocenianymi próbkami na ilorazową skalę ocen punktowych. W pracy, przy realizacji tej procedury przyjęto założenia V modelu ocen porównawczych Thurstone'a [113] oraz rozwiązanie Mostellera [114].

Założenia:

1) Istnieje zbiór próbek $S(i \in \{1,\ldots,n\} : s_i)$ rozmieszczony na subiektywnym continuum obserwacji Q (rys.5.1.a.).
2) Rozkład procesu dyskryminacyjnego (tzn. rozkład subiektywnych reakcji na próbkę s_i w L niezależnych eksperymentach) jest normalny.

3) Istnieje możliwość podania oceny preferencji próbki i nad j: $(i \prec j)$ lub $(i \succ j)$.

4) Odchylenia standardowe w rozkładzie procesów dyskryminacyjnych oraz korelacja między reakcjami na próbki są równe.



Rys.5.1. Rozkłady procesów dyskryminacyjnych i różnic pomiędzy nimi
a) Subiektywne continuum procesów dyskryminacyjnych
b) continuum różnic pomiędzy procesami dyskryminacyjnymi.

Przyjmując, że x_i oraz x_j są reakcjami na próbkę, odpowiednio, s_i oraz s_j i uwzględniając założenia 1-4 otrzymamy:

$$\bar{x}_i = Q_i \qquad i \in \{1, \ldots, n\} \tag{5.2.1.}$$

$$\bar{x}_j = Q_j \qquad j \in \{1, \ldots, n\} \tag{5.2.2.}$$

$$\mathrm{Var}\,(x_i) = \mathrm{Var}\,(x_j) = \sigma^2 = \mathrm{const} \tag{5.2.3.}$$

$$r_{i,j} = \frac{\mathrm{cov}\,(x_i, x_j)}{\sigma_j \sigma_i} = \varrho = \mathrm{const} \tag{5.2.4.}$$

Z danych doświadczalnych otrzymujemy macierz $[N]$, której elementami $n_{i,j}$ jest liczba decyzji preferujących próbkę i-tą nad j-tą. Macierz $[N]$ unormowaną względem liczby niezależnych eksperymentów nazywamy macierzą preferencji $[P]$, gdzie:

$$p_{i,j} = \frac{n_{i,j}}{L} \tag{5.3.}$$

L - liczba niezależnych eksperymentów

Zakładając, że $p_{i,j}$ jest estymatorem $P\,(x_i > x_j)$ i uwzględniając założenia 1-4 oraz (5.2.1.) do (5.2.4.) możemy zapisać:

$$p_{i,j} \cong P\,(x_i > x_j) = \frac{1}{\sqrt{2\pi}} \int_{-D_{i,j}}^{\infty} e^{-0.5y^2}\, dy \tag{5.4.}$$

gdzie:

$D_{i,j} = Q_i - Q_j$

y - standaryzowana zmienna losowa

Zgodnie z rozwiązaniem podanym przez Mostellera [114] poszczególne wartości oczekiwane $\bar{x}_i = Q_i$ procesów dyskryminacyjnych obliczamy z zależności:

$$Q_i = \frac{1}{n}\left[\sum_{j=1}^{n} D_{j,i_o} - \sum_{j=1}^{n} D_{j,i}\right] \tag{5.5.}$$

gdzie:

$D_{j,i_o} = Q_j - Q_{i_o} = Q_j$

Q_{i_o} - wartość oczekiwana rozkładu procesu dyskryminacyjnego dla najniżej ocenionej próbki.
Dla uproszczenia arbitralnie założono $Q_{i_o} = 0$.

Po wyliczeniu z (5.5.) zbioru Q ($i \in \{1,\ldots,n\} : Q_i$) przeliczono wartości Q_i celem otrzymania ilorazowej skali jakości próbek s_i w przedziale $\langle 0;10 \rangle$:

$$\forall_i \; Q_i^{(10)} = \frac{10}{\max_j \{Q_j\}} \; Q_i \qquad (5.6.)$$

Podane procedury wyliczano za pomocą opracowanego programu RANK ORDER w języku wewnętrznym mikrokomputera TEK-31. Przejście z macierzy [P] na macierz [D] realizowano rozwiązując całkę w zależności (5.4.) metodą całkowania numerycznego przez bisekcję[32].

5.2.3.2.2. Procedury statystyczne

Szczegółowe omówienie statystycznej analizy wyników testu A-B zamieszczono w pracach [110, 111]. W rozdz.5.2.3.2.2. ograniczono się do podania podstawowych założeń i zależności.

Stosowane procedury statystyczne miały na celu określenie, czy przy zadanym poziomie istotności zachodzi:

- zgodność słuchaczy w ocenie zbioru S
- możliwość szeregowania zbioru S

Dodatkowo określono istotność różnic pomiędzy średnimi ocenami próbek s_i oraz s_j, $i,j \in \{1,\ldots, n\}$.

Ocena zgodności słuchaczy (P1)

W tej procedurze wylicza się:

1. Liczbę par w zbiorze słuchaczy zgodnie orzekających, że próbka s_i jest preferowana nad s_j

$$J_{i,j} = C \binom{n_{i,j}}{2} \qquad (5.7.)$$

gdzie:

$n_{i,j}$ - por. (5.3.)

---

[32] Ze względu na nietypowy język programu RANK ORDER zrezygnowano z jego zamieszczenia.

2. Ogólną liczbę par zgodnych słuchaczy

$$J = \sum_{i=1}^{n} \sum_{j=1}^{n} J_{i,j} \qquad (5.8.)$$

3. Miarę stopnia przypadkowej zgodności:

$$\delta = \frac{4J}{m-2} - (m-3)\nu \qquad (5.9.)$$

gdzie:

δ - zmienna losowa o rozkładzie χ^2

i ν stopniach swobody,

$$\nu = \frac{m(m-1)\, n\, (n-1)}{2(m-2)}$$

m,n - liczność zbioru, odpowiednio, słuchaczy i próbek

4. Zmienną losową Z o rozkładzie normalnym, standaryzowanym:

$$Z = \sqrt{2\delta} - \sqrt{2\nu - 1} \qquad (5.10.)$$

5. Prawdopodobieństwo przypadkowej zgodności:

$$P\,(X > Z)\,, \quad [\%] \qquad (5.11.)$$

Ocena możliwości szeregowania (P2)

W procedurze P2 wylicza się:

1. Sumę ocen preferujących s_i nad s_j (w zbiorze słuchaczy):

$$R_i = \sum_{j=1}^{n} n_{i,j} \qquad (5.12.)$$

2. Wariancją rozkładu zmiennej losowej R:

$$\mathrm{Var}\,(R) = \frac{1}{n}\left[\sum_{i=1}^{n} R_i^2 - \frac{nm^2\,(n-1)^2}{4}\right] \qquad (5.13.)$$

3. Współczynnik zgodności szeregowania (współczynnik Kendalla):

$$w = \frac{\mathrm{Var}\,(R) - 1}{\mathrm{Var}_{max} + 2} \qquad (5.14.)$$

gdzie:

$$\mathrm{Var}_{max} = \frac{1}{12}\, m^2\,(n^2 - 1)$$

4. Zmienną losową F o rozkładzie Snedecora, stanowiącą miarę możliwości szeregowania:

$$F = \frac{w\,(m-1)}{1-w} \qquad (5.15.)$$

o liczbie stopni swobody: $\nu_1 = n - 1 - 2/m$

$\nu_2 = (m-1)\,(n-1) - 2/m$

Ocena możliwości zróżnicowania próbek (P3)

W procedurze P3 wylicza się:

1) Macierz T zmiennych losowych o wyrazach:

$$t_{i,j} = \frac{(R_i - R_j)/(m \cdot n - m)}{\sqrt{\frac{R_i - R_j}{m \cdot n - m}\left(1 - \frac{R_i - R_j}{2(mn - n)}\right)}} \qquad (5.16.)$$

Zakłada się, że $t_{i,j}$, $i, j \in \{1, \ldots, n\}$ mają rozkład t-Studenta o $\nu = mn - m - 2$ stopniach swobody.

Kryteria oszacowania wyników testu A-B

P1:

Kryterium zgodności słuchaczy jest wartość $P(X > Z)$ wyrażone zależnością (5.11.). Przyjmuje się [110, 111]:

K1
- $P(X > Z) < 0.3$ [%] - zgodność pewna
- $0.3 \leq P(X > Z) \leq 4.6$ [%] - zgodność prawdopodobna
- $P(X > Z) > 4.6$ [%] - zgodność wątpliwa

P2:

Kryterium możliwości szeregowania jest wartość zmiennej F (5.15.). Przyjmuje się [110, 111]:

K2
- $F > F_{0.01}$ - możliwość szeregowania pewna
- $F_{0.05} \leq F \leq F_{0.01}$ - możliwość szeregowania prawdopodobna
- $F \leq F_{0.05}$ - możliwość szeregowania wątpliwa

$F_{0.01}$, $F_{0.05}$ - wartości zmiennej F dla rozkładu Snedecora dla poziomów istotności, odpowiednio $\alpha = 0.01$ i $\alpha = 0.05$.

P3:

Kryterium możliwości różnicowania próbek s_i oraz s_j jest wynik testowania hipotezy:

$$H_o : R_i = R_j \qquad (5.17.)$$

Przyjmuje się [110, 111]:

K3 $\begin{cases} t_{o.o5} \leqslant \begin{cases} t_{i,j} > t_{o.o1} & \text{- możliwość różnicowania pewna} \\ t_{i,j} \leqslant t_{o.o1} & \text{- możliwość różnicowania prawdopodobna} \\ t_{i,j} < t_{o.o5} & \text{- możliwość różnicowania wątpliwa} \end{cases} \end{cases}$

gdzie:

$t_{o.o1}$, $t_{o.o5}$ - wartości zmiennej t-Studenta, odpowiednio, dla poziomu istotności $\alpha = 0.01$ i $\alpha = 0.05$.

$t_{i;j}$ - opisano zależnością (5.16.)

$t_{i,j} \geq t_{\alpha}$ - oznacza, że na poziomie α hipotezę H_o (5.17.) można odrzucić

Procedury P1 do P3 obliczano za pomocą opracowanego programu ASTY napisanego w języku FORTRAN 1900. Listę instrukcji programu oraz przykładowy wydruk wyników zamieszczono w dodatku D.2.

6. EKSPERYMENTY ZASADNICZE

6.1. Wprowadzenie

Celem eksperymentów zasadniczych był dobór, a następnie ustalenie zbioru optymalnych parametrów dziedziny reguł RF (Def.11. i 12.) oraz zbioru optymalnych parametrów syntezy dźwięków mowy (p-t 3. rozdz.1.5.2.). Eksperymenty zasadnicze podzielono na eksperymenty dotyczące optymalizacji parametrów syntezy (rozdz.6.3.2.) oraz doboru i optymalizacji parametrów reguł RF (rozdz.6.3.3.). Ze względu na dużą liczbę eksperymentów omówiono tylko najważniejsze i podano ich końcowe wyniki w postaci skal rangowych.

6.2. Metodyka realizacji eksperymentów zasadniczych

6.2.1. Generacja materiału eksperymentalnego (fraz syntetycznych)

Model syntezy oraz stosowany zbiór reguł realizacyjnych R omówiono w rozdziałach 3. i 4. Słownik fonemów P utworzono z samogłosek sylabicznych - tab.2.1. oraz 4 spółgłosek dźwięcznych - tab.6.1. Metodę doboru parametrów formantowych samogłosek podano w rozdz.2.2.1., natomiast parametry spółgłosek dobrano w oparciu o analizę teoretyczną i wyniki eksperymentów przeprowadzonych przez Kacprowskiego i Mikiela [31] oraz na podstawie własnych badań [37] , w trakcie których ustalono również współczynnik w_1 długości transientu (4.3.), (4.5.):

$$w_1 = 0.25 \qquad (6.1.)$$

Czasy trwania τ_i^j poszczególnych fonemów we frazie ustalono dla każdej frazy oddzielnie w oparciu o wyniki badań Frąckowiak-Richter [116, 117] . Modyfikator w_3 czasu trwania ostatniej samogłoski we frazie ustalono eksperymentalnie i wynosił on:

$$w_3 = 1.15 \qquad (6.2.)$$

Amplitudy A_F^j przyjęto dla wszystkich fonemów identyczne:

$$\forall_j \; A_F^j = 1 \qquad (6.3.)$$

i nie stosowano reguły RK (4.4.)

Tabela 6.1. Zespolone częstotliwości formantowe spółgłosek dźwięcznych stosowanych w eksperymentach zasadniczych[*)].

Spółgłoska	F1	F2	F3	F4
m	125 + j 200	95 + j 830	95 + j 2700	125 + j 3700
n	160 + j 300	125 + j 1230	125 + j 2800	140 + j 3600
j	60 + j 400	80 + j 2200	95 + j 2950	160 + j 3800
l	65 + j 370	95 + j 1500	100 + j 2760	175 + j 3500

[*)] Częstotliwość zespolona filtru piątego formantu była zarówno dla samogłosek, jak i spółgłosek stała i wynosiła F5 = 180 + + j 4500 Hz.

6.2.2. Materiał eksperymentalny[33)]

W badaniach stanowiących przedmiot omówionych w pracy eksperymentów zasadniczych stosowano zbiór SF pięciu fraz (4 frazy typu C-V-C-V oraz jedna fraza V-C-V-C-V):

$$SF\ (k \in \{1,\ldots,5\} : s_{Fk}) \qquad (6.4.)$$

gdzie po uwzględnieniu (4.1.):

$s_{F1} = \{e_1(-P_s, 9), e_2(+P_s, 2), e_3(-P_s, 10), e_4(-P_s, 1)\}$ / j ó l a /

$s_{F2} = \{e_1(-P_s, 7), e_2(+P_s, 2), e_3(-P_s, 9), e_4(-P_s, 4)\}$ / m ó j e /

$s_{F3} = \{e_1(-P_s, 9), e_2(+P_s, 4), e_3(-P_s, 7), e_4(-P_s, 3)\}$ / j ó m u /

$s_{F4} = \{e_1(-P_s, 7), e_2(+P_s, 3), e_3(-P_s, 10), e_4(-P_s, 1)\}$ / m ú l a /

$s_{F5} = \{e_1(-P_s,3), e_2(-P_s,10), e_3(+P_s,4), e_4(-P_s,9), e_5(-P_s,10)\}$ / u l ó j e /

$e_i(j) = p_j \in P$

$P\ (j \{1,\ldots,10\} : p_j) = p_1 = /a/, p_2 = /o/, p_3 = /u/, p_4 = /e/, p_5 = /i/, p_6 = /t/, p_7 = /m/, p_8 = /n/, p_9 = /j/, p_{10} = /l/$

33) Por. uwagi zawarte w rozdziale 1.5.3.2. oraz Zał.7, s.31.

6.2.3.Przygotowanie i ocena materiału eksperymentalnego

Frazy syntetyczne rejestrowano na taśmie magnetofonowej (rys.2.2.). W przypadku testów ocen punktowych stosowano technikę identyczną, jak w eksperymentach wstępnych (rozdz.2.3.). Przy przygotowywaniu materiału do testów porównań w parach (A-B lub A-C) każdą parę rejestrowano dwukrotnie, odstęp między próbkami w parze wynosił 0.5 s, odstęp między daną parą i jej powtórzeniem ok. 1 s, odstęp między kolejnymi dwójkami różnych par 3 s. Pary wgrywano w kolejności losowej przy zachowaniu zasady, aby każda próbka znajdowała się tyle samo na pierwszej, co na drugiej pozycji w parze. Odsłuchy realizowano w warunkach identycznych, jak podano w rozdz.2.3. z tym, że w testach porównań w parach zamiast kilkuosobowej, wyszkolonej ekipy słuchaczy oceny dokonywała 10-12 osobowa ekipa złożona ze studentów Wydziału Elektroniki o prawidłowym otologicznie słuchu, co stwierdzono na podstawie badań audiometrycznych. Przed przystąpieniem do właściwych pomiarów ekipa przeszła 8-godzinny trening (4 sesje treningowe).

Testy z udziałem kilkunastoosobowej ekipy odsłuchowej uznano za testy formalne i ich wyniki poddawano statystycznej analizie wiarygodności metodami omówionymi w rozdz.5. W dalszej analizie brano pod uwagę tylko wyniki tych testów, gdzie stwierdzono pewną zgodność słuchaczy w ocenie oraz pewną możliwość szeregowania. Wyniki (nielicznych zresztą) testów nie spełniających tego kryterium nie zostały w pracy zamieszczone.

