

W. Jassem, G. Demenko

**TRANSKRYPCJA FONETYCZNA
DLA POTRZEB AKUSTYKI MOWY
I JEJ IMPLEMENTACJA
NA DRUKARCE MOZAIKOWEJ**

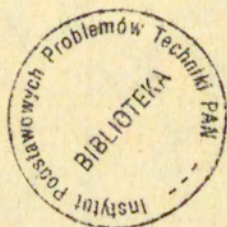
28/1988

P. 269



WARSZAWA 1988

Praca wpłynęła do Redakcji dnia 25 listopada 1988 r.



56751



N a p r a w a c h r ę k o p i s u

Instytut Podstawowych Problemów Techniki PAN

Nakład 170 egz. Ark.wyd. 2,39 Ark.druk. 3,5

Oddano do drukarni w lipcu 1988 r.

Nr zamówienia 404/88

Warszawska Drukarnia Naukowa, Warszawa,
ul. Śniadeckich 8

WIKTOR JASSEM

GRAŻYNA DEMENKO

Pracownia Fonetyki Akustycznej

IPPT PAN

TRANSKRYPCJA FONETYCZNA DLA POTRZEB AKUSTYKI MOWY
I JEJ IMPLEMENTACJA NA DRUKARCE MOZAIKOWEJ

STRESZCZENIE

W pracach z zakresu akustyki mowy zachodzi często konieczność stosowania transkrypcji fonetycznej. Jednak brak jest dotychczas określenia relacji między znakami fonetycznymi i ich ciągami (transkrypcją) a sygnałem mowy z uwagi na podwójny pseudo-paradoks ciągłości/nieciągłości oraz zmienności /inwariancji. Przedstawiono wstępną próbę rozwiązania tego pseudoparadoksu oraz wstępną i w znacznej mierze hipotetyczną próbę określenia relacji między akustycznym sygnałem mowy a transkrypcją.

Zdefiniowano pod programem LETTRIX oraz CHI-WRITER ponad 60 specjalnych znaków fonetycznych. Sporządzono odrębne kody i zbiory znaków fonetycznych dla języka polskiego oraz kilku reprezentatywnych języków o zasięgu międzynarodowym. W tekście niniejszej pracy używamy transkrypcji pod edytorem CHI-WRITER, zaś w Tablicach wyprowadzamy je pod WORD-STAREM. Przedstawiono próbki zastosowania opracowanego programu wykonane za pomocą drukarki mozaikowej sterowanej z mikrokomputera typu PC XT.

Wstęp

W pracach dotyczących problemów akustyki mowy, a szczegól-
nie w pracach akustyczno-fonetycznych, zachodzi często ko-
nieczność posłużenia się transkrypcją, głównie z uwagi na
rozbieżność między elementami pisma w formie drukowanej lub rę-
cznej (litery, znaki interpunkcyjne, spacje międzywyrazowe
itd.) a elementami fonetycznymi (głoski, fonemy, akcenty,

jednostki rytmiczne itd.). Powstaje tu jednak zasadniczy dylemat polegający na braku sformułowania relacji pomiędzy znakami fonetycznymi transkrypcji a fragmentami sygnału akustycznego mowy. Dwa główne problemy polegają na podwójnym pseudo-paradoksie bardzo znacznej zmienności sygnału akustycznego w porównaniu z niewielką liczbą inwariantnych znaków transkrypcyjnych oraz kwazi ciągłym charakterem tego sygnału a dyskretną naturą transkrypcji. W niniejszej pracy dokonuje się wstępnej próby rozwiązania tego pseudoparadoksu. Konieczne w tym celu jest, chociażby ogólnikowe omówienie zróżnicowań intra- i intersegmentalnych oraz podstawowych zjawisk supra-segmentalnych w ścisłym związku z transkrypcją.

Transkrypcja fonetyczna, w różnych jej postaciach, została stworzona przez fonetyków-językoznawców w zasadzie na podstawie audytywnej (słuchowej), 'impresjonistycznej' analizy mowy, która to analiza nosi często znamiona subiektywizmu. Takie pojęcie jak fonem lub alofon zostały zdefiniowane i są stosowane przez fonetyków-lingwistów w oparciu o obserwację słuchową i różne przesłanki teoretyczne, bez uwzględnienia, a nieraz nawet wbrew wynikom akustycznej analizy mowy. Praca niniejsza jest prawdopodobnie pierwszą próbą - wstępną i w znacznym stopniu hipotetyczną - opisowych (tzn. na danym etapie nie sformalizowanych) definicji pojęć fonetycznych, na których opierają się różne transkrypcje, w szczególności transkrypcja fonematyczna, z koniecznymi uzupełnieniami, w oparciu wyłącznie o właściwości i zależności między cechami AKUSTYCZNYMI mowy.

Drugi cel niniejszej pracy jest ściśle pragmatyczno-wdrożeniowy. W chwili powstawania jej koncepcji (lato 1986) nie było programu, który umożliwiłby zapis w postaci transkrypcji fonetycznej za pomocą drukarki mozaikowej sterowanej z mikrokomputera typu PC-XT/AT. W numerze 66 Working Papers in Phonetics (UCLA) z maja 1987 ukazał się artykuł P. Ladefogeda (Ladefoged 1987), z którego wynika, że istnieje już program (który można zrealizować przy użyciu drukarek laserowych) do transkrypcji sterowanej z komputera Macintosh. Podczas 11 Kongresu Nauk Fonetycznych Ladefoged poinformował jednego z autorów, że w danym momencie (sierpień 1987) nie jest mu znany

odpowiedni program dający się implementować pod systemem MS-DOS (BIOS) na komputerze IBM XT/AT lub kompatybilnych. Ta informacja Przewodniczącego Permanent Council for the Organization of International Congresses of Phonetic Sciences stała się dla autorów niniejszej pracy dodatkową motywacją.

Zdefiniowaliśmy ponad 60 znaków specjalnych korzystając z programu LETTRIX, a odpowiednie teksty w wersji ortograficznej i w transkrypcji przygotowaliśmy pod edytorem WORDSTAR. Popularny obecnie program edycyjny CHI-WRITER łączy właściwości tamtych dwóch i umożliwia bezpośrednie ukazanie transkrypcji na monitorze w trybie graficznym. Jednak możliwości definicyjne znaków specjalnych są większe w programie LETTRIX niż w programie CHI-WRITER.

1. Procesy fonetyczne

Na poszczególnych poziomach komunikacji lingwistycznej rozróżnić można następujące procesy fonetyczne :

I. Psychologiczne nadawcy: intencje realizacyjne elementów segmentalnych i suprasegmentalnych.

II. Neurologiczne nadawcy (a) w ośrodkowym układzie nerwowym oraz (b) w układzie odprowadzającym (sterowanie narządów mowy).

III. Fizjologiczne nadawcy (artykulacja - ruchy narządów mowy).

IV. Aerodynamiczne artykulacyjne (przepływ powietrza przez tor artykulacyjny).

V. Akustyczne w wytwarzaniu (pobudzenie, filtrowanie i promieniowanie).

VI. Akustyczne w przenoszeniu (przenoszenie bezpośrednie przez otaczające nadawcę i odbiorcę powietrze lub pośrednie przez urządzenia transmisyjne i/lub zapisująco-odtwarzające).

VII. Aerodynamiczne w odbiorze (ucho zewnętrzne).

VIII. Akustyczno-mechaniczne w odbiorze (ucho środkowe i wewnętrzne).

IX. Neurologiczne odbiorcy (a) w układach odprowadzających, (b) w ośrodkowym układzie nerwowym.

X. Psychologiczne odbiorcy (percepcja z dekodowaniem i rozpoznawaniem).

W skrajnie skrótowy i uproszczony sposób można fonetyczne aspekty komunikacji lingwistycznej opisać następująco:

W psychice nadawcy powstaje intencja (zamiar) przekazania komunikatu o określonym znaczeniu i określonej formie (graficznej, fonicznej lub innej). Ta intencja jest realizowana poprzez aktywizację tych sfer koartykalnych w ośrodku mowy, które związane są z daną formą foniczną lub graficzną.¹ Przy założeniu, że wybrana została forma foniczna, przez nerwowy układ odprowadzający przekazywane są sygnały - odpowiadające wybranym formom - które sterują narządami mowy i powodują powstawanie określonych ich ruchów. Te ruchy z kolei wytwarzają określone procesy akustyczne dwójakiego rodzaju: (a) pobudzenia (np. turbulencja, drgania (kwazi) periodyczne, impulsy) i (b) modulacji (zjawiska związane z przepływem powietrza przez tor głosowy). Z akustycznego punktu widzenia zjawiska fonetyczne opisuje się więc tak na etapie wytwarzania (produkcji mowy przez nadawcę), jak i przesyłania. W szczególności, wytworzony sygnał akustyczny może na drodze do odbiorcy ulec przetworzeniu (np. na sygnał elektryczny), zniekształceniu (linearnemu lub nielinearnemu) albo zakłóceniu (przez dodatkowe, zazwyczaj niepożądane źródło akustyczne). Analiza i przetwarzanie w uchu środkowym i wewnętrznym oraz na nerwowej drodze doprowadzającej dokonuje się na sygnałach nie zdekodowanych. Przetwarzanie z dekodowaniem, możliwym dzięki sprzężeniu zwrotnemu z układem pamięciowym odbiorcy, odbywa się na poziomie ośrodkowego układu nerwowego, a jego psychologicznym korelatem jest percepcja.

Transkrypcja fonetyczna związana jest dotychczas w sposób najbardziej bezpośredni z psychologicznym poziomem przetwarzania, przede wszystkim u odbiorcy. Nie stanowi ona oczywiście części normalnego odbioru w komunikacji, lecz należy do opisu analitycznego opartego przede wszystkim na uświadomionym i odpowiednio zorganizowanym przetwarzaniu psycholingwistycznym. Tak więc związek transkrypcji z akustycznymi zjawiskami mowy jest dotychczas pośredni.