6.2.4.Program eksperymentów zasadniczych

6.2.4.1.Eksperymenty zasadnicze I : Optymalizacja parametrów syntezy dźwięków mowy (związanych z pobudzeniem krtaniowym)

EZ I (1) - Dobór optymalnych parametrów kształtu impulsów pobudzenia krtaniowego dla pięciu typów funkcji kształtu impulsu (funkcje f_A do f_E - rys.2.4.).

EZ I (2) - Dobór optymalnej funkcji kształtu impulsów pobudzenia krtaniowego przy ustalonych w EZ I (1) optymalnych wartościach współczynników kształtu.

EZ I (3) - Optymalizacja parametrów kształtu dla optymalnej funkcji kształtu ustalonej w EZ I (2).

EZ I (4) - Dobór reguły RF_3 sinusoidalnej dewiacji F_0 (3.3.).

EZ I (5) - Optymalizacja parametrów reguły RF_3.

6.2.4.2. Eksperymenty zasadnicze II : Optymalizacja parametrów dziedziny podzbioru RF reguł realizacyjnych R

EZ II (1) - Dobór parametrów reguł akcentowych (reguły RF_4 (3.4.), RF_5 (3.5.), RT_3 (4.7.)).

EZ II (2) - Dobór parametrów reguły intonacyjnej RF_6 (3.6.),

EZ II (3) - Optymalizacja parametrów reguły RF_6.

EZ II (4) - Optymalizacja parametru "a" reguły RF_2 generacji podstawowego konturu częstotliwościowego (3.2.) oraz ustalenie w RF_6 funkcji intonacyjnych F_{k,α_r}, $r \in \{1,2\}$.

EZ II (5) - Optymalizacja parametrów reguły RF_1 generującej podstawowy kontur amplitudowy.

EZ II (6) - Optymalizacja parametrów reguł akcentowych RF_4, RF_5, RT_3.

6.3. Eksperymenty zasadnicze I i II

6.3.1. Uwagi wstępne

W Eksperymentach I parametry dziedziny podzbioru reguł RF ustalono w oparciu o wyniki eksperymentów EW1 do EW5 (rozdz.2) oraz wyniki eksperymentów wstępnych, dotyczących syntezy fraz według reguł RF, opublikowanych w pracy [118] i uzupełnionych dodatkowymi badaniami eksperymentalnymi[34].

[34] Opisu tych badań nie zamieszczono.

Parametry reguły RF_1 : $H = 0.1$, $H_3 = 0.75$, $H_2 = 0.2$, $T_N = 38$ ms, $T_o = 96$ ms, $T_{o1} = 0.4\ T_o$, $T_{o2} = 0.3\ T_o$

Parametry reguły RF_2 : $a = 0.6$

Parametry reguły RF_3 : $A_D = 1.2$, $F_D = 7$ Hz

Parametry reguły RF_4 : $A_a = 0.15$, $q = 1.7$

RF_5 oraz RF_6 nie stosowano.
Parametry reguł RK i RT podano w rozdz.6.2.1.

6.3.2. Opis i wyniki Eksperymentów I

6.3.2.1. Eksperyment EZ I (1)

W badaniach Rosenberga [75] oraz we wstępnych eksperymentach EW3 i EW4 arbitralnie założono, że optymalne wartości parametrów kształtu impulsu pobudzenia krtaniowego (t_o, t_c) nie zależą od typu funkcji kształtu i porównywano, w badaniach nad doborem optymalnej funkcji kształtu, różne funkcje z parametrami (t_o, t_c) uznanymi dla jednego typu funkcji za optymalne (np. w EW3 i EW4 badano (t_o, t_c) tylko dla f_C). W eksperymentach zasadniczych zrezygnowano z tego założenia i w ramach niniejszego eksperymentu EZ I (1) szukano optimum w zbiorze (t_o, t_c), odrębnie dla każdej z 5-ciu badanych funkcji kształtu.

Badany zbiór 11-tu kombinacji parametrów kształtu (t_o, t_c) (tab.6.1.) dobrano w oparciu o wyniki eksperymentu EW3 (rozdz. 2.5.3., rys.2.8.) tak rozmieszczając je na płaszczyźnie t_o X t_c, aby zachodziło względnie równomierne rozłożenie ocen w skali rangowej otrzymanej z testu A-B[35]. Zbiorem badanych funkcji były funkcje f_A, f_B, f_C, f_D i f_E pokazane na rys.2.4. Materiałem eksperymentalnym była fraza syntetyczna $s_{F2} \in SF$ (6.4.). Pięć jednastoelementowych zbiorów frazy s_{F2}:

---

35) To założenie wynika z przyjętej w rozdz.5 metody przejścia z macierzy preferencji na skalę rangową, gdzie w przypadku zbyt dużej nierównomierności w ocenach następuje delinearyzacja skali.

$$s_{F2}(f_A) = \left[i = j \in \{1,\ldots,11\} : s_{F2_i}(t_{o_i}, t_{c_j}, f_A) \right] \qquad (6.5.)$$

.

.

.

$$s_{F2}(f_E) = \left[i = j \in \{1,\ldots,11\} : s_{F2_i}(t_{o_i}, t_{c_j}, f_E) \right]$$

poddano, każdy oddzielnie, formalnym odsłuchom metodą testu A-B. Dla każdej pary liczba niezależnych ocen preferencyjnych wynosiła L = 24 (12 słuchaczy x 2 powtórzenia). Wyniki eksperymentu EZ I (1) w postaci ilorazowych skal rangowych z przedziału $<0; 10>$ podano w tabl.6.1.

Tabela 6.1. Wyniki ocen preferencyjnych pięciu funkcji kształtu badanych w EZ I (1) dla różnych wartości (t_o, t_c).

Nr	t_o	t_c	Funkcja kształtu impulsu				
			f_A	f_B	f_C	f_D	f_E
1	0.6	0.19	3.1	4.27	4.17	6.8	6.22
2	0.41	0.06	7.3	10.0	6.08	8.8	8.35
3	0.41	0.2	4.5	9.48	10.0	10.0	8.65
4	0.41	0.4	0.0	1.96	5.0	0.0	2.53
5	0.26	0.18	10.0	7.63	7.26	10.0	10.0
6	0.11	0.3	9.18	7.11	9.21	8.6	9.87
7	0.11	0.8	7.4	6.8	6.52	9.03	8.75
8	0.45	0.45	2.16	0.0	1.99	1.6	0.0
9	0.35	0.25	5.9	5.46	5.9	7.2	7.24
10	0.11	0.45	7.0	7.11	6.24	5.27	6.18
11	0.05	0.05	1.9	0.72	0.0	2.40	1.22

===== pierwsze miejsce

——— drugie miejsce

6.3.2.2. Eksperyment EZ I (2).

Badano 10-elementowy zbiór fraz syntetycznych S_{F2} utworzony z próbek, które w EZ I (2) otrzymały, dla danej funkcji kształtu, najwyższe oceny (po 2 najlepsze próbki dla każdej funkcji kształtu):

$$S_{F2} = \{s_{F2_5}(f_A),\ s_{F2_6}(f_A),\ s_{F2_2}(f_B),\ s_{F2_3}(f_B),\ s_{F2_3}(f_C),\ s_{F2_6}(f_C),\ s_{F2_3}(f_D),\ s_{F2_5}(f_D),\ s_{F2_5}(f_E),\ s_{F2_6}(f_E)\} \quad (6.6.)$$

gdzie:

$s_{F2_i}(f_A \ldots f_D)$ - por. (6.5.) oraz tab.6.1.

Wyniki testu A-B, przeprowadzonego dla zbioru (6.6.), podano w tab.6.2. (liczba niezależnych ocen L = 24).

Tabela 6.2. Wyniki ocen preferencyjnych dla pięciu funkcji kształtu badanych w EZ I (2) dla dwóch najlepszych kombinacji parametrów (t_o, t_c).

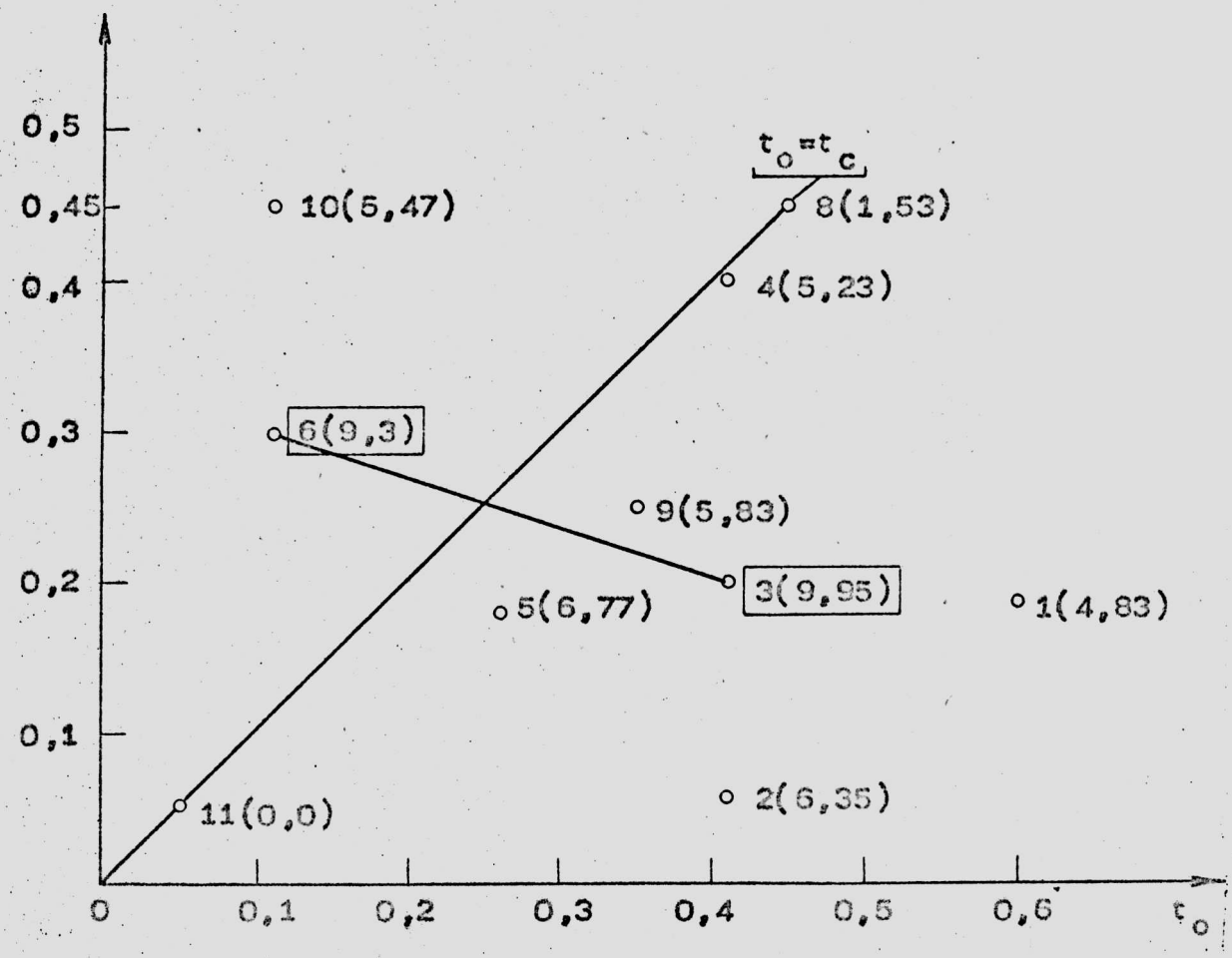
Nr	Funkcja kształtu	t_o	t_c	Wyniki testu A-B
1	f_A	0.26	0.18	5.22
2	f_A	0.11	0.3	6.02
3	f_B	0.41	0.06	8.12
4	f_B	0.41	0.2	9.5
5	f_C	0.41	0.2	10.0
6	f_C	0.11	0.3	8.88
7	f_D	0.41	0.2	2.31
8	f_D	0.26	0.18	0.0
9	f_E	0.26	0.18	7.39
10	f_E	0.11	0.3	6.27

6.3.2.3. Eksperyment EZ I (3)

Dla optymalnej funkcji kształtu impulsu pobudzenia krtaniowego - funkcja f_c (por. EZ I (2), tab.6.2.) - ponownie przebadano wpływ parametrów (t_o, t_c) na naturalność brzmienia. Zbiór kombinacji parametrów (t_o, t_c) był identyczny, jak w EZ I (1). Materiałem eksperymentalnym był podzbiór Sb(SF) czterech fraz syntetycznych:

$$Sb(SF) = \{s_{F1}, s_{F2}, s_{F3}, s_{F4}\} \qquad (6.7.)$$

Rozkład badanych kombinacji (t_o, t_c) na płaszczyźnie t_o X t_c pokazano na rys.6.1., a wyniki przeprowadzonego testu A-B podano w tab.6.3.



Rys.6.1. Rozkład wartości 11 kombinacji parametrów (t_o, t_c) badanych w eksperymentach EZ I (1) i EZ I (3).
(W nawiasach obok numeru kombinacji - tab.6.1. i 6.3. - podano średnie wartości ocen preferencyjnych x_{AB} uzyskanych w EZ I (3) przez daną kombinację).

Tabela 6.3. Wyniki ocen preferencyjnych dla optymalnej funkcji kształtu f_C badanej w EZ I (3) dla różnych kombinacji parametrów (t_o, t_c).

Nr	t_o	t_c	Zbiór fraz				
			s_{F1}	s_{F2}	s_{F3}	s_{F4}	$\bar{x}_{AB}$
1	0.6	0.19	4.6	4.17	4.55	6.0	4.83
2	0.41	0.06	6.1	6.08	7.82	5.4	6.35
3	0.41	0.2	10.0	10.0	10.0	9.8	9.95
4	0.41	0.4	5.92	5.0	5.3	4.71	5.23
5	0.26	0.18	7.5	7.26	6.04	6.3	6.77
6	0.11	0.3	8.65	9.21	9.44	10.0	9.3
7	0.11	0.8	4.82	6.52	7.85	6.02	6.3
8	0.45	0.45	0.66	1.99	2.97	0.52	1.53
9	0.35	0.25	4.47	5.9	7.06	5.9	5.83
10	0.11	0.45	6.58	6.24	5.48	3.58	5.47
11	0.05	0.05	0.0	0.0	0.0	0.0	0.0

===== pierwsze miejsce

——— drugie miejsce

$\bar{x}_{AB}$ - średnia ocena w zbiorze fraz

Celem zbadania możliwości uśrednienia ocen preferencyjnych w zbiorze fraz obliczono, dla każdej dwuelementowej kombinacji (s_{Fi}, s_{Fj}) w zbiorze Sb(SF) wyrażonym zależnością (6.7.), współczynnik korelacji skal rangowych Spearmana:

$$r_s = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \qquad (6.8.)$$

gdzie:

$d_i = x_i - y_i$

x_i - i-ta ranga w pierwszej skali

y_i - i-ta ranga w drugiej skali

n - liczność zbioru ocen na skali x i y.

następnie testowano hipotezę:

$$H_o: \; r_s = 0 \qquad (6.9.)$$

Dla skal s_{F1} do s_{F2} z tab.6.3. hipoteza (6.9.) została odrzucona na poziomie istotności podanym w tab.6.4. W nawiasach poniżej poziomu istotności podano wartość r_s dla danej kombinacji (s_{Fi}, s_{Fj}).

Tabela 6.4. Poziomy istotności α, na których dla skal z tab.6.3. odrzucono hipotezę (6.9.)

	s_{F1}	s_{F2}	s_{F3}	s_{F4}
s_{F1}	-	0.0001 (0.91)	0.008 (0.74)	0.012 (0.72)
s_{F2}		-	0.0002 (0.9)	0.0013 (0.84)
s_{F3}			-	0.0026 (0.81)
s_{F4}				-

6.3.2.4. Eksperyment EZ I (4)

W eksperymencie EZ I (4) jako dane przyjęto optymalną funkcję kształtu impulsu f_C z optymalną kombinacją parametrów $(t_o, t_c) = (0.41 ; 0.2)$.

Dla zbioru Sb (SF) czterech fraz syntetycznych

$$Sb\,(SF) = \{s_{F1}, \ldots, s_{F4}\} \qquad (6.10.)$$

badano, czy i ewentualnie dla jakich wartości parametrów A_D i F_D reguła RF_3 niskoczęstotliwościowej dewiacji F_o poprawia naturalność brzmienia. Dla każdej frazy ze zbioru (6.10.) wygenerowano jej 20 wariantów z różnymi wartościami parametrów $(A_D ; F_D)$. Wartości badanych kombinacji (A_D, F_D) podano w tab.6.5. Cztery 20-to elementowe zbiory fraz poddano badaniom odsłuchowym metodą testu A-C. Próbką odniesienia C była fraza,

gdzie nie stosowano reguły RF_3. W EZ I (4) liczba niezależnych ocen wynosiła L = 24. Wynik testowania hipotezy:

$$H_o : \bar{x}(S_i) > \bar{x}(C) \qquad (6.11.)$$

podano w tabeli 6.5. Kryterium przyjęcia hipotezy (6.11.) było jej spełnienie (w sensie statystycznym - por. 5.2.3.1.) dla co najmniej trzech z czterech badanych dla danej kombinacji (A_D, F_D), fraz syntetycznych.

Tabela 6.5. Wyniki testu A-C dla różnych kombinacji parametrów (A_D, F_D) reguły RF_3.

Nr	A_D	F_D	Hipoteza H_o	Nr	A_D	F_D	Hipoteza H_o
1*	0.7	3.0	przyjęta	11*	1.2	9.0	przyjęta
2*	1.2	3.0	przyjęta	12	2.5	9.0	odrzucona
3	2.5	3.0	odrzucona	13	4.0	9.0	odrzucona
4	4.0	3.0	odrzucona	14	8.0	9.0	odrzucona
5	8.0	3.0	odrzucona	15*	0.7	15.0	przyjęta
6	1.2	6.0	odrzucona	16	1.2	15.0	odrzucona
7*	1.7	6.0	przyjęta	17	2.5	15.0	odrzucona
8	2.5	6.0	odrzucona	18	4.0	15.0	odrzucona
9	4.0	6.0	odrzucona	19	8.0	15.0	odrzucona
10	8.0	6.0	odrzucona	20*	0.7	6.0	przyjęta

*wartości parametrów A_D i F_D poddane badaniom w następnym eksperymencie EZ I (5).

6.3.2.5. Eksperyment EZ I (5)

W eksperymencie EZ I (4) dla sześciu kombinacji (A_D, F_D) przyjęto hipotezę, że zastosowanie reguły RF_3 poprawia naturalność brzmienia fraz syntetycznych. W eksperymencie EZ I (5) metodą testu A-B badano, która z tych kombinacji jest optymalna. Dla każdej z trzech fraz z podzbioru Sb (SF) wyrażonego zależnością:

$$Sb(SF) = \{s_{F1}, s_{F2}, s_{F3}\} \qquad (6.12.)$$

utworzono 7-elementowe zbiory, złożone z wariantów frazy o sześciu kombinacjach (A_D, F_D) oraz z próbki odniesienia C. Liczba niezależnych ocen wynosiła, dla każdej pary w teście A-B, L = 24.
Wyniki eksperymentu EZ I (5) podano w tab.6.6.

Tabela 6.6. Wyniki testu A-B (oceny preferencyjne) dla zbioru sześciu kombinacji (A_D, F_D) badanych w eksperymencie EZ I (5).

Nr	A_D	F_D	Zbiór fraz			
			s_{F1}	s_{F2}	s_{F3}	$\bar{x}_{AB}$
1	0.0	-	0.0	0.72	0.0	0.24
2	0.7	3.0	7.97	7.36	6.75	7.36
3	1.2	3.0	8.98	8.06	10.0	9.0
4	0.7	6.0	1.85	0.0	1.22	1.02
5	1.7	6.0	10.0	10.0	9.76	9.92
6	1.2	9.0	5.0	3.15	2.23	3.46
7	0.7	15.0	3.93	2.31	3.42	3.22

$\bar{x}_{AB}$ - średnia ocena w zbiorze fraz.

Celem sprawdzenia, czy jest dopuszczalne uśrednienie skal rangowych otrzymanych w EZ I (5) dla fraz s_{F1} do s_{F3} (tab.6.6.), obliczono, dla każdej kombinacji skal (s_{F_i}, s_{F_j}) w zbiorze Sb (SF) - zal. (6.12.), współczynnik korelacji r_s Spearmana. Hipoteza o braku korelacji między skalami s_{F1} do s_{F3} testowana w oparciu o wyliczoną wartość r_s, została dla wszystkich kombinacji odrzucona. Poziomy istotności α, na których odrzucono hipotezę o braku korelacji, oraz wyliczone wartości r_s podano w tab.6.7. (wartości r_s umieszczono w nawiasach).

Tabela 6.7. Poziomy istotności, na których w EZ I (5) odrzucono hipotezę o braku korelacji pomiędzy skalami rangowymi z tab.6.6. oraz wartości współczynnika korelacji r_s pomiędzy tymi skalami.

	s_{F1}	s_{F2}	s_{F3}
s_{F1}	-	0.00045 (0.96)	0.0068 (0.93)
s_{F2}	-	-	0.0025 (0.89)
s_{F3}	-	-	-

6.3.2.6. Uwagi końcowe dotyczące Eksperymentów I

1. Eksperymenty zasadnicze I, dotyczące optymalizacji paramentów syntezy dźwięków mowy związanych z pobudzeniem krtaniowym, nie wniosły istotnych zmian we wnioskach ustalonych w oparciu o wyniki eksperymentów wstępnych (rozdz.2.5.7.). Ponownie uzyskano 2 maksima naturalności brzmienia na płaszczyźnie parametrów kształtu t_o X t_c (rys.6.1.), najwyżej oceniano próbki generowane z użyciem funkcji f_B i f_C, itd.
2. Eksperymenty EZ I (1) do (5) wykazały, co ma ważne następstwa praktyczne, że zmiany naturalności brzmienia w funkcji parametrów kształtu impulsów pobudzenia krtaniowego nie zależą od ciągu elementów fonetycznych we frazie (por. współczynniki r_s w tab.6.4. i 6.7. dla skal, odpowiednio, z tab.6.3. i 6.5.).

6.3.3. Opis i wyniki Eksperymentów II

6.3.3.1. Eksperyment EZ II (1)

W pracach dotyczących akcentu zgodnie uznaje się, że akustycznymi korelatami akcentu jest wydłużenie akcentowanej samogłoski, podwyższenie jej poziomu oraz podwyższenie częstotliwości podstawowej w obrębie akcentowanej sylaby [49, 51, 54, 65, 103, 119-120] . Trójwymiarowość akcentu przyjmowana jest bez zastrzeżeń we wszystkich przytoczonych pracach, nie ma natomiast zgodności co do wagi jego poszczególnych korelatów. Jassem i in. [54] badając percepcję akcentu dla syntetycznych bodźców mowopodobnych wykazał, że w mowie polskiej o wrażeniu akcentu głównie decyduje podwyższenie F_0 (wartość progowa podwyższenia 9 %), aczkolwiek wydłużenie czasowe i podwyższenie intensywności (o wartościach większych, odpowiednio, od 20 % i 6 dB) również wywołują wrażenie akcentu. Podobne wyniki dla naturalnej mowy angielskiej uzyskali Brown i McGlone [49] oraz Cheung i in. [120] . Nowakowska [65] eksperymentalnie dowiodła, że do prawidłowej identyfikacji akcentu wymagana jest analiza nie tylko zmian w przebiegu F_0, ale również intensywności i wydłużenia czasowego (decyduje jednak F_0).

Z omówionych prac wynika:

1. Wrażenie akcentu można wywołać przez szczególne wyróżnienie akcentowanego fragmentu w dziedzinie częstotliwości (podwyższenie F_0), energii (podwyższenie intensywności) lub czasu (wydłużenie czasowe).
2. W mowie naturalnej akcent jest sygnalizowany równocześnie w trzech wymienionych w p-cie 1. dziedzinach.
3. Wagę poszczególnych korelatów akcentu można następująco uszeregować według malejącego znaczenia: podwyższenie F_0 - podwyższenie intensywności - wydłużenie czasowe.

W eksperymencie EZ II (1) jako cel przyjęto ustalenie, które kombinacje wartości parametrów (A_a, A_f, q), występujących w regułach akcentowych RF_4 (3.4.) i RF_5 (3.5.), zapewniają jednoznaczne uzyskanie wrażenia akcentu, a następnie które kombina-

cje spełniające to kryterium zapewniają w opinii słuchaczy, uzyskanie naturalnie brzmiącego akcentu[36]. Pierwszy etap eksperymentu przeprowadzono metodą testu M_1 (por.rozdz.5.2.2.1.), a drugi w oparciu o test A-B. Badano wartości (A_a, A_f, q) z przedziału:

$$A_a \in \{0 ; 0.45\}$$
$$A_f \in \{0 ; 35\} \text{ Hz} \qquad (6.13.)$$
$$q \in \{1 ; 1.15\}$$

Pozostałe parametry reguł RF były identyczne jak w Eksperymentach I (rozdz.6.3.1.). W EZ II (1) oraz w pozostałych Eksperymentach II stosowano optymalną funkcję kształtu f_0 z optymalną parą parametrów $(t_o, t_c) = (0.41 ; 0.2)$ oraz $A_D = 1.7$, $F_D = 6$ Hz. Ocenie metodą testu M_1 poddano 26 wariantów frazy s_{F1} z różnymi kombinacjami wartości (A_a, A_f, q) zmienianymi w przedziałach (6.13.). Następnie w etapie drugim badano 10 wariantów frazy s_{F2} z parametrami (A_a, A_f, q), dla których w etapie pierwszym uzyskano najlepsze rezultaty (tzn. najwyższe wartości estymatora $\tilde{P}$ ($s_{F1} \in S [+ P_s]$), gdzie: $S [+ P_s]$ zbiór fraz z subiektywnie odbieraną cechą akcentu, P - procent głosów, $P > 0.5$ (por. odsyłacz [30]).

Wyniki eksperymentu EZ II (1) wykazały:

1. Preferowane są wartości $A_a \approx 0.1$ [37].
2. Niskie oceny otrzymują próbki z samym podwyższeniem częstotliwościowym (tzn. próbki gdzie $A_f \neq 0$, $A_o = 0$, $q = 1$).
3. W odniesieniu do wartości wydłużenia czasowego q nie uzyskano jednoznacznych rezultatów.

---

[36] A_a - odpowiada podwyższeniu intensywności.
A_f - odpowiada podwyższeniu częstotliwości podstawowej F_o.
q - odpowiada wydłużeniu czasowemu.

[37] Stosowana w syntezie reguła RF_2 zapewniała dla $A_a = 0$ wzrost częstotliwości podstawowej, np. dla $A_a = 0.1$ wzrost F_o wynosił ok. 9 Hz przy a = 0.6 i F0 = 120.

6.3.3.2. Eksperyment EZ II (2)

W eksperymencie EZ II (2) jako dane przyjęto:
1. Parametry syntezy jak w EZ II (1).
2. Parametry reguł RF, RK i RT jak w EZ II (1).

Dla frazy $s_{F1} \in SF$ (6.4.) badano 54 warianty pytających funkcji intonacyjnych F_{k, α_1} oraz 24 warianty twierdzących funkcji intonacyjnych F_{k, α_2} [38]. Zbiory funkcji intonacyjnych:

$$F_{k,1} \; (i \in \{1,\ldots 54\} : F^i_{k, \alpha_1}) \qquad (6.14.1.)$$

$$F_{k,2} \; (i \in \{1,\ldots 24\} : F^i_{k, \alpha_2}) \qquad (6.14.2.)$$

utworzono na podstawie konturów intonacyjnych, dla których w badaniach Majewskiego i Blasdell'a [64] oraz Studdert-Kennedy i Hadding [79] wygenerowane frazy syntetyczne zostały przez 90 % populacji słuchaczy zakwalifikowane do zbioru pytań lub twierdzeń. Dodatkowo uwzględniono kontury intonacyjne podane przez Renowskiego [56, 121], uzyskane w oparciu o analizę mowy naturalnej.

Ze względu na brak jednoznacznych danych dotyczących akcentu oraz ze względu na występowanie w większości konturów pytających tzw. punktu zwrotnego z lokalnym maksimum F_o (por. kontury 11-30, 34-36, 40-42, 46-48, 52-54, pokazane na rys.3.1. i 3.2. w dodatku D.3.) regułę akcentową RF_5 (3.5.) wykorzystano łącznie z RF_6 do kształtowania konturu intonacyjnego. Reguły RF_4 oraz reguły RF_2 nie stosowano. Zatem w EZ II (2) kontur $F_{oc}(t)$ w obrębie frazy kształtowano wyłącznie funkcją intonacyjną F_{k, α_2} oraz lokalnym podwyższeniem częstotliwości uzyskiwanym przez zastosowanie RF_5.

---

38) W pracy rozważano tylko dwa podstawowe typy intonacji, (pytającą i twierdzącą), gdyż w mowie polskiej dla pewnych struktur składniowych różnicują one interpretację znaczenia zdań [56, 64].

Za pomocą F_{k,α_r} kształtowano również narastanie $F_0(t)$ na początku frazy, gdyż brak tej cechy pogarszałby naturalność brzmienia fraz syntetycznych (por. eksperymenty wstępne - rozdz.2.). Przykłady konturów intonacyjnych badanych w EZ II (2) pokazano na rys.6.2. i 6.3. Pozostałe kontury ze zbioru $F_{k,1}$ (6.14.1.) i $F_{k,2}$ (6.14.2.) pokazano na rys.3.1. i 3.2. zamieszczonych w dodatku D.3.

W omawianym eksperymencie kilkuosobowa grupa wyszkolonych słuchaczy miała za zadanie ocenić:

a) które z prezentowanych fraz są pytaniami (twierdzeniami),

b) w przypadku pozytywnej decyzji $s_{F1}(F^i_{k,\alpha_r}) \in S(+\alpha_r)$

gdzie:

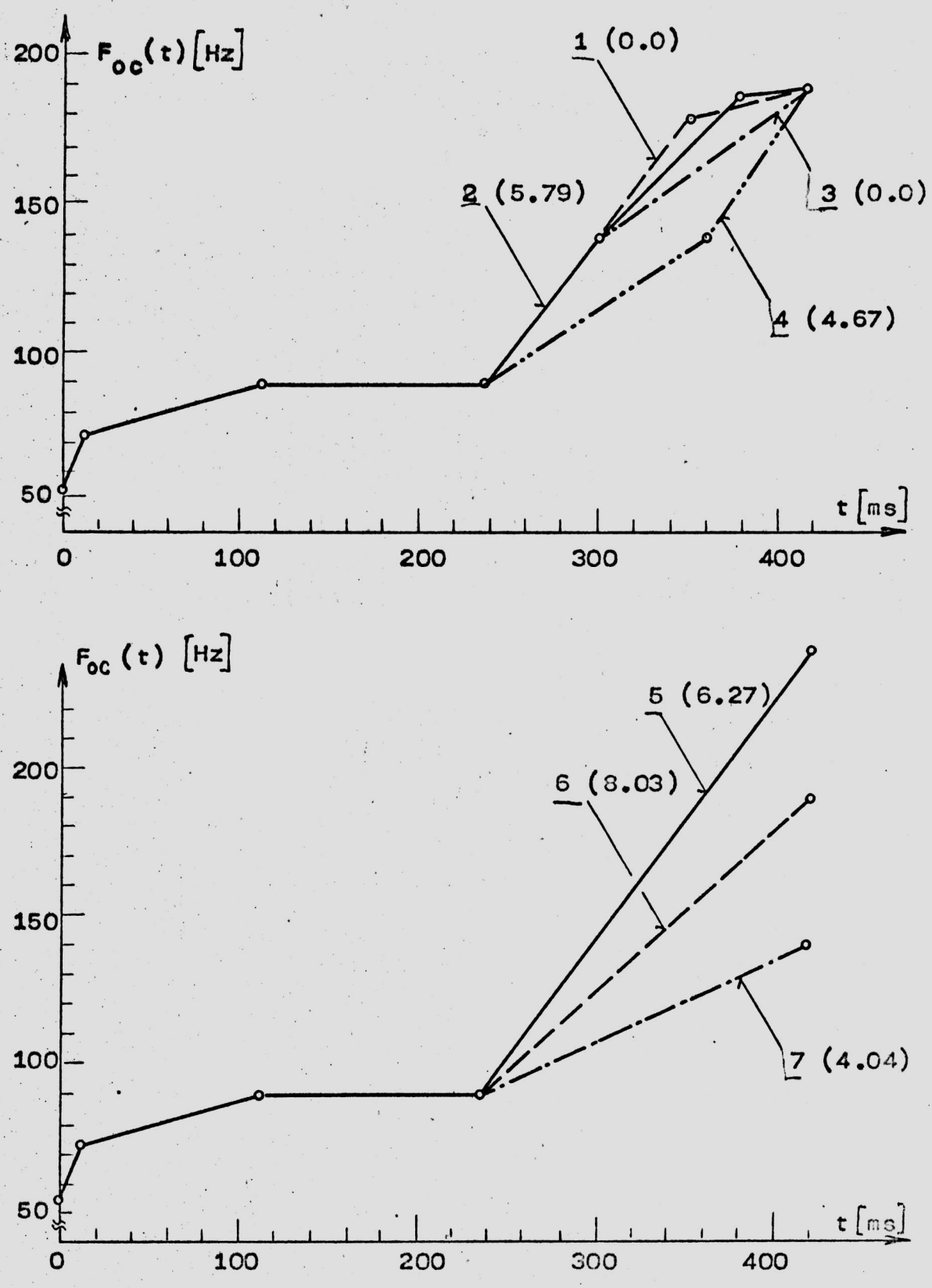
$s_{F1}(F^i_{k,\alpha_r})$ - fraza s_{F1} wygenerowana z zastosowaniem F^i_{k,α_r}