2. Pisownia konwencjonalna i transkrypcja fonetyczna

Już od dawna (por. Albright 1958) stosowana jest transkrypcja fonetyczna w badaniach i opisie dźwiękowej formy języka, przede wszystkim w odróżnieniu od jej formy ortograficznej. Jakkolwiek w językach posiadających formę alfabetyczno-graficzną istnieje zawsze jakiś związek między elementami fonetycznymi a graficznymi, to jednak relacje między tymi dwoma typami elementów mogą być względnie proste (np. w języku fińskim lub węgierskim) poprzez dość regularne ale złożone (np. w jęz. polskim), aż do bardzo skomplikowanych i nieregularnych (np. w jęz. angielskim). Konieczność stosowania transkrypcji w miejsce przyjętej pisowni, skodyfikowanej w ortografii poszczególnych języków, wypływa z następujących okoliczności :

(1) Dany grafem² może w określonym języku odpowiadać różnym pojedynczym głoskom : Pol. <G> /g, j, k/ : droga /droga/, drogi /droji/, dróg /druk/ (jednoznacznie z druk /druk/); ang. <A> /e, e a, o, o, i/ : man /men/, many /meni/, pass /pas/, what /wot/, halt /holt/, cottage /kotidz/. W jęz. włoskim samogłoski przednie: półprzymknięta i półotwarta są zapisywane ortograficznie tak samo, mianowicie za pomocą grafemu <E>. Istnieje kilkadziesiąt wyrazów o tej samej pisowni, lecz różnej wymowie: arena /arena/ "piasek", arena /arena/ "arena"; cera /tjera/ "wosk", cera /tjera/ "cera" (na policzkach), collega /kol'lega/ "(on) łączy", collega /kol'lega/ "kolega", credo /kredo/ "wierzę", credo /kredo/ "wyznanie wiary", messe /messe/ "msze" (l. mn.), messe /messe/ "żniwa" itd. Podobnie jest w tym języku kilkadziesiąt par wyrazów o tej samej pisowni, zawierających grafem <O>, wymawianych bądź jako samogłoska tylna półprzymknięta bądź jako półotwarta : conservatori /konserva'tori/ "konserwatyści", conservatori /konserva'tori/ "konserwatoria" (np. muzyczne), corso /korso/ "kurs", "trasa", corso /korso/ "korsykański" sorta /sorta/ imiesłów od sorgere "powstawać", sorta /sorta/ rodzaj", torta /torta/ "tort", torta /torta/ imiesłów od torgere "podnosić" itp.

(2) Określony grafem może oznaczać różne ciągi głosek: pol. <A> /om, on, op, on/ : trąba /tromba/, kąt /kont/ wziąć /wzjont/, bąk

/bonk/. Może także oznaczać albo pojedynczą głoskę albo ciąg głosek: ang. *wind* /wind/ "wiatr", ale *wind* /wajnd/ "kręcić".

(3) Określone mu ciągowi grafemów mogą odpowiadać pojedyncze głoski albo (różne) ciągi głosek: pol. <RZ> /ʒ,ʃ,rz/ *wierzba* /vjezba/, *kucharz* /kuxaf/, *marznąć* /marznontɕ/. Ang. <MB> /m,mb/ *plumber* /plama/, *number* /namba/. <OUR>: /awa,ua,o/ *scour* /skawa/, *tour* /tua/, *four* /fo/. Szczególnie drastyczny przypadek stanowi ciąg czterech grafemów <OUGH> w j. ang.: *thought* /θot/, *bough* /baw/, *rough* /raf/, *through* /θruw/, *cough* /kof/, *thorough* /θara/, *hough* /hok/, *though* /ðaw/.

(4) Określona głoska (ściślej: fonem, zob. dal.) danego języka może być oddawana przez różne pojedyncze grafemy lub jeden albo więcej różnych ciągów grafemów. Polska spółgłoska zwarta dźwięczna dźwięczno-podniebienna ma 4 różne formy graficzne: <D>, <Z>, <ZI> lub <S> *groźba* /grozba/, *grozić* /groziɕ/, *ziarno* /zarno/, *prośba* /prozba/. Angielska samogłoska tylna średnia zaokrąglona jest zapisywana odmiennie w każdym z następujących wyrazów: *ball* /bol/, *walk* /wok/, *ward* /wod/, *cause* /koz/, *taught* /tot/, *ought* /ot/, *raw* /ro/, *glory* /glori/, *hoarse* /hos/, *cord* /kod/, *cure* /kjo/. Francuska samogłoska półotwarta nosowa jest zapisywana różnie. Następujące wyrazy: *tin*, *tins*, *tint*, *teins*, *teint* oraz *thym* są w wymowie identyczne: /tɛ/.

(5) Jeśli większa liczba języków używa takiego samego alfabetu, to określony grafem może odpowiadać zupełnie odmiennym głoskom w poszczególnych językach. Grafem <J> na przykład oznacza samogłoskę niesylabiczną w pol.: *jeść* /jeɕɕ/, spółgłoskę dźwięczną palatalną w niem.: *jemand* /je:mant/, spółgłoskę siabą, dźwięczno-palatalną, zwartotrącą w ang.: *jet* /dʒet/, spółgłoskę dźwięczną trącą zadźwięczną we franc.: *jouer* /ʒwe/, a spółgłoskę trącą bezdźwięczną tylnojęzykową w hiszpańskim: *viajero* /vja'xero/.

(6) W kilku językach używających określonego alfabetu takie same lub bardzo podobne głoski mogą być w piśmie oddawane zupełnie inaczej. Zwarto-trącej spółgłosce dźwięcznej dźwięczno-palatalnej (lub dźwięczno-palatalnej) odpowiada w piśmie <DE> w jęz. pol.: *drożdże* /drozɕɕe/, <J> w ang.: *cajole* /ka'dʒawl/, zaś <XH> w albańskim: *Hoxha* /hoda/.

Pewnym - jak widać z powyższych przykładów niezbyt

bliskim - odpowiednikiem grafemu jest fonem. O jego definicji i znaczeniu będzie mowa w dalszym ciągu. Przyjmując narazie tylko bardzo ogólnikowo, że fonem jest określona klasa podobnych, ale jednak różnych głosek, można stwierdzić, że w pisowniach poszczególnych języków już w okresie (choćby odległym), kiedy w nim zaczęto stosować któryś z alfabetów, zawsze istniała silniejsza lub słabsza, ale wyraźna tendencja do zachowania zasady przypisania danego grafemu jednemu fonemowi. Jednakże dla określonych celów, również w zakresie akustyki mowy, zachodzi czasem potrzeba zaznaczenia w piśmie lub w druku różnicowań głoskowych w obrębie fonemu, czego już z reguły normalna ortografia nie uwzględnia. Angielski fonem spółgłoskowy lateralny obejmuje kilka wariantów (takie warianty nazwiemy niżej alofonami), wyraźnie się różniących cechami akustycznymi (częstotliwością i przebiegiem czasowym formantów). W ortografii tych różnicowań się nie zaznacza. Natomiast w licznych pracach z zakresu automatycznego rozpoznawania mowy wspomina się o co najmniej dwóch odmianach fonemu lateralnego i niejednokrotnie zachodzi konieczność odróżnienia tych odmian w druku przy omawianiu przebiegu i wyników rozpoznawania. Trzeba wówczas stosować dwa znaki transkrypcyjne [l] oraz [l̥]. W polskim języku wyróżnia się 6 fonemów samogłoskowych. Jednak funkcjonalnie różnych samogłosek jest znacznie więcej (por. np. Steffen-Batogowa 1975, str. 13-14). Różnicowania dotyczą tak częstotliwości jak i szerokości wstęgi (i poziomu) formantów. Nieuwzględnienie ich np. w regułach syntezy mowy polskiej - jak wskazują bieżące prace - nie jest dopuszczalne, a możliwość oznaczenia w określonych przypadkach każdej odmiany odrębnym znakiem fonetycznym jest, jeśli nie absolutnie konieczne, to bardzo dogodne.

3. Podwójny pseudo-paradoks; ciągłość/nieciągłość i zmienność/ inwariancja

W rozważaniach teoretyczno-fonetycznych i teoretyczno-lingwistycznych w okresie od schyłku ubiegłego stulecia do lat czterdziestych lub pięćdziesiątych obecnego wieku znakomita

większość specjalistów rozróżniała tzw. "fizykalistyczny" oraz tzw. "lingwistyczny" punkt widzenia. We wcześniejszym okresie utożsamiano traktowanie "lingwistyczne" z psychologicznym (Baudoïn de Courtenay, Rudnicki, Trubecki we wcześniejszych pracach itd.). Później stworzono mit "autonomiczności" lingwistyki, która miała być nauką "ani fizykalną ani psychologiczną" (Trubetzkoy 1949, str. 1-15). Taki podział punktów widzenia był dogodnym unikiem umożliwiającym akceptację obserwowanej podwójnej sprzeczności: (1) odcinek mowy, który w analizie lingwistycznej jest reprezentowany przez sekwencję dyskretnych elementów zapisywanych kolejnymi znakami transkrypcji okazuje się w analizie akustycznej z regułą zjawiskiem o charakterze ciągłym, w którym niepodobna wyznaczyć miejsc stanowiących granice pomiędzy kolejnymi elementami. (2) Znaki lingwistyczne (w formie transkrypcji) stanowią w danym języku zamknięty i skończony zbiór inwariantnych elementów kodu fonetycznego (lub fonologicznego), podczas gdy analiza akustyczna mowy ujawnia, że - na ile w ogóle znakom lingwistyczno-fonetycznym można przypisać jakieś fragmenty sygnału - fragmenty takie są "nieskończenie różne" (por. np. Hammarberg 1976). Tę drugą sprzeczność można by rozwiązać stosując ontologiczną teorię abstrakcji, ale takie próby podejmowano na gruncie fonetyki (lub lingwistyki) nader rzadko³. Chociaż paradoks ciągłości/nieciągłości również może być rozpatrywany filozoficzno-metodologicznie, to jednak konkretne próby jego przezwyciężenia nie mają takich podstaw teoretycznych. Były one podejmowane na zasadzie operacyjnej wszędzie tam, gdzie trzeba było wykonać na sygnale akustycznym pomiary tzw. iloczasu fonetycznego, tj. rozciągłości czasowej głosek. Podawano szereg pragmatycznych reguł segmentacji, a wśród autorów zachodziła pod tym względem duża zgodność. Operacyjne reguły segmentacji akustyczno-fonetycznej podają m. in. Fant i Lindblom (1961), Peterson i Lehiste (1960), a w polskiej literaturze Frąckowiak-Richter (1973).

Jeden z najbardziej gwałtownych ataków przeciw istnieniu możliwości teoretycznie uzasadnionej segmentacji fonetyczno-akustycznej stanowią oparte na metafizycznych przesłankach artykuły R. Hammarberga (1969) oraz (1982). W drugim

z nich stwierdza się kategorycznie, że "segmentacja, która jest oczywiście swego rodzaju kategoryzacją, jest zjawiskiem czysto umysłowym. W świecie fizycznym segmentów poprostu nie ma" (str. 124, tłum.aut.). Ostrą odprawę dała Hammarbergowi wkrótce C.A. Fowler (1983) ale ta replika ma charakter teoretyczno-filozoficzny z argumentacją o charakterze dedukcyjnym, bez empirycznego dowodu na słuszność wysuwanej tezy, że segmenty są realne tak fizycznie (akustycznie), jak i psychologicznie (percepcyjnie).

Teoria i praktyka segmentacji akustycznego sygnału mowy na odcinki jednoznacznie skorelowane z głoskami lub fonemami stają się problemem węzłowym w automatycznym rozpoznawaniu mowy ciągłej. Jak długo ARM ogranicza się do identyfikacji elementów nielicznego leksykonu składającego się z kilkudziesięciu do kilkuset haseł (zwłaszcza wyrazów wypowiedzianych w izolacji), jednostką wzorcową może być całe hasło. Mamy tu do czynienia z rozpoznawaniem globalnym. Ale w rozpoznawaniu mowy ciągłej magazynowanie wzorców całych wyrazów staje się albo niemożliwe albo całkowicie nieekonomiczne i niepraktyczne z uwagi na ograniczenia wynikające z pojemności pamięci szybkiego dostępu, czasu wykonania rozkazów oraz czasu dostępu do pamięci zewnętrznej. Proponuje się wprowadzić jako element rozpoznawania takie fragmenty sygnału mowy jak sylaby, półsylaby, diady i in., lecz liczne systemy (a także różne metody syntezy) oparto na segmentacji głoskowej, którą stosowano już u zarania tych poczyniń (Por. Fry 1950, Otten 1964). Fonetyczny lub (fonematyczny) kod jest powszechnie uważany za najekonomiczniejszy w rozpoznawaniu ciągłej mowy przy dużym lub nieograniczonym słowniku i znacznej swobodzie składniowej. Typowe dla lat osiemdziesiątych wyniki segmentacji otrzymał Cohen (1981): "More than 70% of the segmental boundaries were successfully marked within 20 ms of the hand-assigned positions." (str.1436). Zachodzi pytanie, czy taki rezultat, bądź co bądź odległy od 100%, sugeruje, że istotnie fonetyczne granice międzysegmentalne można w sygnale akustycznym wyznaczyć tylko w pewnych przypadkach, a zatem że nie można utrzymać kategorycznej tezy o segmentalnym charakterze tego sygnału, czy też że mamy tu do czynienia z niedoskonałością środków technicznych

albo nie w pełni trafnym sposobem postępowania. Odpowiedź na to zasadnicze pytanie próbuje dać cykl prac wykonanych w ostatnich latach w PFA: Jassem (1984), Domagała (1984), Jassem, Kubzdela, Domagała (1986), Jassem, Domagała (1986). W ostatniej z tych prac wykazano doświadczalnie, że możliwe jest wykrycie 94% granic międzysegmentalnych w statystycznie reprezentatywnym tekście polskim przedstawionym w postaci skwantowanego amplitudowo i częstotliwościowo widma dynamicznego.

Porównując wyniki segmentacji otrzymane w PFA niemal banalnie prostą metodą i przy użyciu prymitywnych środków technicznych, z rezultatami segmentacji uzyskanymi ostatnio w zagranicznych ośrodkach zajmujących się ARM (Por. np. Dixon & Martin 1978, Lea 1980, Haton 1982) wypada stwierdzić, iż przeprowadzane w PFA dowodzenie segmentalnej natury akustycznego sygnału mowy jest bardzo bliskie pełnej realizacji. Tak więc, choć niewiele jeszcze wiadomo na temat problemu ciągłości/nieciągłości na różnych etapach łączności za pomocą mowy, które wymieniono na początku niniejszej pracy, w każdym razie pierwszy paradoks na "złączu" poziomu akustycznego z percepcyjno-lingwistycznym wydaje się być bliski rozwiązania, co pozwoli spełnić pierwszy warunek automatycznego przypisania ciągu fonetycznych znaków transkrypcyjnych ściśle określonym fragmentom akustycznego sygnału mowy.

Dруга sprzeczność: zmienność/inwariancja jest bardziej odległa od wyjaśnienia. Różnego rodzaju doświadczenia wykazują niezbicie, iż w naturalnym procesie percepcji są, lub mogą być, identyfikowane kolejne segmenty fonetyczne. W normalnych warunkach komunikacji identyfikacja wszystkich kolejnych jednostek segmentalnych nie jest konieczna z uwagi na znaczną redundancję na poziomie fonetycznego kodu (byłaby zresztą prawdopodobnie niemożliwa przy zbyt małej pojemności pamięci sensorycznej i pamięci krótkotrwałej w stosunku do szybkiego przepływu informacji na poziomie segmentalnym), natomiast nie ulega wątpliwości, że identyfikowane są co najmniej niektóre segmenty. O tym zaś, że w krańcowym przypadku, gdy brak jest informacji 'zstępującej' ('top down'), n.p. przy odsłuchach logatomów, zachodzi dekodowanie wszystkich (lub prawie wszystkich) kolejnych segmentów fonetycznych (por. przegląd

problematyki w: Łobacz (1985). Na obecnym etapie wiedzy teoretycznej i technicznej w zakresie ARM odwzorowanie percepcyjnej identyfikacji segmentów jest wciąż nieosiągalne.

W rozdz. 5 przedstawiono podstawy, na których można by oprzeć zasady przypisywania zmiennemu w znacznym zakresie sygnałowi akustycznemu, w sposób jednoznaczny, ciągu dyskretnych znaków transkrypcyjnych spośród zamkniętego zbioru elementów, co -- poza aspektami teoretycznymi, zwłaszcza w zakresie lingwistyki -- mogłoby mieć istotne znaczenie dla automatycznego segmentalnego rozpoznawania mowy ciągłej oraz syntezy z tekstu (text-to-speech synthesis).

4. Źródła zmienności

Zgodnie z treścią rozdz. 3 zakładamy obecnie, że sygnał mowy podzielony został wzdłuż osi czasowej na fragmenty ograniczone (a) początkiem wypowiedzi, (b) wewnętrznymi granicami międzysegmentalnymi oraz (c) końcem wypowiedzi. Fragmenty te nazywamy segmentami. Dowolne dwa segmenty tej samej lub dwóch różnych wypowiedzi określonego języka mogą być pomiarowo ekwiwalentne lub nieekwiwalentne. Kryterium ekwiwalencji stanowią progowe wartości parametrów pomiarowych. Wyznaczenie tych wartości jest jednym z najpilniejszych zadań fonetyki akustycznej, tym bardziej że nie ma dotychczas zgodności nawet co do tego, jakie mianowicie powinny to być parametry. Zależą one w sposób krytyczny od metod analizy. Jeśli na przykład sygnał mowy przetwarza się na kwantowane w czasie i częstotliwości widmo dynamiczne, to parametrami będą kolejne pasma częstotliwościowe, a wielkością pomiarową poziom sygnału w danym paśmie i danym przedziale czasowym. Jeżeli analiza odbywa się metodą predykcji liniowej, to parametrami są poszczególne współczynniki predykcyjne, a wielkością obserwowaną wartość poszczególnych parametrów w kolejnych punktach czasowych. Istnieje oczywiście szereg innych możliwości. Nie jest obecnie jeszcze wiadomo, jak ma być ustalona wartość progowa, ale jedną z narzucających się możliwości jest oparcie się na danych percepcyjnych. Byłyby tu wymagane szeroko zakrojone prace badawcze.

Założywszy, że podjęta została decyzja, iż dane dwa segmenty nie są, w sensie wyżej zaproponowanym, ekwiwalentne ('takie same'), należy ustalić, jakie źródło lub źródła zmienności wywołują (lub generują) obserwowaną różnicę. Uzgadniamy przy tym a priori, że mamy tu do czynienia z mową prawidłową (fizjologiczną) w odróżnieniu od mowy patologicznej o dowolnej etiologii, którą pozostawiamy poza zakresem poruszanej tutaj problematyki.

Rozróżniamy przede wszystkim zmienność losową od zmienności systematycznej. Jeżeli różnicę między dwoma segmentami nie możemy przypisać żadnemu ze znanych systematycznych źródeł zmienności, które zostaną wstępnie określone poniżej, to różnicę uważamy za losową.

4.1. Różnice koartikulacyjne.

Pewne cechy artykulacji związane z wytwarzaniem ciągów (sekwencji) segmentów tworzących naturalną wypowiedź mają charakter uniwersalny, tzn. nie są charakterystyczne dla określonego języka lub jakiejś ograniczonej liczby języków, lecz wynikają z ogólnych anatomicznych i fizjologicznych właściwości narządów mowy.

Wartości parametrów fonetyczno-akustycznych z reguły nie są stałe w obrębie poszczególnych segmentów. Bez względu jednak na to, jak dalece segment ma charakter niestacjonarny, wartości parametrów w obrębie danego segmentu zależą w mniejszym lub większym stopniu od wartości w segmentach sąsiednich lub bardzo bliskich. Wynika to przede wszystkim z bezwładności narządów artykulacyjnych, które nie mogą w sposób skokowy przechodzić z określonej pozycji na następną. Przy tym różne narządy artykulacyjne i ich części mają niejednakową bezwładność zależną tak od ich masy, jak i od prędkości i sposobu funkcjonowania mięśni sterujących ich ruchami. Jak widać na ryc.1 początkowe segmenty wyrazów *lis*, *lew*, *las*, *los*, *luz* nie są ekwiwalentne pomiarowo.

Różnice są tutaj zależne od kontekstu fonetycznego w sposób ogólny. Nie są one charakterystyczne dla języka polskiego. Analogiczne zróżnicowania występują w innych językach, jak to widać z porównania np. z wyrazami ang. *lick*, *lack*, *lock*, *look*

lub niem. *lies, las, los*. W danym przypadku te współzależności wynikają stąd, że przednia część języka artykułuje w znacznej mierze niezależnie od części środkowej i tylnej. Ruchy części przedniej o mniejszej masie są stosunkowo szybkie, zaś środkowej i tylnej są wolniejsze. Pozycja lub ruch języka w czasie wymawiania segmentu samogłoskowego są częściowo przygotowane już w obrębie segmentu poprzedniego. Tego rodzaju współzależności między sąsiednimi (lub bardzo bliskimi) segmentami, które są ich wynikiem, określamy jako różnice koartikulacyjne. Takimi są zatem różnice pomiędzy początkowymi segmentami w wyrazach: *lis, lew, las, los, luz*.