$S(+\alpha_r)$ - zbiór fraz odbieranych subiektywnie jako pytania ($r = 1$) lub twierdzenia ($r = 2$)

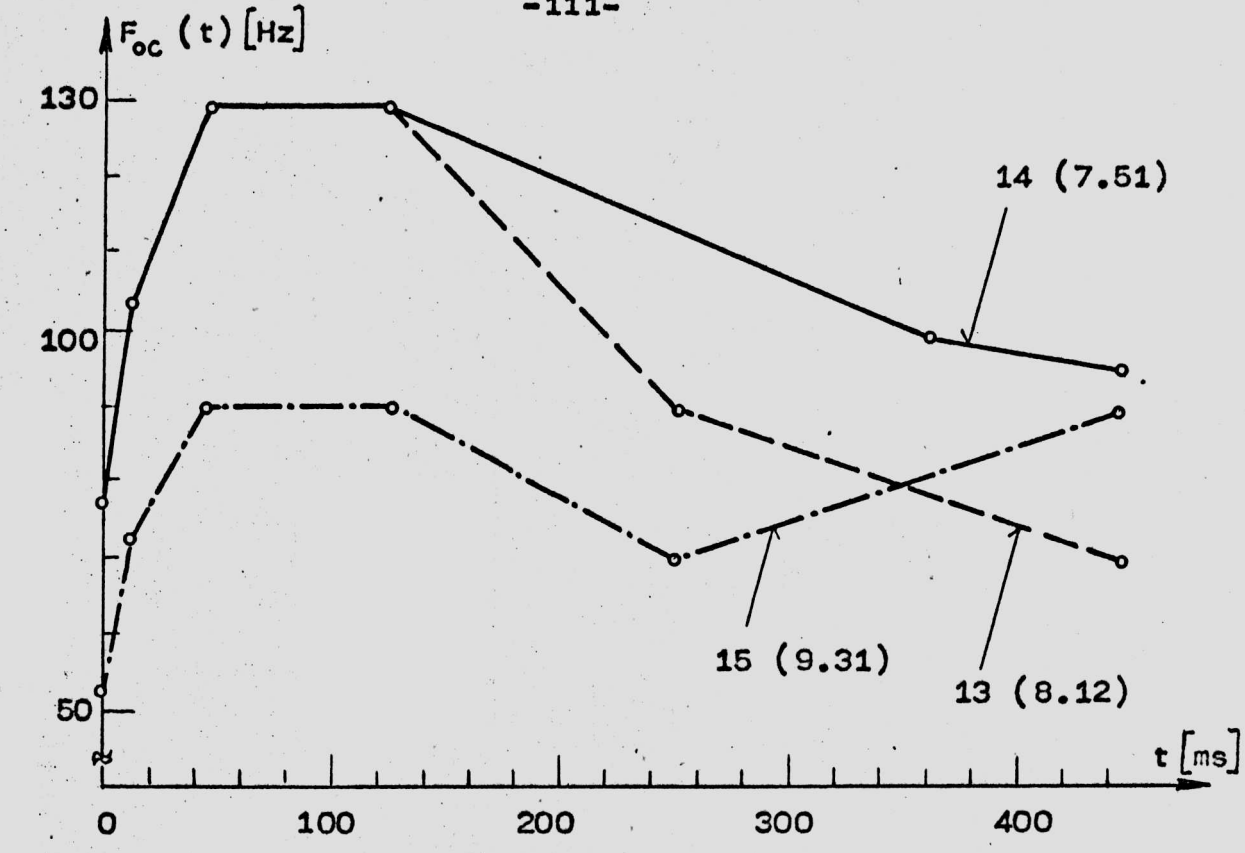
oceniano naturalność pytania (twierdzenia) metodą testu ocen punktowych.

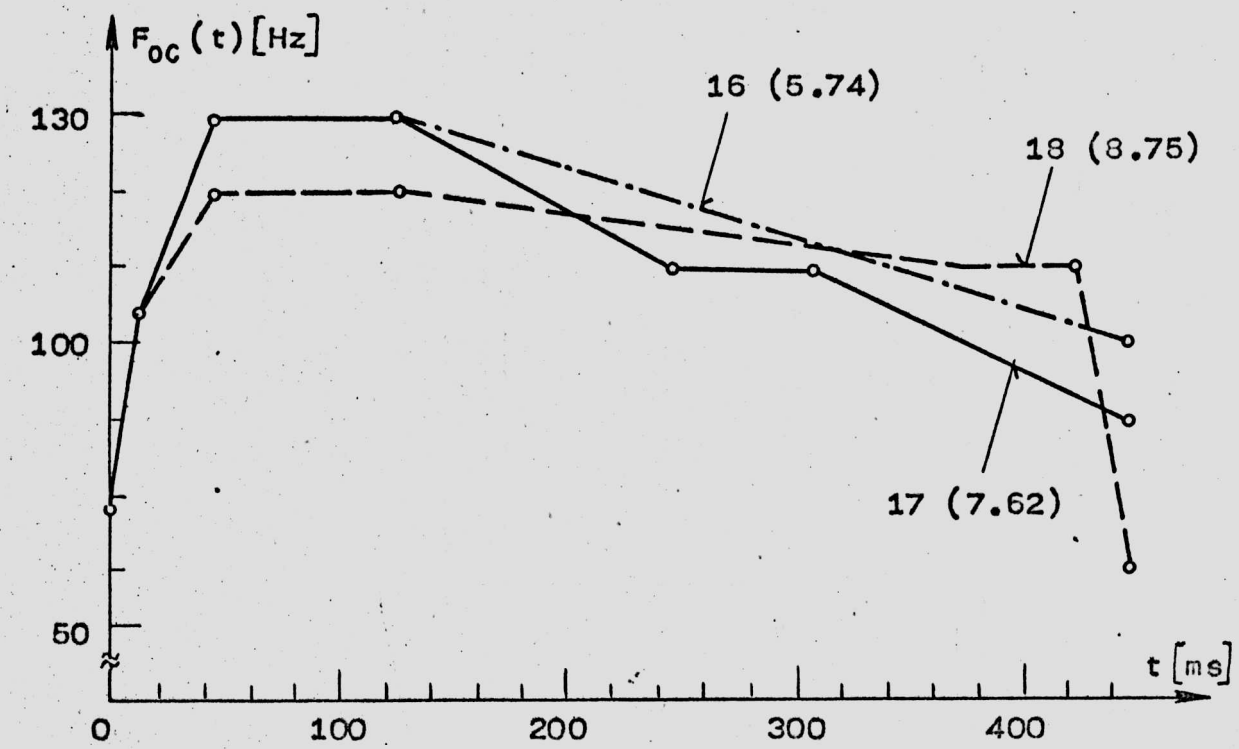
W wyniku testu M_1 (punkt a) z 54 konturów pytających odrzucono 21 o numerach: 1, 3, 11, 19, 20, 21, 23, 24, 27, 33, 34, 35, 36, 37, 39, 40, 41, 42, 43, 49, 54. Frazy generowane z tymi konturami odbierano jako twierdzenia, zawołania, frazy wykrzyknikowe, frazy ze zdziwieniem itp. Wszystkie 24 kontury oznajmiające były odbierane prawidłowo.

Do dalszych badań wybrano po 4 kontury z grupy pytań i twierdzeń, dla których frazy syntetyczne $s_{F1}(F^i_{k,\alpha_r})$ otrzymały najwyższe oceny. Wybrane kontury pokazano na rys. 6.4.a) oraz 6.4.b).

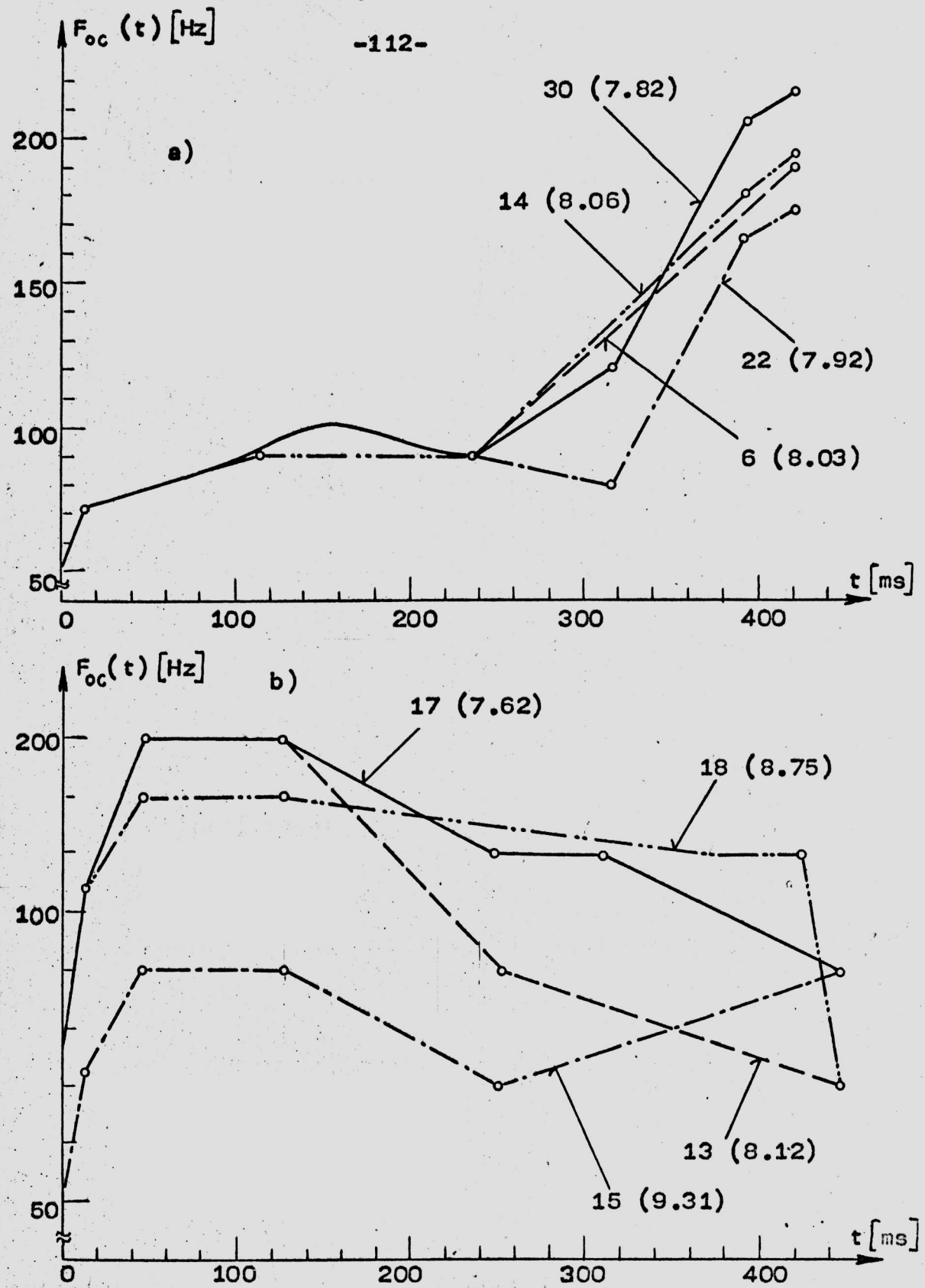


Rys.6.2. Przykłady pytających konturów intonacyjnych badanych w eksperymencie EZ II (2). Obok numeru konturu podano w nawiasach uzyskaną ocenę punktową.





Rys.6.3. Przykłady twierdzących konturów intonacyjnych badanych w eksperymencie EZ II (2). Obok numeru konturu podano w nawiasach uzyskaną ocenę punktową.



Rys.6.4. Kontury intonacyjne uznane w eksperymencie EZ. II (2) za optymalne: a) kontury pytające b) kontury twierdzące.

6.3.3.3.Eksperyment EZ II (3)

W eksperymencie EZ II (3) badano zbiór 11 konturów pytających i 11 konturów twierdzących. Każdy z tych zbiorów składał się z czterech konturów intonacyjnych uznanych w EZ II (3) za optymalne oraz z akcentowych wariantów tych konturów (łącznie 8 konturów). Dodatkowo do zbioru wchodziły 3 kontury stanowiące pośrednie warianty 4-ch konturów z EZ II (2). Warianty akcentowe konturów otrzymywano przez zastosowanie reguł RF_4 , RF_5 i RT_5 o parametrach:

$\underline{RF_5}$: $A_a = 0.1$

$\underline{RF_6}$: $A_f = 8$ Hz

$\underline{RT_5}$: $q = 1.7$

Reguły RF_2 nie stosowano. Pozostałe parametry miały wartości jak w EZ II (2). Zbiór konturów $F_{oc}(t)$ badanych w EZ II (3) pokazano na rys.6.5. (kontury pytające) i 6.6. (kontury twierdzące).