4.2 Specyficzne różnice kontekstowe

Jeżeli dystrybucja dwóch, trzech lub większej liczby (rzadko przekraczającej pięć) segmentów jest tego rodzaju, że żaden z nich nie występuje nigdy w danym języku w tym samym otoczeniu fonetycznym, co którykolwiek z pozostałych, to różnicę między takimi segmentami określamy jako kontekstową. Między samogłoskami w *sieć* (fonetycznie ekwiwalentne z *siedź*) oraz *szewc* zachodzą znaczne różnice widmowe, jak to widać z ryc.2. Pierwsza z nich jest, według klasyfikacji artykulacyjnej, półprzymknięta, a druga półotwarta. Odpowiednie różnice w częstotliwościach formantowych są na ryc.2 wyraźnie widoczne. Pierwsza występuje w polskim języku tylko w bezpośrednim obustronnym otoczeniu głosek palatalnych (w tym przypadku spółgłosek odpowiadających wersji graficznej *si--* oraz *--ć*), w których druga nie występuje nigdy. Na odwrót, w obustronnym otoczeniu niepalatalnym może wystąpić półotwarta, ale nigdy półprzymknięta samogłoska przednia. Z punktu widzenia kontekstu fonetycznego te dwie głoski wzajemnie się wyłączają.

Tego samego rodzaju samogłoski istnieją w języku francuskim, ale tam mogą występować w tym samym otoczeniu fonetycznym (tzn. ekwiwalentnym kontekście fonetycznym), na przykład na końcu wyrazów *manger* i *mangeait* /mäʒe, mäʒe/. (por. ryc. 3). Zatem w jęz. franc. omawiane samogłoski nie wyłączają się wzajemnie. Powyższe przykłady ukazują, że różnice kontekstowe są specyficzne dla określonych języków. Relacje między nimi nie są uniwersalne.

4.3 Różnice dystynktywne

Między dwoma segmentami zachodzi różnica dystynktywna, jeżeli oba mają niektóre otoczenia segmentalne wspólne. Różnica między spółgłoskami bezdźwięcznymi trącymi -- dźwięczną oraz dźwięczną-palatalną jest w polskim języku dystynktywna: Wymówione jako cała wypowiedź wyrazy *proszę* oraz *prosię* lub *kasza* i *Kasia* świadczą, że obie wymienione spółgłoski mogą występować w fonetycznie ekwiwalentnym otoczeniu segmentalnym. Jednakże nie istnieje analogiczna para, której jednym członem byłby wyraz *ruszać* (tzn. nie ma wyrazu **ruszać*)⁴ ani para, której jednym członem byłby wyraz *siano* (nie ma wyrazu **szano*). Polskie samogłoski: przednia wysoka i środkowa wysoka ilustrują różnicę dystynktywną, gdyż z jednej strony istnieją pary wyrazów *bit:byt*, *pił:pył*, *piśk:pysk* itp., ale nie można dobrać w analogiczny sposób pary do wyrazu *lis* (nie ma **lys*), ani do *tyśka* (nie ma wyrazu **tyśka* ani **tyśka*)⁵.

4.4. Różnice osobnicze

Cechy akustyczne segmentów są w znacznym stopniu uzależnione od indywidualnych, tj. osobniczych właściwości narządów mowy. Tradycyjne opisy fonetyczne różnych języków ograniczały się do określania głosek w terminach anatomiczno-fizjologicznych cech artykulacyjnych. Były to (i nadal są w zakresie wiedzy elementarnej) opisy o charakterze jakościowym przy czym bierze się pod uwagę cechy fizjologiczne w skali makro, rozróżniając na przykład spółgłoski wargowe, zębowe, dźwięczne, podniebienne, gardłowe itd., to znaczy stosując określenia odwołujące się do arbitralnie wyróżnionych części toru głosowego.

1

Fonetyka akustyczna posługuje się (tak jak to czyni akustyka - dyscyplina fizyczna) opisem ilościowym. Już w latach pięćdziesiątych stwierdzono, że określone zjawiska fonetyczne, które w zapisie tradycyjnym uważano za "takie same" (bo ich opis kwalitatywny był taki sam) okazują się wyraźnie odmienne m.in. na skutek nie dających się pominąć różnicowań międzyosobniczych. Najbardziej znana z pierwszych prac fonetyczno-akustycznych wskazujących na różnicowania między-

osobnicze i opartą na dostatecznie obszernym materiale doświadczalnym jest wielokrotnie cytowany artykuł Petersona i Barneya (1952). Została ona poprzedzona wstępnymi badaniami z tego zakresu Petersona (1951). Literatura dotycząca tak dystyngtywnych jak i osobniczych zróżnicowań w zakresie zjawisk segmentalnych, a szczególnie częstotliwości formantów samogłoskowych jest już bardzo obszerna (np. Summerfield & Haggard (1975), Pols (1977), Papçun (1980), Lobanov (1971)). Ocena formantów samogłoskowych jako zarazem cech lingwistycznych i osobniczych w polskim języku przedstawiono w pracach: Jassem, Krzyško, Stolarski (1984) oraz Jassem (1986). Problemem, którego dotychczas do końca nie rozpoznano, jest interakcja pomiędzy zróżnicowaniami lingwistycznymi oraz osobniczymi w akustycznym sygnale mowy. Wszystkie akustyczne parametry "krótkoterminowe" (szybkoszienne) są podatne na zmienność tak jednego, jak i drugiego rodzaju.

4.5 Różnice fakultatywne

Określone segmenty mogą pozostawać w relacji wariantów fakultatywnych, jeżeli ich dystrybucja jest taka, że każdy z nich występuje w ekwiwalentnych kontekstach (otoczeniach) fonetycznych. W wyrazach angielskich *four*, *law*, *war* /fo, 'lo, 'wo/ itp. wymawiana jest bądź samogłoska tylna pośrednia między półotwartą a półzamkniętą ze słabą labializacją, bądź tylna półotwarta z silną labializacją. W dowolnym otoczeniu (np. w *caught* - równobrzmiące z *court* /kot/ *laud* - równobrzmiące z *lord* /lod/) występuje bądź jedna, bądź druga samogłoska.

4.6 Różnice wywołane czynnikami suprasegmentalnymi

Zjawiska fonetyczne dzieli się najogólniej na segmentalne i suprasegmentalne. Dziedziną pierwszych są w zakresie analizy sygnału akustycznego kolejne, wyżej zdefiniowane segmenty. Dziedziną drugich są ciągi segmentów, z których najkrótszym (w sensie rozciągłości czasowej) jest sylaba, a najdłuższym całkowita wypowiedź. Dziedzinami pośrednimi taksonomicznie mogą być jednostki rytmiczne lub frazy intonacyjne. Z definicji do czynników suprasegmentalnych nie zalicza się zróżnicowań międzyosobniczych wynikających z uwarunkowań fizjologicznych. Czynniki suprasegmentalne dotyczą wszystkich czterech ogólnych

cech psychoakustycznych: głośności, wysokości, barwy i organizacji czasowej oraz ich korelatów fizyko-akustycznych i mają charakter lingwistyczny lub paralingwistyczny. W pierwszym przypadku mają one właściwości systemowe. Na przykład intonacje frazowe poszczególnych języków stanowią określone systemy przebiegów wysokości tonu (fizycznie: częstotliwości podstawowej). W zakresie organizacji czasowej jako element systemowy można wymienić różne sposoby realizacji izochronizmu. Przepuszczalnie zróżnicowania głośności nie wykazują właściwości systemowych (nie stanowią systemu). Mamy tutaj do czynienia ze zjawiskami paralingwistycznymi (np. znaczne podwyższenie wysiłku głosowego związane z emocją albo z odległością od słuchacza). Również suprasegmentalne zróżnicowania barwy nie są przepuszczalnie systemowe, a zatem są paralingwistyczne (np. jasna barwa głosu związana z uczuciem radości). Zjawiska suprasegmentalne w zakresie głośności i barwy oraz ich fizyko-akustyczne korelaty nie zostały jeszcze należycie zbadane, tym bardziej zatem ograniczona jest wiedza o interakcji tych czynników z cechami segmentalnymi. Liczne badania różnych języków wykazały natomiast, iż dość powszechnie do lat 50-tych (60-tych) utrzymywana teza, iż akcent związany jest z głośnością (natężeniem, amplitudą, poziomem natężenia itd.) jest błędna. Jest to prawdopodobnie problem, w zakresie którego badania fonetyczno-akustyczne w sposób najbardziej drastyczny wykazały błędność pewnej teorii lingwistycznej.

Zjawiska suprasegmentalne wchodzi w różne interakcje ze zjawiskami segmentalnymi. Jak wyżej wspomniano, wiedza w tym zakresie jest nader ograniczona. Wspomnieć jednak wypada dla zilustrowania zagadnienia, że relacje między izochronizmem suprasegmentalnym a iloczusem głoskowym w jęz. ang. ukazano w pracy Jassem, Hill, Witten (1986). Analogiczne relacje badała w jęz. pol. L. Richter (1987). Łobacz (1983) analizowała wpływ tempa mowy na przebiegi formantowe samogłosek polskich.

Braki wiedzy na temat interakcji między zjawiskami suprasegmentalnymi a segmentalnymi stanowią jedno z najpoważniejszych utrudnień w automatycznej identyfikacji elementów segmentalnych w mowie ciągłej oraz związanej z tym automa-

tycznej transkrypcji.

Można dokonać bardziej finezyjnej klasyfikacji zróżnicowań segmentalnych niż to uczyniono powyżej w odcinkach 4.1...4.6. Liczne typy zróżnicowań omawia Labov (1986). Można je jednak podciągnąć pod któreś z tych, jakie omówiono, z konieczności w sposób ogólny, powyżej w niniejszym rozdziale.