Zbiory konturów intonacyjnych badano dla podzbioru Sb (SF) dwóch fraz syntetycznych typu C-V-C-V oraz V-C-V-C-V:

$$Sb\ (SF) = \{s_{F1}\ ,\ s_{F5}\} \qquad (6.15.)$$

Cztery 11-to elementowe zbiory fraz syntetycznych:

$$s_{F1}\ (F_{k,1}) = [i \in \{1,\ldots,11\} : s_{F1}\ (F^i_{k,\alpha_1})]$$

$$s_{F1}\ (F_{k,2}) = [i \in \{1,\ldots,11\} : s_{F1}\ (F^i_{k,\alpha_2})]$$

$$s_{F5}\ (F_{k,1}) = [i \in \{1,\ldots,11\} : s_{F5}\ (F^i_{k,\alpha_1})]$$

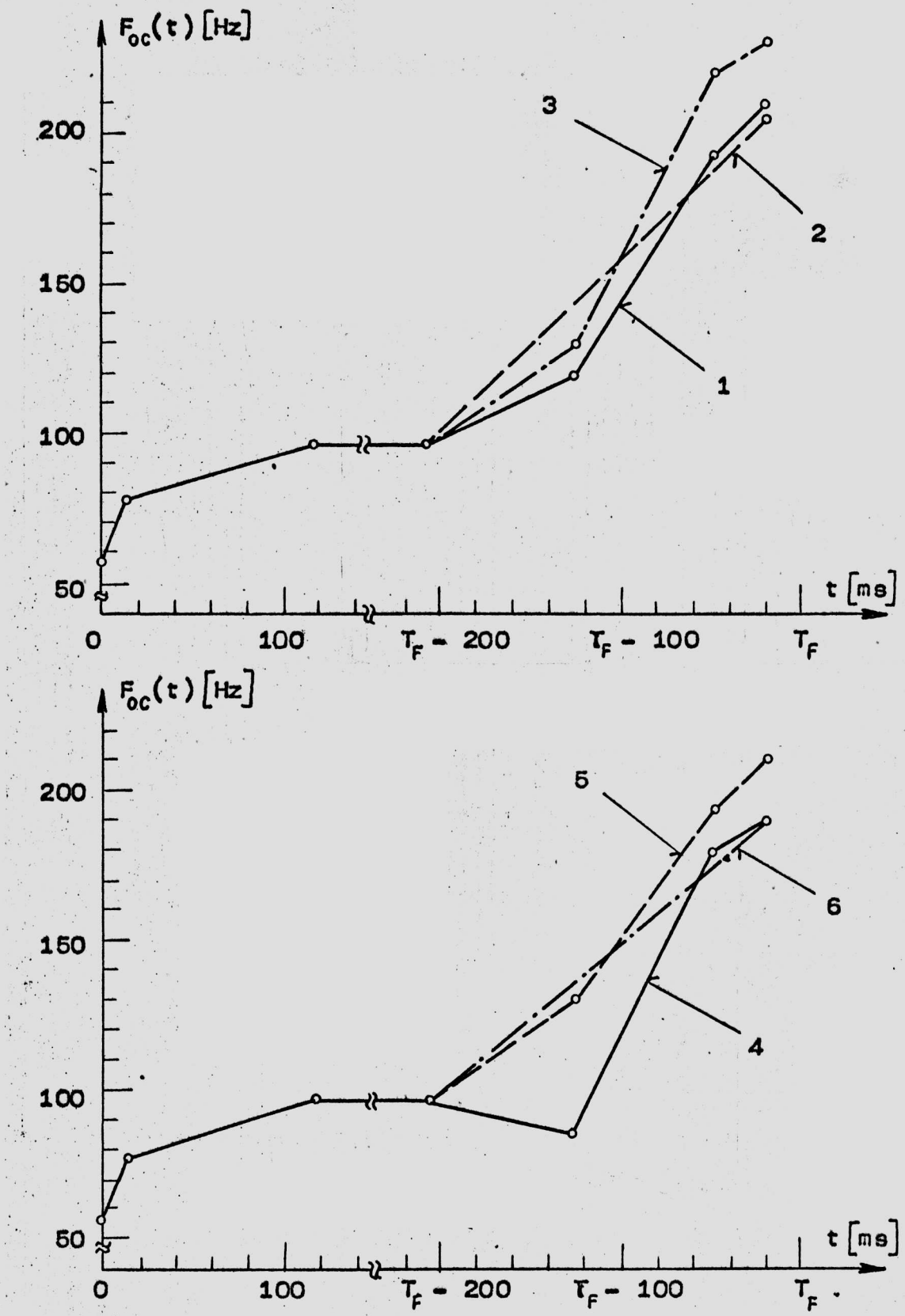
$$s_{F5}\ (F_{k,2}) = [i \in \{1,\ldots,11\} : s_{F5}\ (F^i_{k,\alpha_2})]$$

poddano ocenie subiektywnej metodą formalnego testu A-B (liczba niezależnych ocen L = 12). Wyniki tego testu w postaci skal ocen preferencyjnych podano w tab.6.7. Optymalny kontur pytający i twierdzący pokazano na rys.6.7.

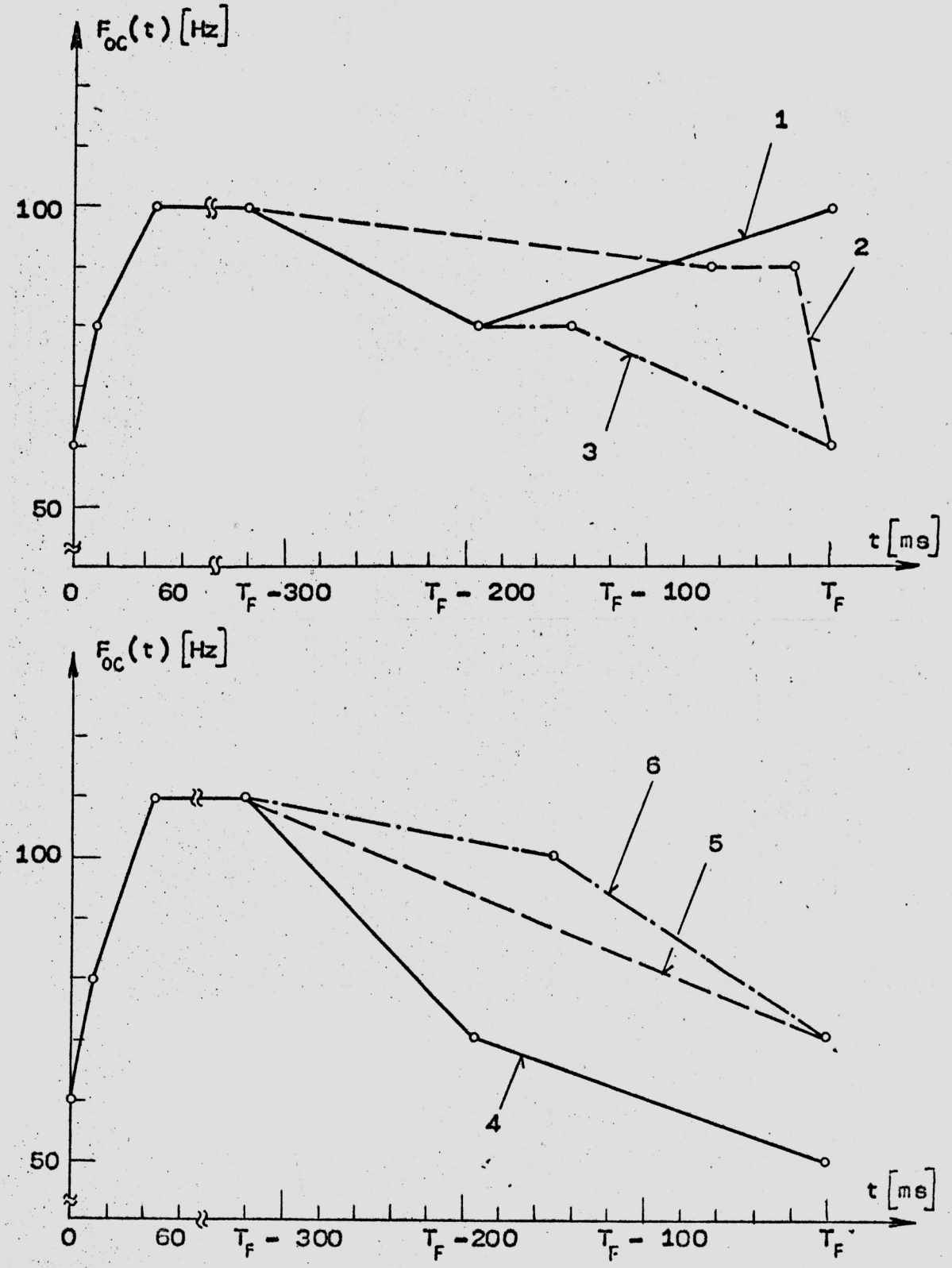
Tabela 6.7. Wyniki testu A-B dla konturów intonacyjnych badanych w eksperymencie EZ II (3).

Twierdzenia				Pytania			
Nr konturu	Akcent[1]	Oceny (w skali 0-10)		Nr konturu	Akcent[1]	Oceny (w skali 0-10)	
		s_{F5} V-C-V-C-V	s_{F1} C-V-C-V			s_{F5} V-C-V-C-V	s_{F1} C-V-C-V
1	-	0.0	3.14	1	-	7.33	9.16
1	+	1.54	4.57	1	+	10.0	10.0
2	+	2.25	0.88	2	+	0.05	2.93
2	-	2.48	3.58	2	-	0.0	5.21
3	-	9.32	7.43	3	-	0.45	0.0
3	+	7.36	7.34	3	+	1.86	1.72
4	+	7.48	9.93	4	+	7.15	9.35
4	-	10.0	10.0	4	-	7.29	9.72
5	-	4.43	5.04	5	-	4.93	9.87
5	+	3.87	5.37	5	+	3.49	7.09
6	-	0.1	0.0	6	-	9.36	5.17

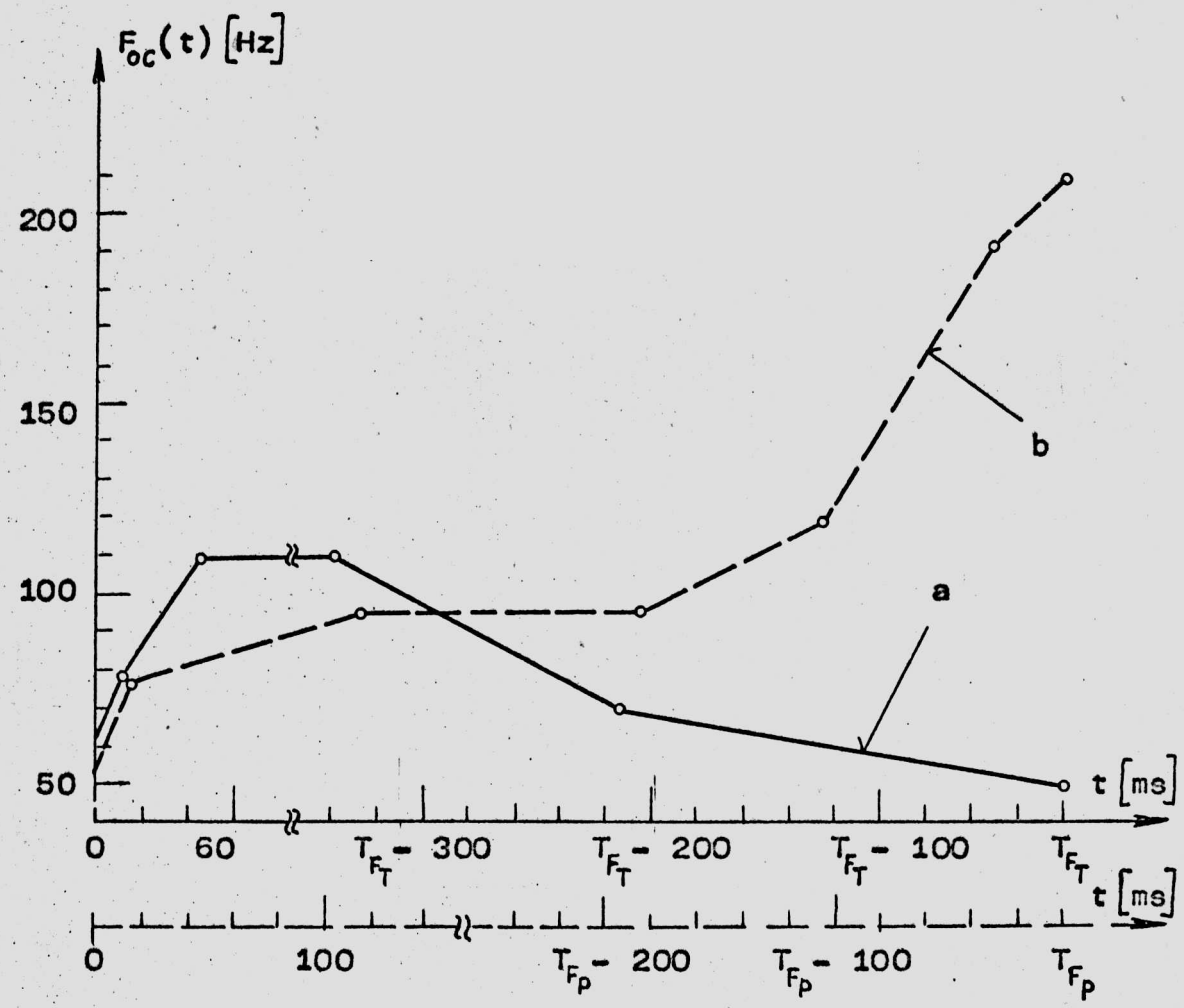
1) "+" - wariant akcentowany
"-" - wariant bez akcentu



Rys.6.5. Pytające kontury intonacyjne badane w eksperymencie EZ II (3).



Rys.6.6. Twierdzące kontury intonacyjne badane w eksperymencie EZ II (3).



Rys.6.7. Optymalny kontur intonacyjny dla frazy oznajmiającej (a) i pytającej (b).

Współczynnik korelacji r_s Spearmana, obliczony dla skal s_{F1} i s_{F5} z tab.6.7. (oddzielnie dla konturów pytających i twierdzących), wynosił:

kontury pytające $r_s\ (s_{F1}/s_{F5}) = 0.62$
kontury twierdzące $r_s\ (s_{F1}/s_{F5}) = 0.918$

Poziom istotności, na którym można odrzucić hipotezę o braku korelacji pomiędzy skalami s_{F1}/s_{F5} wynosił:

kontury pytające = 0.04
kontury twierdzące = 0.00007

6.3.3.4. Eksperyment EZ II (4)

W eksperymencie EZ II (2) i EZ II (3) badano kontury intonacyjne $F_{oc}(t)$ bez konsekwentnego stosowania reguł RF, np. w EZ II (2) regułę akcentową RF_5 stosowano łącznie z RF_6 do kształtowania konturu intonacyjnego. Reguły RF_2 w tych eksperymentach nie stosowano, gdyż wymagałoby to uwzględnienia jeszcze jednego parametru "a" występującego w tej regule. Celem eksperymentu EZ II (4) było ustalenie optymalnej wartości parametru "a" reguły RF_2 generacji podstawowego konturu częstotliwościowego (3.2.), przy równoczesnym ustaleniu funkcji intonacyjnych $F_{k,\alpha_1}(t)$ i $F_{k,\alpha_2}(t)$ występujących w regule RF_6 (3.6.). W eksperymencie EZ II (4) wynikowy kontur intonacyjny $F_{oc}(t)$ generowano stosując kolejno reguły $RF_1 \longrightarrow RF_2 \longrightarrow RF_6$. Badano wartości parametru "a" ze zbioru: {0.3, 0.4, 0.5, 0.6, 0.7, 1.0, 2.0, 100.0} . Dla każdej wartości parametru "a" wyliczano parametry funkcji intonacyjnych $F_{k,\alpha_r}(t)$ w ten sposób, aby w końcowym, intonowanym segmencie frazy (320 ms dla fraz oznajmiających i 180 ms dla fraz pytających) funkcje $F_{k,\alpha_r}(t)$ miały identyczny kształt jak optymalny kontur pytający (funkcja $F_{k,\alpha_1}(t)$) lub twierdzący (funkcja $F_{k,\alpha_2}(t)$), ustalony w EZ II (3) - rys.6.7. W pozostałych segmentach fraz (tzn. gdzie $F_{k,\alpha_r}(t) = \text{const}$) przebieg $F_0(t)$ generowała funkcja RF_2. Materiał eksperymentalny, technika jego oceny oraz liczba niezależnych ocen były identyczne jak w EZ II (3). Wyniki otrzymane

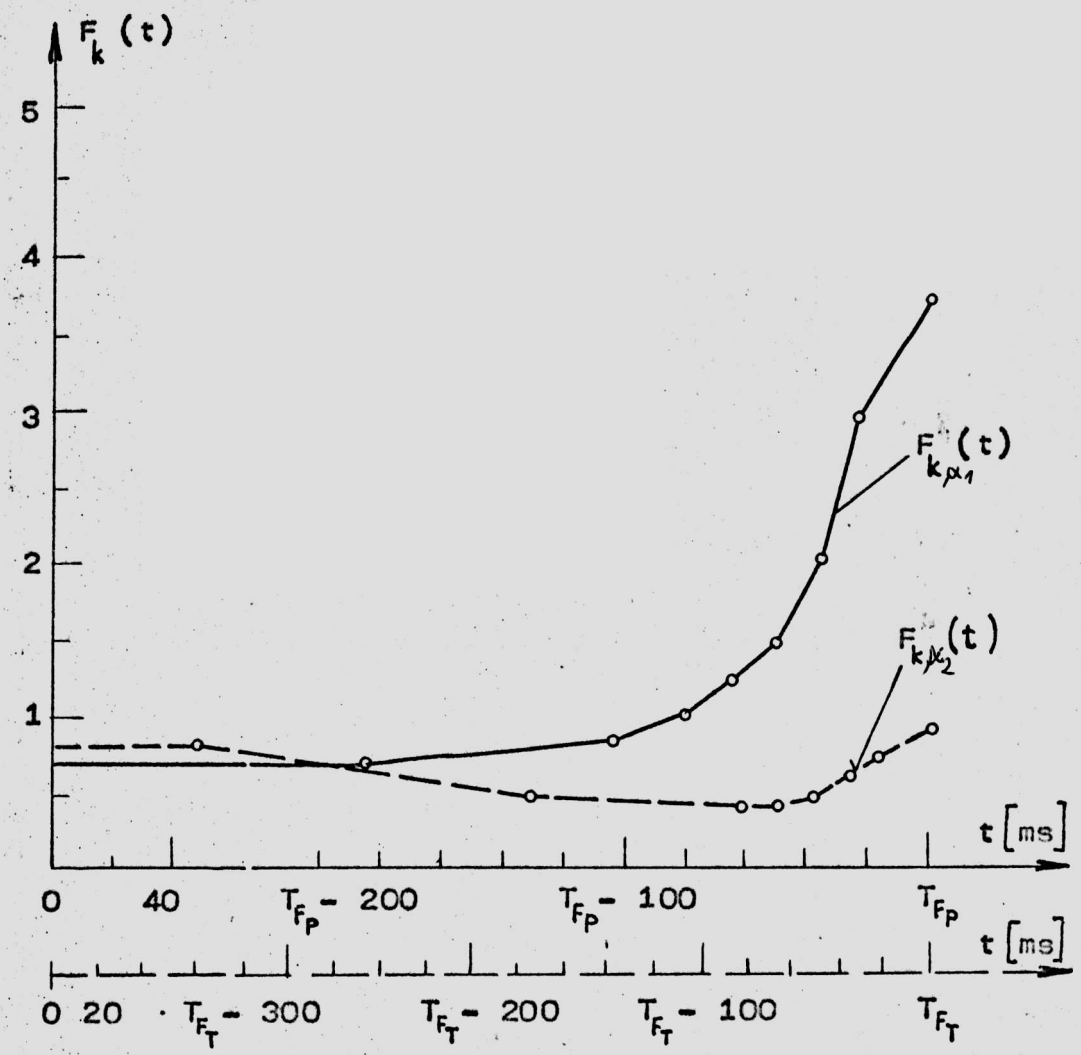
z testu A-B podano dla badanych fraz i badanego zbioru wartości parametru "a" w tab.6.8. Optymalną funkcję intonacji pytającej F_{k,α_1} oraz twierdzącej F_{k,α_2} pokazano na rys.6.8. Optymalna wartość parametru "a" reguły RF_2 wynosi 0.5.

Tabela 6.8. Wyniki testu A-B dla różnych wartości parametru "a" reguły RF_2 badanych w EZ II (4).

"a"	Oceny w skali 0-10			
	Frazy oznajmiające		Frazy pytające	
	s_{F5} V-C-V-C-V	s_{F1} C-V-C-V	s_{F5} V-C-V-C-V	s_{F1} C-V-C-V
0.3	0.0	2.42	1.47	0.0
0.4	6.75	4.30	3.86	3.54
0.5	10.0	10.00	10.0	9.49
0.6	8.76	8.20	4.12	5.3
0.7	7.79	6.73	4.04	2.77
1.0	2.47	0.0	0.0	4.4
2.0	5.52	5.88	5.76	3.37
100.0[1)]	8.77	8.99	7.68	10.0

[1)] Dla a = 100 funkcje intonacyjne F_{k,α_1} i F_{k,α_2} były identyczne jak optymalny kontur, odpowiednio, pytający i twierdzący otrzymany w EZ II (3).

Współczynniki korelacji r_s Spearmana dla skal z tab.6.8. oraz poziomy istotności α, na których można odrzucić hipotezę o braku korelacji pomiędzy tymi skalami, podano w tab.6.9.



Rys.6.8. Optymalna funkcja intonacji pytającej $F_{k,\alpha_1}(t)$ oraz twierdzącej $F_{k,\alpha_2}(t)$.

Tabela 6.9. Poziomy istotności α, na których można odrzucić hipotezę o braku korelacji między skalami z tab. 6.8. oraz wartości r_s dla tych skal (r_s podano w nawiasach poniżej α).

	s_{F1} oznajm.	s_{F5} oznajm.	s_{F1} pytaj.	s_{F5} pytaj.
s_{F1} oznajmiaj.	-	0.0003 (0.95)	0.01 (0.83)	0.03 (0.76)
s_{F5} oznajmiaj.	-	-	0,001 (0.92)	0.08 (0.64)
s_{F1} pytaj.	-	-	-	0.1 (0.59)
s_{F5} pytaj.	-	-	-	-

6.3.3.5. Eksperyment EZ II (5)

Celem eksperymentu EZ II (5) była optymalizacja parametrów H, T_N, T_o, H_2, H_3, T_{o1}, T_{o2} reguły generacji podstawowego konturu amplitudowego RF_1 (3.1.). W etapie pierwszym eksperymentu EZ II (5) badano parametry narastania konturu amplitudowego (H i T_N) przy pozostałych parametrach o wartościach podanych w rozdz.6.3.1. W drugim etapie badano parametry opadania konturu amplitudowego (T_o, T_{o1}, H_2, H_3) przy optymalnych wartościach H i T_N, ustalonych w etapie pierwszym. W obydwu etapach badano frazy s_{F1} i s_{F5} w wariancie oznajmiającym i pytającym (łącznie 4 frazy). Stosowano optymalne funkcje intonacyjne F_{k,α_1} i F_{k,α_2} oraz optymalną wartość parametru "a" reguły RF_2 (a = 0.5). W odsłuchach stosowano test A-B, gdzie liczba niezależnych ocen wynosiła L = 16 (2 sesje z udziałem 8 słuchaczy).

Etap pierwszy

Parametry H i T_N badano dla 9-ciu dwuelementowych kombinacji ich wartości (H_i, T_{Nj}) tworzących zbiór:

$$TH1\left[(i \in \{1,\ldots,3\}\ ,\ j \in \{1,\ldots,3\}\ :\ (H_i, T_{Nj})\right] \qquad (6.16)$$

gdzie:

$H_1 = 0.1$; $H_2 = 0.2$; $H_3 = 0.3$

$T_{N1} = 40$, $T_{N2} = 50$, $T_{N3} = 60$ [ms]

Otrzymano optymalne wartości kombinacji (H, T_N) różne dla fraz oznajmiających i pytających:

frazy pytające - $(H, T_N)'_{opt} = (0.2\ ,\ 60\ [ms]\ )$

frazy oznajmiające- $(H, T_N)''_{opt} = (0.1\ ,\ 50\ [ms]\ )$

Uwzględniając dodatkowo wyniki eksperymentów wstępnych przyjęto:

$$(H, T_N)_{opt} = (0.15\ ,\ 50\ [ms]\ ) \qquad (6.17.)$$

Etap drugi

Celem zmniejszenia liczby badanych parametrów, w tym etapie przyjęto: $T_{o1} = 0.4\ T_o$; $T_{o2} = 0.3\ T_o$ (por.rozdz.6.3.1.).

Badano parametry T_o, H_2, H_3 dla 9-ciu trzyelementowych kombinacji ich wartości podanych w tab.6.10., gdzie zamieszczono również wyniki testów A-B uśrednione w zbiorze badanych 4-ch fraz syntetycznych.

W etapie drugim przy generacji fraz syntetycznych nie stosowano reguły RF_2, gdyż to pociągałoby za sobą konieczność każdorazowej modyfikacji funkcji intonacyjnych przy każdej zmianie parametrów (T_o, H_2, H_3). Kontur intonacyjny oraz przebieg narastania konturu F_0 na początku frazy kształtowano funkcjami F_{k,α_r}.

Analiza otrzymanych wyników wskazuje, że są preferowane dłuższe czasy opadania oraz $H_3 = 0.75$ i $H_2 = 0.25$. Ponieważ te wyniki pokrywają się z wartościami uzyskanymi w badaniach cytowanych w rozdz.6.3.1., zatem za optymalne przyjęto podane tam wartości:

$$T_o = 96\text{ms},\ H_3 = 0.75\ ,\ H_2 = 0.2$$

To pozwoliło uniknąć przeliczania funkcji intonacyjnych F_k, α_r celem utrzymania stałego, optymalnego przebiegu konturów intonacyjnych dla nowych wartości parametrów opadania konturu A_o.

Tabela 6.10. Wyniki testu A-B dla dziewięciu kombinacji (T_o, H_2, H_3) badanych w drugim etapie eksperymentu EZ II (5).

Lp.	T_o [ms]	H_3	H_2	$\bar{x}_{AB}$
1	60	0.40	0.15	3.87
2	80	0.40	0.15	6.75
3	99	0.40	0.15	7.14
4	60	0.70	0.15	0.00
5	80	0.70	0.15	4.03
6	99	0.70	0.15	8.62
7	60	0.75	0.25	5.12
8	80	0.75	0.25	9.77
9	99	0.75	0.25	10.00

$\bar{x}_{AB}$ - średnia ocena w zbiorze fraz. Oceny średnie ponownie normowano w skali $\langle 0 ; 10 \rangle$.

6.3.3.6. Eksperyment EZ II (6)

W eksperymencie EZ II (6) stosowano:

1. Pełny zbiór reguł RF.
2. Regułę RF_1, RF_2, RF_3 i RF_6 z ustalonymi uprzednio optymalnymi parametrami.

Materiałem eksperymentalnym był zbiór fraz s_{F1} i s_{F5} w wariancie oznajmiającym i pytającym. Badano parametry reguł akcentowych A_a, A_f, q. W oparciu o wyniki eksperymentu EZ II (1) ustalono zbiór 8 trzyelementowych kombinacji wartości tych parametrów. Ocenę subiektywną przeprowadzono metodą testu A-B,

przy czym liczba niezależnych ocen wynosiła L = 16. Badane wartości parametrów A_o, A_f, q podano w tab.6.11., gdzie również zamieszczono wyniki testu A-B dla frazy V-C-V-C-V (s_{F5}) oraz C-V-C-V (s_{F1}).

Tabela 6.11. Wyniki testu A-B dla ośmiu kombinacji wartości parametrów A_a, A_f, q reguł akcentowych.

Lp.	A_a	A_f Hz	q	$\bar{x}(s_{F1})$	$\bar{x}(s_{F5})$
1	0.09	0.0	1	10.0	9.85
2	0.13	0.0	1	9.6	10.0
3	0.0	10.0	1	1.5	4.0
4	0.0	0.0	1	6.5	5.8
5	0.0	0.0	1.15	3.2	4.8
6	0.0	10.0	1.15	0.0	0.0
7	0.13	0.0	1.15	5.3	6.2
8	0.09	0.0	1.15	7.1	5.8

$\bar{x}(s_{F1})$, $\bar{x}(s_{F5})$ - średnie dla wariantu pytającego i oznajmiającego frazy. Średnie ponownie normowano w skali $\langle 0 ; 10 \rangle$

Dla wartości parametru $A_a \neq 0$ stosowanie reguły RF_2 powodowało równoczesne podwyższenie częstotliwości o 12 Hz dla $A_a = 0.13$ i o 8 Hz dla $A_a = 0.09$ (przy a = 0.5, F0 = 120 Hz).

Analiza wyników eksperymentu EZ II (6) wskazuje, że słuchacze najwyżej oceniali próbki z podwyższeniem amplitudowym 0.09 - 0.13 bez wydłużenia czasowego. Zdecydowanie najniższe oceny uzyskały próbki z podwyższeniem częstotliwościowym bez podwyższenia amplitudowego.

6.3.3.7. Uwagi dotyczące Eksperymentów II

1. Eksperymenty EZ II (2) i EZ II (3) wykazały, że realizacja intonacji pytającej i twierdzącej zachodzi w stosunkowo krótkim, końcowym odcinku fraz (180 ms dla intonacji pytającej i 320 ms dla intonacji twierdzącej). Zbliżone rezultaty uzyskał Renowski [56, 121] dla mowy naturalnej.
2. Stosunkowo nieduże zmiany w przebiegu konturu intonacyjnego w istotny sposób różnicują subiektywną ocenę intonacji (por. kontury 2 i 1 z rys.6.2. oraz 16 i 17 z rys.6.3.).
3. W subiektywnej ocenie akcentu słuchacze wyraźnie preferują dwuwymiarowy akcent uzyskiwany przez synchroniczne podwyższenie F_0 i A_0 (por. EZ II (6) - tab.6.11.).
4. Ocena intonacji oraz wpływu czasowego przebiegu $A_0(t)$ na naturalność fraz syntetycznych nie zależy w istotny sposób od typu frazy - skale ocen preferencyjnych uzyskane w EZ II (3) i EZ II (4) dla fraz typu V-C-V-C-V oraz C-V-C-V wykazują silną korelacj) (tab.6.9.). Silną korelację stwierdza się również między skalami ocen wpływu $A_0(t)$ na naturalność dla fraz o różnej intonacji. Można zatem postawić hipotezę, że ocena subiektywna zmian przebiegu $A_0(t)$ nie zależy ani od typu frazy, ani od typu intonacji.

7. PODSUMOWANIE

7.1. Optymalne parametry syntezy dźwięków mowy związane z pobudzeniem krtaniowym

1) Typ funkcji kształtu pobudzenia krtaniowego (por. rys.2.4.):

$$f_c(t) = \begin{cases} a/2\ (1 - \cos\ \pi\ t/T_o) & 0 \leqslant t < T_o \\ a\ .\ \cos\ \pi\ (t-T_o)/T_c & T_o \leqslant t \leqslant T_o + T_c \\ 0 & T_o + T_c < t \leqslant T \end{cases} \qquad (7.1.)$$

gdzie:

T - okres, T_o - czas narastania impulsu,
T_c - czas opadania impulsu pobudzenia.

2) Współczynniki kształtu (t_o, t_c):

$$t_o = T_o/T = 0.41$$
$$t_c = T_c/T = 0.2 \qquad (7.2.)$$

7.2. Optymalne parametry dziedziny reguł RF:

1. Parametry występujące w regule RF_1 (3.1.):

$H = 0.15$, $H_3 = 0.75$, $H_2 = 0.2$, $T_N = 50$ ms, $T_o = 96$ ms,
$T_{o1} = 0.4\ T_o$, $T_{o2} = 0.3\ T_o$

2. Parametry występujące w regule RF_2 (3.2.):

$a = 0.5$ $\quad F0 = 140$ Hz

3. Parametry występujące w regule RF_3 (3.3.):[39)]

$A_D = 1.7$ $\quad F_D = 6$ Hz

4. Parametry występujące w regule RF_4:

$A_a = 0.11$, $q = 1.7$

[39)] Sinusoidalna dewiacja F_o, ze względu na brak powiązania z poziomami syntezy mowy, winna należeć do zbioru parametrów syntezy dźwięków mowy. W pracy zaliczono ją do zbioru reguł RF i powiązano poprzez obligatoryjne zastosowanie z regułą RF_2 (por.7.3.). W eksperymentach parametry reguły RF_3 - A_D i F_D - traktowano jako parametry syntezy i optymalizowano je w ramach Eksperymentów Zasadniczych I.

5. Parametry występujące w regule RF_6:

a) Współrzędne węzłów funkcji liniowej aproksymującej optymalną funkcję intonacji pytającej $F_{k,\alpha_1}(t)$ podano w tab.7.1.

Tabela 7.1.

Nr węzła	1	2	3	4	5.
t [ms]	0	T_F - 180	T_F - 105	T_F - 99	T_F - 79
$F_{k,\alpha_1}(t)$	0.69	0.69	0.85	0.89	1.06

Nr węzła	6	7	8	9
t [ms]	T_F - 59	T_F - 44	T_F - 30	T_F
$F_{k,\alpha_1}(t)$	1.4	1.9	2.76	3.46

b) Współrzędne węzłów funkcji liniowej aproksymującej optymalną funkcję intonacji twierdzącej $F_{k,\alpha_2}(t)$ podano w tab.7.2.

Tabela 7.2.

Nr węzła	1	2	3	4	5
t [ms]	0	T_F - 320	T_F - 195	T_F - 99	T_F - 79
$F_{k,\alpha_2}(t)$	0.79	0.79	0.50	0.44	0.46

c.d. tab.7.2.