5. Iloczas

Iloczasem nazywamy rozciągłość czasową segmentów lub określonych (krótkich) sekwencji segmentów (np. głosek lub sylab). W kontekście fonetyki akustycznej iloczas podlega zatem pomiarowi, przy czym dogodną jednostką jest milisekunda. Iloczas konkretnych segmentów zależy od znacznej ilości czynników, między którymi zachodzą interakcje. Do takich czynników należą: stopień otwarcia jamy ustnej przy artykulacji samogłoskowej (otwarte samogłoski są dłuższe od przymkniętych), kontekst fonetyczny następujący (samogłoski są dłuższe przed spółgłoskami trącymi niż przed zwartymi), tempo mowy i inne. W związku z transkrypcją interesujące są w zasadzie tylko takie zróżnicowania iloczynowe, które są dystynktywne w sensie analogicznym do tego, w jakim rozpatrywaliśmy widmowe zróżnicowania cech segmentalnych w rozdz.4. W niektórych językach określony (pod względem barwy) segment może występować z iloczynem względnie długim lub względnie krótkim, przy czym segment względnie długi i względnie krótki mają niektóre (nie wszystkie) konteksty (otoczenia) wspólne. W jęz. niem. (standardowa forma nieregionalna) samogłoska otwarta pośrednia między przednią a tylną występuje z iloczynem względnie długim w *Kahn, las*, a z krótkim w *kann, laß*. Obie formy mają zatem co najmniej niektóre konteksty wspólne. Jednakże do wyrazu *Klasse* (krótka samogłoska) nie można dobrać odpowiednika **Klaße* ani do *Straße* (długa samogłoska) odpowiednika **Strasse* (z krótką samogłoską). Zatem nie wszystkie wszystkie otoczenia są wspólne dla obu. Różnica iloczynowa jest więc dystynktywna. Podobna sytuacja zachodzi w scenicznej (nieco przestarzałej) formie jęz. francuskiego, jak to ilustruje np. para *maitre :mettre*. Dla zaznaczenia dystynktywnej długości

używa się w transkrypcji DWUKROPKA : niem. /ka:n # 'kan, 'la:s # 'las/, franc. /me:tr # 'metr/ (dwukropek sygnalizuje samogłoskę względnie długą).

Istnieje również iloczas kwazidystynktywny, który zachodzi wtedy, gdy dystynktywna różnica barwy łączy się nierozzerwalnie z iloczasową. W jęz. niem. istnieje para *Röslein: Rößlein*. W pierwszym wyrazie występuje samogłoska przednia, półprzymknięta zaokrąglona, długa, a w drugim - przednia, półotwarta zaokrąglona, krótka. Samogłoski pozostają w relacji dystrybucji dystynktywnej, przy czym zróżnicowanie stopnia otwarcia (różnica w częstotliwości F_1) zawsze występuje łącznie z różnicą w długości. Jeżeli w określonym języku zachodzi choćby jedna opozycja wyłącznie iloczasowa (jak w niem., por. wyż.), to dwukropek w transkrypcji zachowuje się także dla zaznaczenia iloczasu kwazidystynktywnego. Stąd /transkrypcja /rø:slain/ i /rœslain/. W jęz. ang. nie ma opozycji wyłącznie iloczasowych. Iloczas jest kwazidystynktywny. Stąd trzy typy transkrypcji: Najdawniejszy (wychodzący obecnie z użycia), w którym pomijało się w określonych opozycjach różnicę barwy, a symbolizowało iloczasową. W drugim typie transkrypcji (stosowanym m.in. przez Gimsona (1962) i Roacha (1983)) transkrypcja sygnalizuje obie współistniejące różnice (barwy i iloczasu). W trzecim typie transkrypcji pomija się kwazidystynktywny iloczas, a zaznacza się tylko różnice barwy samogłoskowej:

Transkrypcja

	Typ I	Typ II	Typ III
<i>l<u>u</u>ke</i>	'lu:k	'lu:k	'luk
<i>loo<u>k</u></i>	'luk	'lok	'lok
<i>caugh<u>t</u></i>	'kɔ:t	'kɔ:t	'kot
<i>co<u>t</u></i>	'kot	'kɔt	'kot'
<i>bea<u>t</u></i>	'bi:t	'bi:t	'bit
<i>bi<u>t</u></i>	'bit	'bit	'bit

6. Rytm. Akcent rytmiczny.

Ilustrując powyżej zasady analizy dystrybucyjnej w płaszczyźnie segmentalnej posługiwaliśmy się, jako przykładami, określonymi wyrazami. Czyniliśmy to w założeniu, że cytowany wyraz stanowi całkowitą frazę fonologiczną (jak to bywa w krótkich, jednowyrazowych wypowiedziach). Wyraz bowiem nie jest jednostką fonetyczną żadnego rzędu, lecz jednostką gramatyczną (składniowo-morfologiczną). Ciągi segmentów w wielu językach układają się w jednostki fonetyczne wyższego rzędu, mianowicie w jednostki rytmiczne. Model rytmu mowy, oparty na analizie słuchowej, zaproponowano w odniesieniu do jęz. ang. w pracy Jassema (1949). Zmodyfikowaną jego wersję sprawdzono około 30 lat później pomiarowo (Jassem, Hill, Witten 1984) stosując analizę regresyjną. Ten sam model sprawdzono z pomyślnym rezultatem i przy zastosowaniu podobnej analizy statystycznej na akustycznym materiale mowy polskiej: Richter (1987). Na podstawie tych prac można rytm mowy, przynajmniej w odniesieniu do jęz. pol. i ang., opisać następująco:

1. Fraza fonologiczna składa się z całkowitej liczby ścisłych jednostek rytmicznych wykazujących tendencję izochronizmu tzn. tendencję do wyrównania czasu trwania, wobec czego kolejne segmenty wchodzące w ich skład są tym krótsze im większa ich liczba składa się na jednostkę rytmiczną.

2. Ścisła jednostka rytmiczna może być poprzedzona anakruzą, tj. ciągiem segmentów nie wykazującym tendencji izochronizmu, przy czym segmenty wchodzące w skład anakruzy mają minimalny iloczas.

3. Ścisła jednostka¹ rytmiczna wraz z ewentualną poprzedzającą ją anakruzą stanowi pełną jednostkę rytmiczną, która zgodnie z polską terminologią prozodyczną nazwiemy zestrojem akcentowym. Transkrypcję ukazującą zestroje akcentowe ukazano na tablicach 2 i 3.

4. Pierwsza lub jedyna sylaba ścisłej jednostki rytmicznej jest sylabą akcentowaną rytmicznie (sylabą o akcencie rytmicznym).

Angielska wypowiedź /ajken'fajnju/ składa się z anakruzy /ajken/ oraz ścisłej jednostki rytmicznej /fajnju/. Akcentowana

jest sylaba /fajn/. Wersja ortograficzna tej wypowiedzi jest dwojaka: *I can fine you* lub *I confine you*. Zachodzi tu zatem różnicowanie gramatyczne i znaczeniowe (semantyczne), które nie znajduje odzwierciedlenia w wymowie, a zatem w sygnale akustycznym. Zasada wzajemnej jednoznaczności wymaga, by brakowi różnicowania fonetycznego odpowiadał brak różnicowania w transkrypcji. Stąd właśnie forma /ajkan'fajnju/ w obu przypadkach. Zasady ścisłej transkrypcji fonologicznej odzwierciedlającej akustyczno-fonetyczne zjawiska rytmu i akcentu bez względu na formę ortograficzną są następujące:

1. Zestrój akcentowy zapisuje się jako nieprzerwany ciąg liter fonetycznych.
2. Spacje oddzielają kolejne zestroje akcentowe.
3. Sylabę akcentowaną rytmicznie zaznacza się przez umieszczenie przed nią znaku akcentu.
4. Jeśli znak akcentu nie występuje na początku zestroju akcentowego, to ciąg liter fonetycznych pomiędzy bezpośrednio poprzedzającą spacją a znakiem akcentu odpowiada anakruzie.

Transkrypcje na Tabl. 2 i 3 ilustrują powyższe zasady.

7. Wysokość tonu. Akcent intonacyjny.

Zróżnicowania tonu (częstotliwości podstawowej) w mowie są wykorzystywane w poszczególnych językach bądź w dwóch płaszczyznach struktury (są to tzw. języki tonalne), bądź tylko w jednej (języki intonacyjne). Dystrybucyjną definicję tego zróżnicowania podano w pracy Jassema (1978). Egzemplifikację dwupłaszczyznowej funkcji wysokości tonu znaleźć można m.in. w monografii Jonesa (1950) w rozdz. XXV, wraz z ukazaniem metod ich oznaczania w transkrypcji. W niniejszym opracowaniu uwzględniamy jeden język typu tonalnego, mianowicie szwedzki. Zróżnicowania są tutaj znacznie prostsze niż w większości azjatyckich i afrykańskich języków tego typu (np. chiński-pekiński, chiński-kantoński, thai, burmański, sudański, hottentocki i in.). Wyróżnimy (por. np. Elert 1964) akcent tonalny złożony (symbol ^x), akcent tonalny prosty (symbol ') oraz akcent zredukowany (symbol :): *tanke*n /^xtanke)n/ ("myśl"), *tanke*n /tanke)n/ ("kanister"), *tanke*nanstengning /^xtanke)n,anstrengning/

("wysiłek umysłowy").

Istnieje kilka systemów znaków transkrypcyjnych służących do oznaczania jednopoziomowej struktury intonacyjnej w takich językach jak polski, niemiecki, angielski, francuski i in. Tylko nieliczni autorzy włączają te znaki do bieżącej transkrypcji segmentalnej (np. Thudicum 1926, Bayer 1929). Ponieważ istnieją wciąż daleko idące rozbieżności zdań co do struktury intonacji, nawet tak szeroko pod tym względem badanego języka jak angielski (standardowa forma brytyjska i amerykańska), w odniesieniu zaś do znakomitej większości języków badania intonacji są bardzo słabo zaawansowane, w obecnym opracowaniu ograniczymy się do jej uwzględnienia tylko w zakresie jej funkcji akcentuacyjnej. Pragniemy przy tym przypomnieć, że konkretne przebiegi ekstrahowanego z sygnału parametru zmiennej czasowo częstotliwości podstawowej wymagają z jednej strony rozpracowania szeregu podstawowych zagadnień natury matematycznej (Surmanowicz-Demenko 1987) oraz że konkretny przebieg parametru F_0 zależy jest od kilkunastu czynników (Jassem, Demenko 1986), których dokładne zbadanie jest warunkiem koniecznym uzyskania możliwości automatycznego rozpoznawania strukturalnych jednostek tonu i intonacji w mowie celem ich automatycznego zapisu transkrypcyjnego.