Nr węzła	6	7	8	9
$t_{[ms]}$	T_F - 59	T_F - 44	T_F - 30	T_F
$F_{k,\alpha_2}(t)$	0.48	0.59	0.76	0.82

7.3. Gramatyka G_{RF} reguł realizacyjnych RF

Przy ustaleniu gramatyki G_{RF} zrezygnowano ze stosowania reguły RF_5 tworzenia akcentowego wariantu podstawowego konturu częstotliwościowego, gdyż jak wykazały wyniki eksperymentu EZ II (6) - rozdz.6.3.3.6. - do uzyskania cechy akcentu we frazach syntetycznych wystarczy zastosowanie reguły RF_4 przed regułą RF_2 (por. punkt 3. w rozdz.6.3.3.7.). Dodatkowo arbitralnie założono przynależność reguły RF_3 do zbioru reguł RF (por. odsyłacz [39]). Przyjęto również, że gramatyka G_{RF} generuje ciąg złożony z parametrów głównych ze zbioru D_{DF} (3.13.), eliminując tym samym zasadniczą przeszkodę w formalizacji G_{RF} jaką jest nieabstrakcyjny charakter terminalnego produktu tej gramatyki, tzn. funkcji czasowych sterujących amplitudą i częstotliwością pobudzenia krtaniowego[40]. Na mocy tego założenia zachodzi:

$$\left\{F_2^i\right\} \subset DF \longrightarrow \left\{D_i\right\} = D_{DF} \qquad (7.3.)$$

co oznacza, iż w dalszych rozważaniach zbiór funkcji czasowych $\left\{F_2^i\right\}$ należących do dziedziny reguł zostanie zastąpiony zbiorem parametrów głównych $\left\{D_i\right\}$ dziedziny tych reguł.

Przyjmijmy również następujące założenia:

Zał.12. Gramatyka G_{RF} stanowi wydzielony zbiór przekształceń gramatyki reguł realizacyjnych G_R i jej argumenty zawierają się w zbiorze argumentów gramatyki G_T.

---

40) Gramatyka generatywna jest z definicji przekształceniem operującym w zbiorze symboli abstrakcyjnych i generującym ciągi symboli abstrakcyjnych.

Zał.13. Aksjomatem gramatyki G_{RF} jest podciąg T_{RF} ciągu T symboli terminalnych komponentu fonologicznego złożony z elementów należących do przecięcia zbioru T zbiorem A_{RF}:

$$V_{RF} = T_{RF} \tag{7.4.1.}$$

$$T_{RF} = T \cap A_{RF} \tag{7.4.2.}$$

gdzie:

V_{RF} - aksjomat gramatyki G_{RF} ,

A_{RF} - skończony alfabet symboli, na których operuje G_{RF},

T - ciąg opisany zależnością (4.1.).

Zał.14. Alfabetem symboli terminalnych gramatyki G_{RF} jest dziedzina DF reguł realizacyjnych RF reprezentowana przez jej zbiór parametrów głównych D_{DF} (3.13.),(7.3.).

Wprowadźmy teraz następującą definicję gramatyki G_{RF}:

Def.16. Gramatyka $G_{RF}\{V_{RF}, A_{RF}, D_{DF}, RF\}$ jest różnowartościowym odwzorowaniem wiążącym jednoznacznie wszystkie możliwe poprawne ciągi T otrzymywane z komponentu fonologicznego z ciągami parametrów głównych $\{D_i\}$ reprezentujących funkcje czasowe sterujące pobudzeniem krtaniowym w procesie syntezy fraz, co przy zadanych A_{RF}, D_{DF} i RF można zapisać:

$$G_{RF} \in \text{Bmap}\left[(V_{RF} = T \cap A_{RF}),\ (\{D_i\} \leftarrow D_{DF})\right]\Big|_{A_{RF},D_{DF},RF} \tag{7.5.}$$

gdzie:

Bmap (A,B) - różnowartościowe odwzorowanie zbioru (ciągu) A w B,

V_{RF}, A_{RF}, D_{DF}, RF - argumenty gramatyki G_{RF}.

Gramatyka G_{RF} (Def.16.) posiada dwie specyficzne właściwości, które można określić jako zdolność "filtracji" ciągu T otrzymywanego z komponentu fonologicznego ($T_{RF} = T \cap A_{RF}$) oraz uwikłanie argumentów (V_{RF} na podstawie (7.4.1.) i (7.4.2.) zależy od A_{RF}, D_{DF} reprezentuje dziedzinę RF przy czym zarówno D_{DF}, jak i RF są argumentami G_{RF}).

Dzięki tym właściwościom uzyskano:

1. Możliwość wydzielenia G_{RF} z gramatyki G_R.
2. Możliwość sformalizowania gramatyki reguł realizacyjnych RF (por.uwagi o dualiźmie reguł RF zamieszczone w rozdz.1.2., s.10 oraz odsyłacz[40]).

W odniesieniu do rozważanego w pracy przypadku syntezy dźwięcznych fraz mowy, gdzie ciągiem sterującym jest ciąg T opisany zależnością (4.1.) oraz po uwzględnieniu założeń podanych w niniejszym rozdziale ustalono następującą kolejność stosowania reguł ze zbioru RF:

$$RF_1 \longrightarrow RF_4 \longrightarrow RF_2 \longrightarrow RF_3 \longrightarrow RF_6 \qquad (7.6.)$$

$$\llcorner\!\!\llcorner [BG \quad + P_s] \lrcorner\!\!\lrcorner$$

$$\llcorner\!\!\llcorner BG\ [+ P_s] /\alpha_r \lrcorner\!\!\lrcorner$$

Ustalanie następstwa $RF_1 \longrightarrow RF_4 \longrightarrow RF_2$, gdzie wszystkie reguły wykonywane są przez podstawienie bezwarunkowe, powoduje generowanie przez G_{RF} tylko akcentowych wariantów grupy wydechowej, tzn. $BG\ [+ P_s]$, co jest zgodne z jednym z podstawowych założeń pracy (Zał.1.) o obligatoryjnej realizacji zestroju akcentowego we frazach mowy polskiej[41]. Reguła RF_6 (3.6.) wykonywana jest tylko po wykryciu w T_{RF} znacznika intonacji α_r.

W gramatyce G_{RF} ze względu na jej wyodrębnienie z całościowej gramatyki reguł realizacyjnych G_R przyjęto, że parametry wspólne z gramatykami reguł RK lub RT (por.rozdz.4.2.) oraz parametry generowane przez gramatyki G_{RK} lub G_{RT} i wykorzystywane w G_{RF} znajdują się w słowniku dostępnym przez G_{RF}. Stąd komunikacja pomiędzy wydzielonymi podgramatykami gramatyki G_R zachodzi w obębie komponentu reguł realizacyjnych KRR poprzez Słownik 1 (rys.4.1.).

---

[41] Z założenia tego wynika, iż w ciągu T (4.1.) musi przynajmniej raz wystąpić element z cechą $+P_s$. W przeciwnym wypadku podana w pracy gramatyka G_{RF} wykaże błąd w ciągu T, gdyż po procedurze przekształcenia $T \longrightarrow T_{RF}$ otrzymany ciąg T_{RF} nie będzie kompletny. Jest to jeszcze jedna specyficzna cecha podanej gramatyki G_{RF}.

Warto tu również przypomnieć, że po przeprowadzeniu procedury optymalizacji parametrów dziedziny reguł realizacyjnych, zbiór parametrów można zastąpić zbiorem ich optymalnych wartości (por. zależności (3.9.) do (3.12.) z rozdz.3.4.), co oznacza zastąpienie w G_{RF} argumentu D_{DF} argumentem D_{DFopt} (3.11.).

7.4. Uwagi końcowe

1. Zamieszczone w pracy wyniki rozważań teoretycznych oraz badań eksperymentalnych wskazują, że założone w rozdz.1.5.2. cele pracy zostały osiągnięte. Podano zbiór reguł pozwalających w prosty sposób sterować częstotliwością podstawową i amplitudą pobudzenia krtaniowego w procesie syntezy fraz mowy polskiej, przeprowadzono procedurę doboru i optymalizacji parametrów występujących w regułach oraz podano czteroargumentową gramatykę G_{RF} zarządzającą wykonaniem tych reguł (cel 1, 3 i 4). Część eksperymentalną przeprowadzono za pomocą specjalnie opracowanego cyfrowego syntezatora formantowego stanowiącego wszechstronne i elastyczne narzędzie do badań nad szeroko rozumianymi regułami realizacyjnymi w syntezie dźwięcznych fraz mowy (cel 2).

2. Za swój oryginalny wkład do rozwoju badań nad syntezą mowy autor uznaje podanie koncepcji przejścia z ciągu abstrakcyjnych symboli terminalnych generowanych przez komponent fonologiczny w zbiór funkcji czasowych sterujących aparatem syntezy dźwięków mowy, opartej o zbiór reguł realizacyjnych stanowiących jeden z argumentów gramatyki G_{RF} zarządzającej wykonaniem tych reguł (w znanej autorowi literaturze nie podejmowano prób kompleksowego rozwiązania tego problemu), podanie sformalizowanego zapisu gramatyki G_{RF} w postaci zmodyfikowanej wersji gramatyki generatywnej skończenie stanowej, co zapewnia kompatybilność struktury komponetu reguł realizacyjnych z komponentem fonologicznym i syntaktycznym, przeprowadzenie szeregu eksperymentów dla syntetycznych samogłosek i fraz mowy polskiej, w wyniku których ustalono optymalne parametry reguł RF z uwzględnieniem zagadnień akcentu i podstawowych typów intonacji. Dodatkowo autor, posługując się metodami analizy strukturalnej, podjął próbę usystematyzowania szeregu pojęć, zjawisk i proble-

mów związanych z generacją pobudzenia krtaniowego w syntezie fraz mowy oraz podał metodykę planowania i estymacji wyników badań eksperymentalnych dotyczących oceny fraz mowy syntetycznej.

3. Praca miała charakter typowo interdyscyplinarny i wymagała zaznajomienia się z szeregiem dziedzin wiedzy, jak lingwistyka strukturalna, fonetyka, filtracja cyfrowa, teoria estymacji statystycznej i skalowania, psychometria, teoria mnogości, synteza dźwięków mowy, zatem nieuniknione było korzystanie, w miarę możności, z gotowych rozwiązań szczegółowych koncentrując się na ich scaleniu i podporządkowaniu założonym celom badawczym.

4. Podana w pracy koncepcja gramatyki reguł RF pozwala na jej uogólnienie na cały zbiór reguł realizacyjnych R przy dowolnej postaci ciągu T, co umożliwi opracowanie pełnej gramatyki G_R do celów syntezy mowy.

5. Podana w pracy koncepcja sterowania parametrami pobudzenia krtaniowego w syntezie fraz pozwala na generowanie dłuższych komunikatów o ile przyjmie się, co jest dopuszczalne w pewnych zastosowaniach praktycznych, że każdy komunikat mówiony składa się z ciągu fraz (oczywiście zostanie tu pominięta ponadfrazowa struktura suprasegmentalna).

6. Jako główne kierunki dalszych badań stanowiących kontynuację niniejszej pracy można wymienić:

- dobór i optymalizację parametrów reguł RK i RT podanych w rozdz.4.2. (por.odsyłacz[26]),
- opracowanie gramatyki G_R reguł realizacyjnych na poziomie fraz mowy syntetycznej,
- stopniowe rozbudowanie zbiorów reguł RF, RK i RT aż do osiągnięcia poziomu syntezy dowolnego tekstu w języku mówionym,
- opracowanie pełnej gramatyki G_R .

7. Niewątpliwie niedosyt może budzić stosunkowo jednostronne wykorzystanie bogatego materiału otrzymanego z subiektywnej oceny próbek mowy syntetycznej, jednak to wynikało z postawionych celów pracy (szukanie optimum optimorum) oraz z faktu praktycznej niemożności rozszerzenia, w ramach tego typu pracy, i tak już obszernej tematyki badawczej.

WYKAZ LITERATURY

1 Saussure F., de, Kurs językoznawstwa ogólnego, PWN, Warszawa 1961. Tłum. K.Kasprzyk.

2 Chomsky N., Current issues in linguistic theory, Mouton, The Hague 1965.

3 Chomsky N., Aspects of the theory of syntax, MIT Press, Cambridge, Mass. 1965.

4 Chomsky N., Topics in the theory of generative grammar, Mouton, The Hague 1966.

5 Lyons J., Wstęp do językoznawstwa, PWN, Warszawa 1976. Tłum. K.Bogacki.

6 Sigurd B., Struktura języka, PWN, Warszawa 1975.Tłum. Z.Wawrzyniak.

7 Lieberman P., Intonation, perception and language, MIT Press., Cambridge, Mass. 1967.

8 Allen J., Reading machines for the blind: the technical problems and the methods adopted for their solution, IEEE Trans. Audio, Electroacust, vol. AU-21, nr 3, 1973, s.259-264.

9 Coker C.H., Umeda N., Browman C.P., Automatic synthesis from ordinary English text, IEEE Trans. Audio, Electroacoust., vol. AU-21, nr 3, 1973, s.293-297.

10 Nye P.W., Hankins J.D., Rand T., Mattingly I., Cooper F.S., A plan for the field evaluation of an automated reading system for the blind, IEEE Trans. Audio, Electroacoacust, vol. AU-21, nr 3, 1973, s.265-267.

11 Klatt D.H., Structure of a phonologicol rule component for a synthesis - by-rule program, IEEE Trans. Acoust. Speech, Signal Processing, ASSP-22, 1976, s.391-398.

12 Allen J., Synthesis of speech from unrestricted text, Proc. IEEE, vol. 64, nr 5, 1976, s.433-442.

13 Umeda N., Linguistic rules for text-to-speech synthesis, Proc. IEEE, vol. 64, nr 4, 1976, s.443-451.

14 Fallside F., Young S., Speech autput from a computer Controlled water-supply network, Proc. IEEE, vol. 125, nr 2, 1978.

15 Cohen M., Massaro D., Real-time speech synthesis, Behavior Res. Method, Instrumentation, vol. 8, nr 2, 1976, s.189-196.

16 Clark J.E., A real-time speech synthesis system, Monitor-Proc. IREE Aust., vol. 38, nr 3, 1977, s.56-69.

17 Kacprowski J., Mikiel W., Recent experiments in parametric synthesis of Polish speech sounds, referat B-5-11, VI International Congress on Acoustics, Tokio, 1968.

18 Kacprowski J., Mikiel W., Realizacja procesu syntezy mowy za pomocą syntezatora SYNFOR II, Prace IPPT PAN, nr 25, 1968.

19 Kacprowski J., Akustyczne aspekty problemu komunikacji człowiek - komputer w języku naturalnym, Archiwum Akustyki, t.7, nr 3-4, 1972, s.201-212.

20 Iivonen A., On the relationships between the main components in the control of speech, Speech Comm. Seminar, vol. 2, Stockholm, 1974, s.163-171.

21 Jassem W., Mowa a nauka o łączności, PWN Warszawa, 1974.

22 Mangold H., Stall D.S., Principles of text controlled Speech Communication with Computers, wyd. L. Bolc, Carl Hanser Verlag, 1978, s.138-181.

23 Wolf D., Entropy of speech signals and rate disfortion function, Int. Zürich Sem. on Digitol Comm., 1974, s. B8 ff.

24 Carlson R., Granström B., A phonetically oriented programming language for rule description, Speech Comm. Seminar, vol. 2, Stockholm, 1974, s.245-254.

25 Friedman J., Morin Y.Ch., Phonological grammar tester: description, Phonetics Lab. Univ. of Michigan, raport NTIS PB 208 708, 1971.

26 Friedman J., Computer exploration of fast-speech rules, IEEE Trans.Acoust. Speech, Signal Processing, vol. ASSP-23, nr 1, 1975, s.100-103.

27 Mattingly I.G., Synthesis by rule of prosodic features, Language and Speech, vol. 9, nr 1, 1966, s.1-13.

28 Majewski W., Hollien H., Formant frequency regions of Polish vowels, J.Acoust. Soc. Amer., vol. 42, nr 5, 1967, s.1031-1037.

29 Kudela K., A study of the optimal formant frequency values of Polish vowels using synthetic speech, Speech Analysis and Synthesis, vol. II, PWN, Warszawa, 1970, s.219-238.

30 Kudela-Dobrogowska K., Further studies of the optimal formant frequency values of Polish vowels, Speech Analysis and Syntesis, vol. 3, PWN, Warszawa, 1973, s.266-285.

31 Kacprowski J., Mikiel W., Simplified rules for parametric synthesis of nasal and stop consonants in C-V syllables ..., Acustica, vol. 16, 1965/66, s.356-364.

32 Kacprowski J., Mikiel W., Realizacja procesu syntezy mowy za pomocą syntezatora SYNFOR II, Prace IPPT PAN, nr 25, 1968.

33 Myślecki W., Zalewski J., Gos A., Badanie wpływu kształtu impulsów pobudzenia krtaniowego na jakość syntetycznych samogłosek polskich, Prace XXII Seminarium z Akustyki, Świeradów, 1975, s.152-157.