W językach rozpatrywanych w niniejszym opracowaniu sylaba akcentowana intonacyjnie stanowi początek dystyngtywnej konfiguracji intonacyjnej (por. Jassem 1978, 1987). W transkrypcji będziemy oznaczać: akcent główny (tzw. intonację rdzenną, ang. nuclear tone) za pomocą znaku " | ", zaś akcent poboczny (w intonacji nierdzennej lub akcent wyłącznie rytmiczny) za pomocą znaku " ; " np. w ang. *Take a walking stick* /tejk a'wokin stik/, w pol. *Tu stoi lampa naftowa* /tu stoi 'lampa naf.tova/. W jez. ang. istnieje opozycja akcentuacyjna: *I saw a moving van* /aj,so a'muwvɪn 'ven/ ("Widziałem wóz meblowy"); *I saw a moving van* /aj'so e,muwvɪn 'ven/ ("Widziałem poruszającą się ciężarówkę").

8. Zarys taksonomii jednostek segmentalnych

W dotychczasowych rozważaniach związek między literami i

znakami występującymi w transkrypcji a sygnałem akustycznym traktowany był ogólnikowo, bądź celem ukazania zawiłych nieraz relacji między formą ortograficzną a fonetyczną (rozdz.2), bądź dla zilustrowania w transkrypcji iloczasu i struktur suprasegmentalnych (rozdz. 5, 6 i 7). Relacje między tymi znakami i literami fonetycznymi a sygnałem akustycznym nie zostały na razie bliżej określone.

Celem niniejszego rozdziału jest dokonanie wstępnej próby taksonomii segmentów akustyczno-fonetycznych, która po przeprowadzeniu dalszych (sugerowanych m. in. w poprzednich rozdziałach) pracach badawczych i eksperymentalno-pomiarowych oraz po odpowiednim sformalizowaniu, mogłaby doprowadzić do stworzenia algorytmu automatycznej transkrypcji sygnału mowy opartej na analizie cyfrowej i do automatycznej syntezy mowy na podstawie fotoelektrycznego lub laserowego odczytu transkrypcji.

1. Dowolna prawidłowa wypowiedź jest sygnałem akustycznym podlegającym analizie różnymi metodami. Metody te umożliwiają opis dowolnego analizowanego fragmentu sygnału w terminach określonych miar i z określoną dokładnością, w szczególności zaś dowolnego fragmentu, który zdefiniowano powyżej jako segment (fonetyczny). Opis każdego segmentu zawiera ilościowo określone wszystkie cechy każdego segmentu, które umożliwiają dalszą jego klasyfikację.

2. Jeżeli dwa segmenty nie wykazują różnic pomiarowych lub wykazują różnice uznane za podprogowe albo losowe, to stosujemy do nich, także jeśli ich liczba jest większa od dwóch klasyfikację przez ekstensję i jako grupę uznajemy je za stanowiące segment fonologiczny najniższego rzędu w układzie hierarchiczno-klasyfikacyjnym.

Dalej będziemy używać pojęcia segmentu w sensie zbiorowym tj. w znaczeniu segmentu fonologicznego, a nie w sensie jednostkowym.

3. Segmenty (fonologiczne), które wykazują wyłącznie różnice koartykulacyjne, stanowią idiofon.

4. Idiofony, pomiędzy którymi zachodzą wyłącznie różnice osobnicze, stanowią unifon.

Segment jednostkowy, fonologiczny oraz idiofon i unifon są elementami uniwersalnymi, tj. są niezależne od fonetycznych

właściwości poszczególnych języków. Kolejne stadia taksonomiczne mają już charakter specyficzny, tj. uwzględniający stan rzeczy w określonych językach.

Z pragmatycznego punktu widzenia różnica między uniwersalnym a specyficznym poziomem klasyfikacyjno-taksonomicznym może być istotna w automatycznym rozpoznawaniu elementów segmentalnych oraz ich automatycznej transkrypcji. Techniczne układy dokonujące rozpoznawania w zakresie płaszczyzny uniwersalnej mogą stanowić standard międzynarodowy. Układy klasyfikujące i rozpoznające na poziomie specyficznym muszą realizować algorytmy skonstruowane oddzielnie dla każdego języka (i ewentualnie dialektu).

5. Każdy język (dialekt) wykorzystuje skończoną liczbę unifonów. Większość z nich występuje w bardzo różnorodnych otoczeniach segmentalnych. Takie unifony określamy jako samodzielne. Niektóre jednak unifony w danym języku występują tylko w bardzo wąsko określonym kontekście fonetycznym i zawsze w określonej kolejności. Na przykład w wyrazie *dbać* występują dwie samogłoski, wyraźnie widoczne na spektrogramie (ryc. 4), z których pierwsza jest znacznie krótsza niż druga i ma artykulację centralną objawiającą się w mniej więcej równych odstępach między F_1 , F_2 i F_3 . Taka samogłoska występuje w polskim języku tylko w nielicznych ściśle określonych kontekstach, m.in. między dwoma segmentami stanowiącymi zwarcia dźwięczne. W polskim języku dwa zwarcia dźwięczne nie występują nigdy bezpośrednio po sobie. W jęz. ang. relacje są dokładnie odwrotne. W wyrazie-wypowiedzi takim jak *abdomen* (por. ryc. 4) dwa segmenty zwarte dźwięczne następują po sobie bezpośrednio, samogłoska zaś bardzo podobna do tej, którą ukazaliśmy powyżej w polskim wyrazie *dbać*, w jęz. ang. jest segmentem samodzielnym i to najczęstszym spośród tzw. segmentów wokalicznych.

Samodzielny unifon oraz ciąg unifonów, które w danej pozycji w danym języku występują zawsze łącznie i w określonej kolejności, nazywamy głoską. Samodzielne unifony są głoskami monosegmentalnymi, a wyżej określone ciągi unifonów, w skład których wchodzi unifony niesamodzielne, są głoskami polisegmentalnymi.

FONEM jest to pojedyncza głoska wchodząca z innymi

głoskami danego języka w relacje dystynktywności, albo taki zbiór głosek ("a", "b", "c"...), z których każda oddzielnie pozostaje w relacji dystynktywności z co najmniej jedną inną głoską danego języka ("a:l...", "b:m...", "c:n"...), ale które względem siebie nie są dystynktywne, a zatem pomiędzy którymi zachodzi relacja bądź wariacji kontekstowej (pozycyjnej) bądź fakultatywnej, przy czym głoski należące do jednego, określonego fonemu, czyli jego ALOFONY muszą mieć określony zbiór wspólnych cech fonetycznych.

Alofony danego fonemu, które nie mają żadnego kontekstu wspólnego, nazywamy alofonami POZYCYJNYMI, zaś alofony, które mają wszystkie konteksty wspólne - FAKULTATYWNYMI.

9. Transkrypcja

W tradycyjnym ujęciu transkrypcja określonego tekstu lub określonej wypowiedzi odnosi się do aspektu psychologicznego sygnału mowy (por. str. 6) odbioru sygnału mowy przez osobę dokonującą słuchowej analizy tego sygnału. Transkrypcja taka może być przeprowadzona przez specjalistę znającego dany język lub przez fonetyka nie znającego języka, jak to ma miejsce w przypadku badania nowych języków i dialektów dotychczas nie opisanych. Poprawność takiej transkrypcji nie ma dotychczas podstaw naukowych i zasadza się na zgodności opinii raczej niż na określonych prawidłach i ogólnych przesłankach psychoakustycznych. Na ogół biorąc, taka zgodność opinii istnieje, dzięki czemu powstają słowniki wymowy jakoteż teksty (z reguły dydaktyczne), w których dłuższe wypowiedzi zapisane są w transkrypcji.

Zgodność opisu fonetycznego w płaszczyźnie segmentalnej (a także suprasegmentalnej) opartego wyłącznie na wrażeniach słuchowych jest jednak względna z uwagi na nieuniknione elementy subiektywizmu (por. str. 6). Poza tym kryteria sprowadzające wyniki analizy słuchowej do kodu fonematycznego nie są dostatecznie uściślone. Stwierdzono ponadto, że w poszczególnych językach pojawiają się sporadycznie unifony nie poddające się regułom interpretacji fonematycznej. Z tych powodów liczne języki, zwłaszcza bardziej znane, zapisywane są

w transkrypcji na kilka czasów sprzecznych sposobów. Niektóre kontrowersje dotyczące analizy i transkrypcji mogą być rozstrzygnięte poprzez włączenie analizy fonetyczno-akustycznej (por. Łobacz 1973).

Jednym z celów niniejszego opracowania jest znalezienie przesłanek do powstania teorii, która umożliwiłaby opracowanie algorytmu automatycznej transkrypcji w oparciu o wyniki analizy akustycznej sygnału mowy w połączeniu z możliwie jak najbardziej określonymi operacjami matematyczno-logicznymi. Tymi operacjami są:

(a) segmentacja;

(b) ekstrakcja określonych parametrów opisujących dowolne fragmenty sygnału, w szczególności jednak wydzielone segmenty fonetyczno-akustyczne:

(c) ustalenie segmentów (α) nie różniących się pomiarowo (β) różniących się o mniej niż z góry założona wartość progowa, (γ) różniących się wyłącznie losowo;

(d) ustalenie różnic między segmentami wynikających wyłącznie z uwarunkowań koartykulacyjnych;

(e) ustalenie różnic między segmentami wynikającymi wyłącznie z uwarunkowań osobniczych;

(f) ustalenie interakcji między cechami segmentalnymi a czynnikami suprasegmentalnymi;

(g) dokonanie analizy dystrybucyjnej unifonów i głosek;

(h) klasyfikacja prowadząca do zaliczenia określonych głosek reprezentowanych w tekście do odpowiednich fonemów, a w obrębie każdego fonemu - do jego alofonów (jeśli fonem obejmuje alofony).

Z definicji i wyjaśnień poprzedniego rozdziału wynika, że w skład określonego fonemu może w jednym przypadku wchodzić tylko jedna głoska, podczas gdy inne fonemy mogą być reprezentowane przez dwie lub więcej głosek. W tym drugim przypadku głoski reprezentujące ten sam fonem są, według dotychczasowych definicji, jego alofonami.

Dla celów transkrypcji dogodnie jest przyjąć zasadę, że w przypadku fonemu reprezentowanego przez jedną tylko głoskę, będziemy taką pojedynczą głoskę też nazywać alofonem tego fonemu. Uprości to taksonomię i oparcie na niej transkrypcję.

Ponieważ w ten sposób każda głoska określonego języka jest alofonem jakiegoś fonemu, transkrypcję, która przypisuje wzajemnie jednoznacznie odrębny znak każdej głosce, nazywamy TRANSKRYPCJĄ ALOFONICZNĄ.

Jeśli każdemu fonemowi danego języka przypisujemy jeden znak i określony tekst zapisujemy na podstawie przypisania każdej głoski określonemu fonemowi, to taką transkrypcję nazywamy FONEMATYCZNĄ.

W pewnych przypadkach może zachodzić potrzeba wyodrębnienia w transkrypcji jednej czy kilku głosek niezależnie od jej (ich) przynależności fonematycznej. Transkrypcję, w której niektóre znaki odnoszą się do głosek bez względu na ich interpretację fonematyczną, a pozostałe symbolizują fonemy, nazywamy KWAZIALOFOFONICZNĄ.

Przykłady: Wymówione w izolacji wyrazy *rzecz*, *wieś*, *sen* oraz *dzień* zawierają jako element samogłoskowy cztery różne segmenty wokaliczne: (1) przedni półotwarty ustny, (2) przedni półprzymknięty ustny, (3) przedni półotwarty nosowy oraz (4) przedni półprzymknięty nosowy. Wszystkie należą jako alofony pozycyjne do fonemu samogłoskowego przedniego o nieskrajnym stopniu otwarcia. Zapis fonematyczny tych wyrazów przedstawia się następująco: /ʒetʃ, 'vjeɕ, 'sɛn, 'dʑɛɲ/, zaś zapis alofoniczny [ʒetʃ, 'vjeɕ, 'sɛn, 'dʑɛɲ] (por. ryc. 5).

Polski fonem spółgłoskowy nosowy palatalny obejmuje m.in. alofon, który jest głoską wokaliczną, przednią wysoką i nosową. Jeśli istnieje potrzeba zasygnalizowania w transkrypcji, że w danym fragmencie tekstu występuje ta szczególna głoska, podczas gdy reszta wypowiedzi interesuje nas jedynie na poziomie fonematycznym, to możemy zastosować transkrypcję kwazialofoniczną: *Czy to jest pański brat?* /tʃitojɛs paŋjɫsci 'brat/.

W literaturze językoznawczej oraz w podręcznikach języków obcych spotkać można kilka systemów transkrypcyjnych. Liczne z nich są niekonsekwentne pod względem poziomu abstrakcji, na którym dokonano transkrypcji: pomieszanu ulega poziom głoskowy (alofoniczny) i fonematyczny, co prowadzi często do nieporozumień. Poza tym stosuje się różne systemy literowe. Polski język jest zapisywany dwoma różnymi systemami literowymi

w transkrypcji: sławistycznym (starszym) i międzynarodowym (nowszy).

Stosowanie alfabetu fonetycznego ustalonego przez Międzynarodowe Stowarzyszenie Fonetyczne (International Phonetic Association) znajduje w ostatnich latach w Polsce coraz szersze uznanie, tak w odniesieniu do języków obcych, jak i do rodzimego. Jedyny słownik wymowy polskiej (M. Karaś i M. Madejowa, 1984) stosuje zasadniczo transkrypcję międzynarodową, chociaż w formie niekonsekwentnej, tak pod względem interpretacji fonematycznej, (pomieszczenia płaszczyzn alofonicznej i fonematycznej), jak i przez włączenie niektórych znaków spoza alfabetu międzynarodowego.

10. Alfabet fonetyczny IPA

Historię powstania międzynarodowego alfabetu fonetycznego IPA przedstawiono w (przestarzałej już co prawda) pracy Albright (1958). Podstawy fonetyczne tego alfabetu znaleźć można w pełnej formie w *The Principles* (1949), a następnie dalsze uzupełnienia w różnych zeszytach *Le Maître Phonétique* oraz *Journal of the International Phonetic Association*.

W bieżącym roku (1987) toczy się dyskusja w sprawie licznych problemów teoretycznych i praktycznych związanych z międzynarodowym alfabetem fonetycznym. Ma on być poddany rewizji na zjeździe członków IPA w r. 1989. Najpilniejsze zagadnienia, wokół których toczy się obecnie dyskusja omówiono w numerze 1 vol.17 *JIPA*: Ladefoged (1987) oraz Henton (1987).

11 i 12 czerwca 1987 odbyła się sesja poświęcona komputerowemu kodowaniu fonetycznemu w związku z koniecznością zebrania dużej bazy danych fonetycznych, głównie dla celów dwóch wielkich zachodnioeuropejskich planów naukowych w zakresie tzw. "technologii mowy" (Speech Technology) z kilkusetmilionowymi nakładami (w przeliczeniu na dol. USA), mianowicie plan Alvey i plan Esprit. Dyskutowano sprawę (a) transliteracji wykorzystującej wyłącznie znaki kodu ASCII o numerach 32-127 (dec) dostępnych bezpośrednio z typowej klawiatury mikrokomputera oraz o transkrypcji kodowanej, tzn. zdefiniowaniu specjalnych liter fonetycznych i przypisaniu ich

w miejsce znaków nie używanych w transkrypcji (przede wszystkim dużych liter i cyfr). Wyniki dyskusji podsumowano w raporcie Wellisa (Wells 1987).

Zostały już zdefiniowane wszystkie liczne znaki fonetyczne, których można używać pod różnymi programami edycyjnymi na mikrokomputerach Macintosh (Ladefoged 1937). Natomiast w chwili pisania niniejszej pracy autorom nie wiadomo o istnieniu alfabetu fonetycznego zdefiniowanego w kodzie nadającym się do realizacji na komputerach typu IBM PC XT i AT, w zastosowaniu do najważniejszych języków europejskich.

W ramach niniejszej pracy zdefiniowano przy użyciu programu LETTRIX 63 znaki fonetyczne nie istniejące w standardowym kodzie ASCII (ae jest w odmianie duńskiej). Wybrano te znaki literowe systemu IPA, które służą do transkrypcji fonematycznej (w kilku jej odmianach) następujących języków: angielski (standard brytyjski i amerykański), francuski, niemiecki, rosyjski, polski, szwedzki i włoski. Wszystkie na nowo zdefiniowane znaki wraz z wykazem, w którym z wymienionych języków dany znak fonetyczny jest używany, widnieją w porządku kwazi-alfabetycznym w Tabelicy 1 a/b.

Dalsze tablice ilustrują różne zastosowania transkrypcji fonematycznej w systemie IPA: Tablica 2 zawiera standardowy tekst (por. *Principles* 1949) w wersji polskiej z transkrypcją ściśle fonologiczną, z uwzględnieniem rytmu i akcentu w zestrojach akcentowych. Jest to zatem transkrypcja bliższa sygnału akustycznego niż tradycyjna (wyrazowa). Na Tabelicy 3 ta sama bajka Ezopa, w wersji angielskiej, pokazana jest w dwóch różnych odmianach transkrypcji. Pierwsza jest transkrypcja wyrazową z zaznaczeniem kwazidystyngtywnego iloczasu. Druga jest ściśle fonologiczna. Tablica 4 zawiera listę fonemów standardowej angielszczyzny brytyjskiej. W pierwszej linii widnieje znak fonemu, w drugiej przykład w wersji ortograficznej, a w trzeciej transkrypcja przykładowego wyrazu, ta sama, co u góry w Tabelicy 4. Ostatni potrójny wiersz zawiera spółgłoski sylabiczne. Tablica 5 pokazuje transkrypcję języka rosyjskiego. Jest to według większości możliwych interpretacji, transkrypcja kwazialofoniczna z wyodrębnionymi

alofonami nieakcentowanymi. Nie zastosowaliśmy jednak wyodrębnienia kursywą, albowiem istnieją i takie interpretacje, w których samogłoski nie akcentowane jęz. ros. stanowią oddzielny system fonemów.

Na Tablicy 6.1 i 6.2 ukazano dla języków: polskiego, angielskiego, francuskiego, rosyjskiego niemieckiego, szwedzkiego i włoskiego specjalne litery fonetyczne potrzebne do transkrypcji. Każdy wiersz składa się z dwóch linii. W górnej podano kod wywołujący ze standardowej klawiatury mikrokomputera, a pod kodem wywołany znak fonetyczny. Kodowanie jest we wszystkich istotnych punktach zgodne z raportem Wellisa (1987), z następującymi uzupełnieniami :

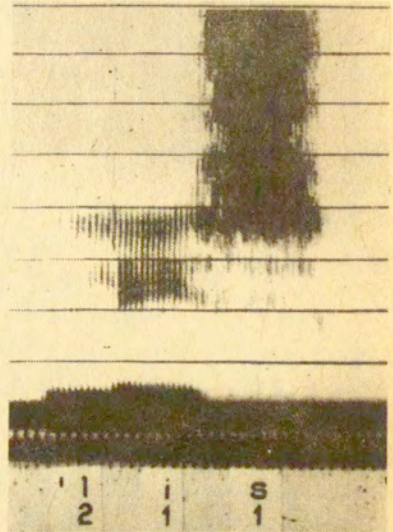
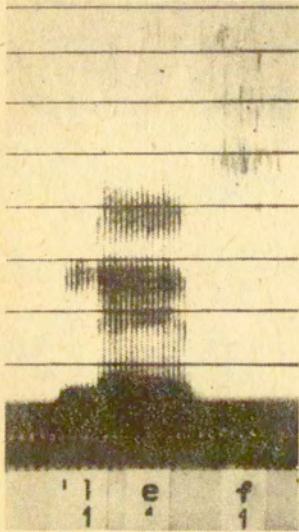
1.1. Afrykaty (fonemy zwarto-trące) zdefiniowano pod pojedynczym kodem. Upraszcza to transkrypcję oraz ewentualną pracę statystyczną, gdyż każdemu naciśnięciu klawisza odpowiada jeden fonem. Odróżnienie afrykat od sekwencji "zwarta plus trąca" konieczne jest dla rozróżnienia takich par jak *czy# trzy /tʃi # 'tʃi/* albo *podzielimy # podziemnymi /pɔdʲɛlimi # podzem'nimi/*. Ponieważ jednak wielu fonetyków unika stosowania ligatur dla oznaczenia afrykat, zakodowaliśmy je cyframi 1...6. Dla wszystkich uwzględnionych tu języków zdefiniowano również znaki akcentu frazowego głównego i pobocznego :

	akcent	akcent
	główny	poboczny
kod	"	:
znak		

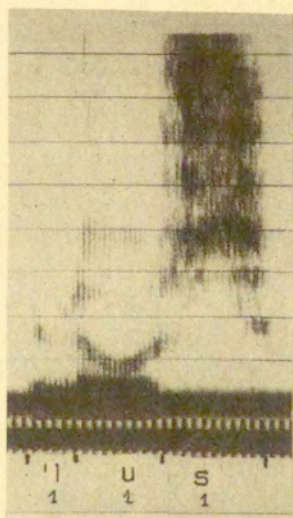
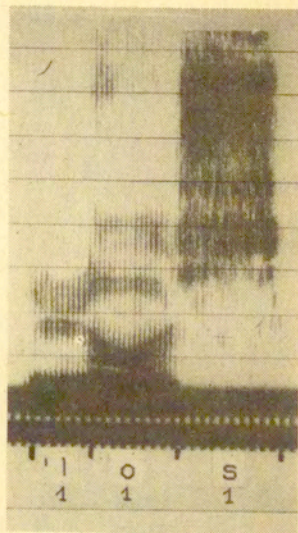
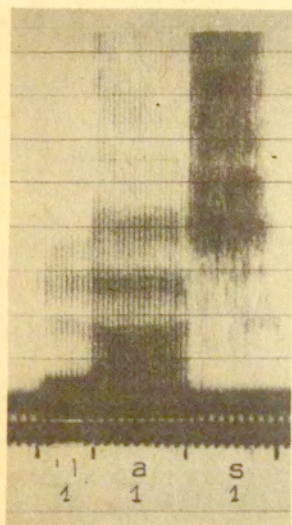
a dla języka szwedzkiego dodatkowo akcent złożony (gravis) (zob. swed.ipa).