34 Myślecki W., Zalewski J., Badanie wpływu zmian częstotliwości podstawowej oraz amplitudy pobudzenia ..., Prace XXII Seminarium z Akustyki, Świeradów, 1975, s.207-211.

35 Zalewski J., Myślecki W., Research on the Selection of Fo for the optimal synthesis of Polish vowels, Referat .... , VIII Int. Congress of Phonetic Sciences, Leeds, 1975.

36 Zalewski J., Myślecki W., Selection of glottal excitalion parameters optimising the naturalness of synthetic speech, Referat, IX Int. Congress of Phonetic Sciences, Kopenhaga, 1978.

37 Gos A., Myślęcki W., Zalewski J., Dynamiczne sterowanie cyfrowym syntezatorem formantowym ..., Prace XXIV Seminarium z Akustyki, Gdańsk, 1977, s.76-79.

38 Jassem W., Podstawy fonetyki akustycznej, PWN, Warszawa, 1973.

39 Kacprowski J., Mikiel W., The terminal-analog speech synthesizer as acoustic output of a computer, Proc. 7-th Int. Congress on Acoustics, Budapest, 1971, Referat 23-C-4.

40 Coker C.H., Cummiskey P., On-line computer control of the formant synthesizer, J.Acoust Soc. Amer., vol. 38, 1965, s.940 (A).

41 Tomlinson R.S., SPASS - an improved terminal-analog speech synthesizer, J. Acoust. Soc. Amer., vol. 38, 1965, s.940 (A).

42 Liljentorants J.C., The OVE Speech Synthesizer, IEEE Trans. Audio Electoacoust., vol. AU-16, nr 1, 1968.

43 Knight J.L., Minicomputer presentation of speech stimuli, Behar. Res. Methods and Instrum., vol. 9, nr 2, 1977, s.169-172.

44 Soskuty O.V., A voice generator on the principle of phonem synthesis, Elektronik, vol. 26, nr 9, 1977, s.44-48.

45 Strube H.W., Analog discrete-time filter for speech synthesis, IEEE Trans. Acoust., Speech and Signal Process., vol. ASSP-25, nr 1, 1977, s.50-55.

46 Gagnon R.T., Voice synthesizer, U.S. Patent 3.908.085, 1975.

47. Allen J., Steingart R.J., A special purpose processor for speech synthesis, 1977 Electro Conf. Record, El Segundo, Calif., USA, 1977, s.1-6.

48 Hewes C.R., Buss D.D., De Wit M., Bradersen R.W., CCOS in speech processing, IEEE Conf. EASCON-77, Arlington, USA, 1977, Referat 23-3A.

49 Brown W.S., Mc Glone R.E., Aeorodynamic and acoustic study of stress in sentence productions, J.Acoust. Soc. Amer., vol. 56, nr 3, 1974, s. 971-974.

50 Lieberman P.,Some acoustic correlates of word stress in American-English,J.Acoust.Soc.Amer.,vol.32,1960, s.451-454.

51 Dłuska M., Prozodia języka polskiego, PWN, W-wa, 1976.

52 Bolinger D.L., Theory of pitch accent in English, Word, vol. 14, 1958, s.109-149.

53 Morton J., Jassem W., Acoustic correlates of Stress, Language and Speech, vol. 8, part 3, 1965, s.148-158.

54 Jassem W., Morton J., Steffen-Batog M., The perception of stress in synthetic speech-like stimuli by Polish listeners, Speech Analisis and Synthesis, ed. W.Jassem, vol. I, PWN W-wa, 1968, s.289-308.

55 Majewski W., Zalewski J., Rola częstotliwości podstawowej w procesie percepcji syntetycznych sygnałów dźwiękowych mowy, Prace Naukowe ITA Politechniki Wrocławskiej, nr 13, 1973, s.37-50.

56 Renowski J., Proces przekazywania wiadomości za pośrednictwem akustycznego sygnału mowy i udział w nim częstotliwości podstawowej głosu, Zeszyty Naukowe Politechniki Wrocławskiej, nr 215, 1969.

57 Rabiner L.R., A model for synthesizing speech by rule, IEEE Trans. Audio Electroacust., vol. AU-17, 1969, s.7-13.

58 Laufer A., A programme for synthesizing Hebrew Speech, Phonetica, vol. 32, 1975, s.292-299.

59 Flanagan J.L., Speech analysis, synthesis and perception, Springer-Verlag, 1965.

60 Rothenberg M., Carlson R., Granström B., Linqvist J., A three-parameter voice source for speech synthesis, Preprints of the Speech Comm. Seminar, Stockholm, 1974, vol. 2, s.235-243.

61 Essen O., von, Fonetyka ogólna i stosowana, PWN, Warszawa, 1967, przekład A.Szulc.

62 Ainsworth W.A., A system for converting English text into speech, IEEE Trans. Audio Electroacust., vol. AU-21, nr 3, 1973, s.288-290.

63 Umeda N., Teranishi R., The parsing program for automatic text-to-speech synthesis ..., IEEE Trans. Acoust. Speech, Signal Process., vol. ASSP-23, nr 2, 1975, s.183-188.

64 Majewski W., Blasdell R., Influence of fundamental frequency cues on the perception of some synthetic intonation contour, J. Acoust. Soc. Amer., vol. 45, nr 2, 1969, s.450-457.

65 Nowakowska W., Percepcja częstotliwości podstawowej i poziomu intensywności w mowie polskiej, Referat, XXIV Sem. z Akustyki, Gdańsk-Władysławowo, 1977.

66 Trzebski A., Fizjologia głosu, rozdz. w pracy zbiorowej p.t. Fizjologia człowieka, Państw. Zakł.Wyd.Lekarskich, Warszawa, 1971.

67 Mitrinowicz-Modrzejewska A., Fizjologia i patologia głosu, słuchu i mowy, Państw.Zakł.Wyd.Lekarskich, Warszawa, 1963.

68 Ishizaka K., Flanagan J.L., Synthesis of voiced sounds from a two-mass model of the vocal cords, Bell System Tech. J., vol. 51, 1972, s.1233-1268.

69 Ishizaka K., Flanagan J.L., Acoustic properties of longitudinal displacement in vocal cord vibration, Bell System Tech. J., vol. 56, nr 6, 1977, s.889-918.

70 Berg J., von den, Myoelastic-aerodynamic theory of voice production, J. Speech Hear. Res., vol. 1, 1958, s.227-244.

71 Coker C., Umeda N., Improvements in or relating to systems for the syntesis of speech from alphanumeric data, Patent 1 380 502, The Patent Office, London, 1972.

72 Cooper F.S., Peterson E., Fahringer G.S., Some sources of characteristic vocoder quality, 52-th Meeting of Acoust. Soc. Amer., referat H 1.

73 Lindqvist J., The voice source studied by means of inverse filtering, STL-QPSR, nr 1, 1970, s.3-9.

74 Sondhi M. M., Measurement of the glottal waveform, J.Acoust. Soc. Amer., vol. 57, nr 1, 1975, s.228-232.

75 Rosenberg A.E., Effect of glottal pulse shape on the quality of natural vowels, J.Acoust. Soc. Amer., vol. 49, nr 2, 1971, s.583-590.

76 Holmes J.N., The influence of glottal waveform on the naturalness of speech from parallel formant synthezizer, IEEE Trans. Audio, Electroacoust, vol.AU-21, nr 3, 1973, s.298-305.

77 Rao P.V., Thosar R.B., A programming system for studies im speech synthesis, IEEE Trans. Acoust. Speech, Signal Process., vol. ASSP-22, nr 3, 1974, s.217-225.

78 Isacenko A., Schadlich H., A model of standard German intonation, The Hague, 1970.

79 Studdert-Kennedy M., Hadding K., Auditory and linguistic process in the perception of intonation contours, Language and Speech, vol.XVI, 1973, s.293-313.

80 Carlson R., Granström B., Lindbolm B., Rapp K., Some timing and fundamental frequency characteristics of Swedish sentences: ..., STL-QPSR, nr 4, 1972, s.11-19.

81 Haavel R., Temporal characteristic of the pitch contour, I i II, Acustica, vol. 34, 1976, s.147-157.

82 Sambur M.R., Rosenberg A.E., Rabiner L.R., Mc Gonegal C.A., On reducing the buzz in LPC synthesis, J.Acoust. Soc. Amer., vol. 63, nr 3, 1978, s.918-924.

83 Kacprowski J., Synteza formantowa dźwięków samogłoskowych i nosowych (podstawy teoretyczne), Archiwum Elektrotechniki, t.XIII, nr 3, 1964, s.661-676.

84 Fant G., On the predictability of formant levels and spectrum envelopes from formant frequencies, For Roman Jakobson, s-Gravenhage, 1956, s.109-120.

85 Rabiner L.Q., Digital-formant synthesizer for speech synthesis studies, J.Acoust. Soc. Amer., vol. 43, 1968, s.822-828.

86 Oppenheim A.V., Speech analysis-synthesis system based on homomorphic filtering, J.Acoust. Soc. Amer., vol. 45, nr 2, 1969, s.458-465.

87 Nakatani L.H., Schaffer J.A., Hearing "words" without words: 'prosodic cues, J.Acoust. Soc. Amer., vol. 63, nr 1, 1978, s.234-245.

88 Miller R.L., Nature of the vocal cord wave, J.Acoust. Soc. Amer., vol. 31, nr 6, 1959, s.667-677.

89 Holmes J.N., An investigation of the volume velocity waveform at the larynx during speech by means of on inverse filter, 4-th Int. Congress on Acoust., Copenhagen, 1962, ref. G 13.

90 Flanagan J.L., Some influences of the glottal wave upon vowel quality, Proc. 4-th Int. Congr. Phonetic Sciences, Helsinki, 1961.

91 De Mori R., Laface P. Makhomine V.A., Mezzalama M., A syntactic procedure for the recognition of glottal pulses in continuous speech, Pattern Recognition, vol.9, 1977, s.181-189.

92 Paille J., Source vocale pour synthétiseurs á formants, Revue d'Acoustique, nr 6, 1969, s.111-114.

93 Carcaud M., Contribution à la synthèse de la 'parole - Simulation de la source vocale, Thesis, Orsay University, France, 1975.

94 Flanagan J.L., Landgraf L.L., Self oscilating source for vocal-tract synthesizer, IEEE Trans. Audio Electroacoust., vol. AU-16, nr 1, 1968, s.57-64.

95 Rabiner L.R., Jackson L.B., Schafer R.W., Coker C.H., A hardware realization of a digital formant speech synthesizer, IEEE Trans. Commun. Tech., vol. COM-19, 1971, s.1016-1020.

96 Rabiner L.R., Speech synthesis by rule: an acoustic domain approach, Bell System Tech. J., nr 47, 1968, s.17-37.

97 Michaels S.B., Strong W.J., Analysis-synthesis of glottal excitation, 70-th Meeting, Acoust. Soc. Amer., referat P8, streszczenie - J.Acoust. Soc. Amer., vol. 38, 1965, s.935.

98 Kacprowski J., Theoretical bases of the synthesis of Polish vowels ..., Speech Analysis and Synthesis, vol. I, ed. W.Jassem, PWN Warszawa, 1968.

99 Flanagan J.L., Note on the design of terminal-analog speech synthesizer, J.Acoust. Soc. Amer., vol. 29, 1957, s.306-310.

100 David E.E., Miller J.E., Mathews M.V., Monaural phase effects in speech perception, Proc. 3-rd Int. Congr. Acoust., Amsterdam, 1961, s.227-229.

101 Leites R.D., Sobol'ew W.N, Eff'ekt vlijan'ia faz na słuchovy'e vospriat'ie sint'ezirovannoj rec'i, Elektrosv'iaź, nr 1, 1974, s.62-65.

102 Rosenberg A.E., Schafer R.W., Rabiner L.R., Effect of smoothing and quantizing the parameters of formant - coded voiced speech, J.Acoust. Soc. Amer., vol.50, nr 6, 1971, s.1532-1538.

103 Fiodorowa N.A., The effect of some acoustic parameters of the synthetic speech signal on the perception of stress by russian listeners, Speech Analysis and Synthesis, ed. W.Jassem, vol.3, PWN Warszawa, 1973.

104 Huggins A.W., Viswanathan R., Makhoul J., Speech-quality testing of some variable-frame-rate (VFR) linear - predictive (LPC) vocoders, J.Acoust. Soc. Amer., vol.62, nr 2, 1977, s.430-434.

105 Flanagan J.L., Pitch discrimination for synthetic vowels, J.Acoust. Soc. Amer., vol.30, nr 5, 1958, s.435-442.

106 Tenney J.C., Discriminability of differences in the rise time of a tone, 63-th Meeting of Acoust. Soc. Amer., referat U8.

107 Takasugi T., Suzuki J., Experimental discussion of SPAC and SPOC for noise reduction, J.Radio Res. Lab., vol.24, nr 113, 1977, s.35-40.

108 Hanson B.A., Donaldson R.W., Subjective evaluation of an adaptive differential voice encoder, IEEE Trans. Commun., vol. COM-26, nr 2, 1978, s.201-208.

109 Pisoni D.B., Lazarus J.M., Categorical and noncategorical modes of speech perception ..., J.Acoust. Soc. Amer., vol.55, nr 2, 1974, s.328-333.

110 Łętowski T., Słuchowa ocena urządzeń elektroakustycznych, Zeszyty Naukowe COB-RR i Tr., 1976.

111 Sołtys Z., Wąsowicz Z., Metody oceny słuchowej, Prace Naukowe ITA Polit. Wrocł., seria: Monografie, nr 35, 1978.

112 Coombs C.H., Dawes R.M., Tversky A., Wprowadzenie do psychologii matematycznej, PWN, Warszawa, 1977.

113 Thurstone L.L., The measurement of values, The University of Chicago Press., 1959.

114 Mosteller F., Remarks on the method of paired comparisons; I., Psychometrica, vol.16, nr 1, 1951, s.3-9.

115 Renowski J., Badanie wpływu zmian częstotliwości podstawowej w mowie naturalnej na wrażenie intonacji, Zeszyty Naukowe Politechniki Wrocławskiej, nr 147, Wrocław, 1967, s.3-12.

116 Frąckowiak-Richter L., The duration of polish vowels, Speech Analysis and Synthesis, vol.3, PWN Warszawa, 1973, s.88-115.

117 Richter L., Czas trwania polskich spółgłosek, Prace XXI Seminarium z Akustyki, Rzeszów 1974, s.42-44.

118 Myślecki W., Zalewski J., Gos A., Reguły generacji pobudzenia krtaniowego w procesie syntezy krótkich fraz mowy polskiej, Prace XXIV Seminarium z Akustyki, Gdańsk-Władysławowo, 1977, s.108-111.

119 Collier R., Psychological correlates of intonation patterns, J.Acoust. Soc. Amer., vol.58, nr 1, 1975, s.249-255.

120 Cheung J.Y., Holden A.D., Minifie F.D., Computer recognition of linguistic stress patterns in connected speech, IEEE Trans. Acoust. Speech, Signal Process., June 1977, s.252-256.

121 Renowski J.,'Badanie wpływu zmian częstotliwości podstawowej w mowie naturalnej na wrażenie intonacji, Zeszyty Naukowe Politechniki Wrocławskiej, nr 147, Wrocław, 1967, s.3-12.

Odbiorcy:

		egz.
1.	Ośrodek Informacji NTITA	1
2.	Biblioteka Główna Politechniki Wrocławskiej	1
3.	Biblioteka Międzyinstytutowa I-21/28	1
4.	prof. dr Janusz Kacprowski IPPT PAN, Warszawa	1
5.	prof. dr hab. Wiktor Jassem IPPT PAN, Poznań	1
6.	prof. dr Wojciech Oszywa WAT Warszawa	1
7.	doc. dr hab. Franciszek Połomski Uniwersytet Wrocławski	1
8.	doc. dr hab. Leonard Bolc Uniwersytet Warszawski	1
9.	doc. dr Wojciech Majewski ITA, Politechnika Wrocławska	1
10.	doc. dr Janusz Zalewski ITA, Politechnika Wro Wrocławska	1
11.	mgr inż. Władysław Mikiel dr inż. Ryszard Gubrynowicz IPPT PAN,Warszawa	1
12.	Biblioteka Główna Politechniki Gdańskiej	1
13.	Komisja Przewodów Doktorskich ITA, Politechnika Wrocławska	3
14.	Egzemplarze autorskie	3
	Razem	18