Ryc. 6 ukazuje spektrogramy wycinków polskiej i angielskiej wersji bajki Ezopa z literami transkrypcji fonematycznej i zaznaczonymi granicami międzysegmentalnymi. Cyfry pod znakami transkrypcji oznaczają liczbę segmentów składającą się na głoskę reprezentującą dany fonem.

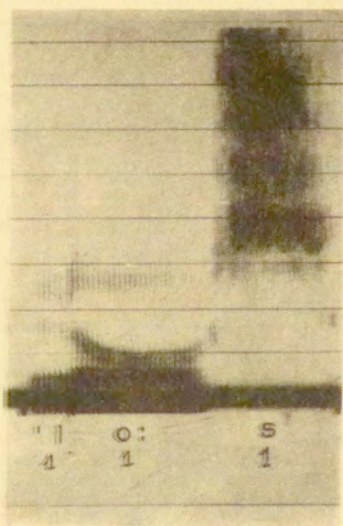
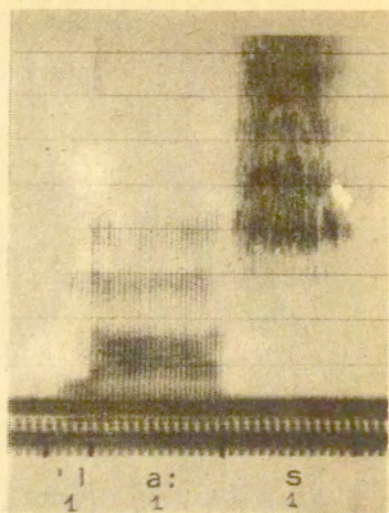
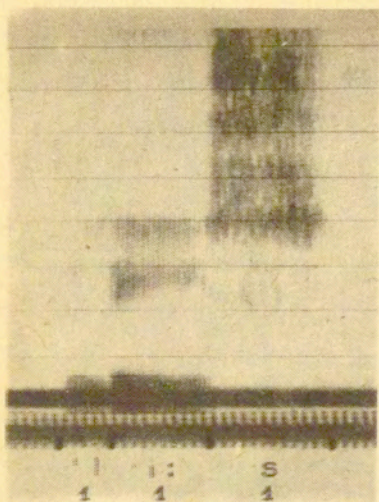
Tablica 7 ukazuje w powiększeniu (uzyskanym pod programem LETTRIX) zdefiniowane przez nas litery fonetyczne w piśmie prostym i w kursywie.



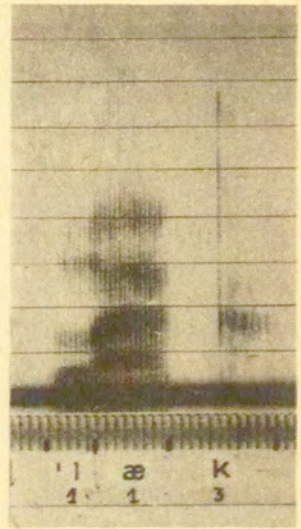
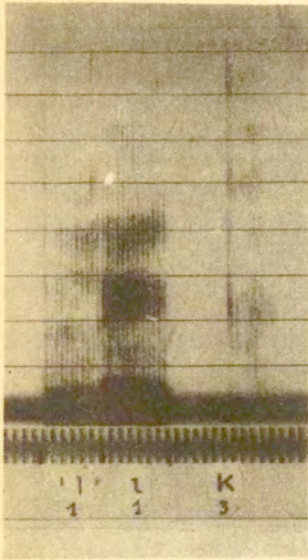
Ryc. 1. Polskie wyrazy lew, lis.



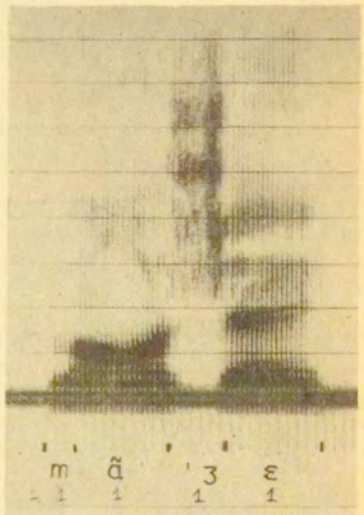
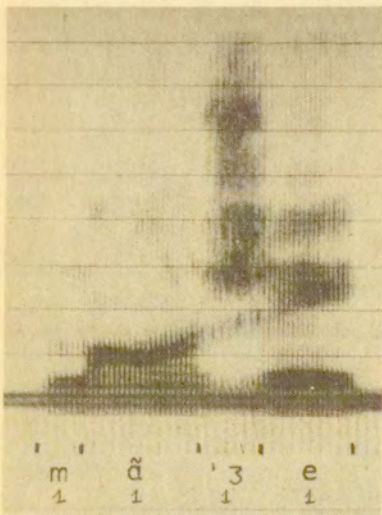
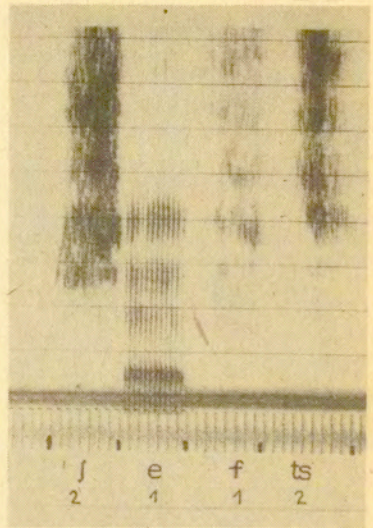
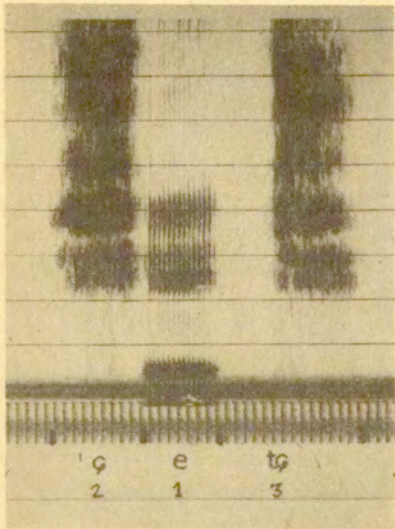
Ryc. 1. Zróżnicowania koartykulacyjne:
polskie wyrazy las, los, luz.



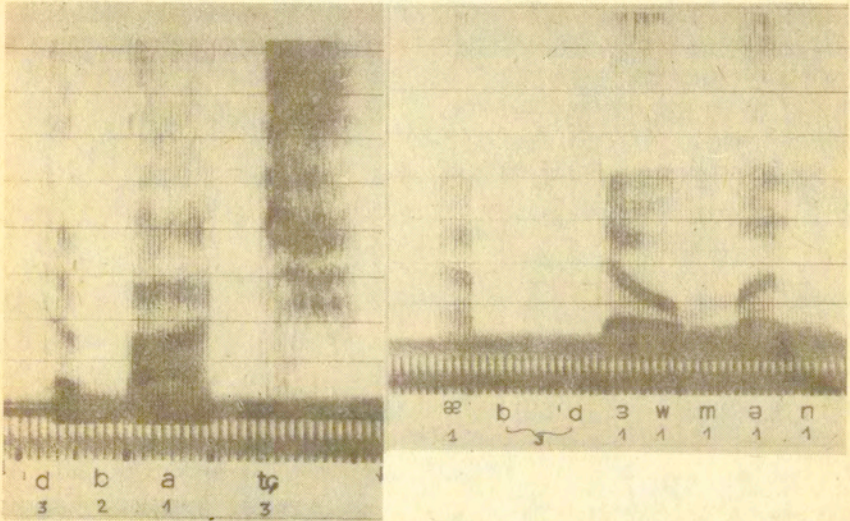
Ryc. 1. Niemieckie wyrazy *Lies*, *Las*, *Los*.



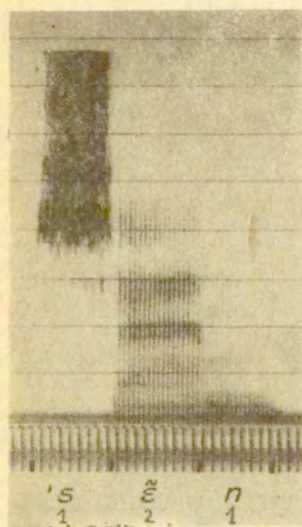
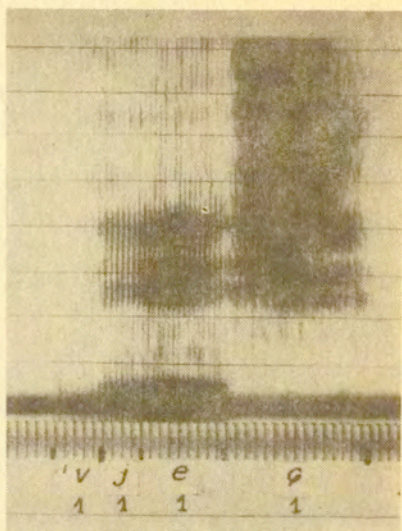
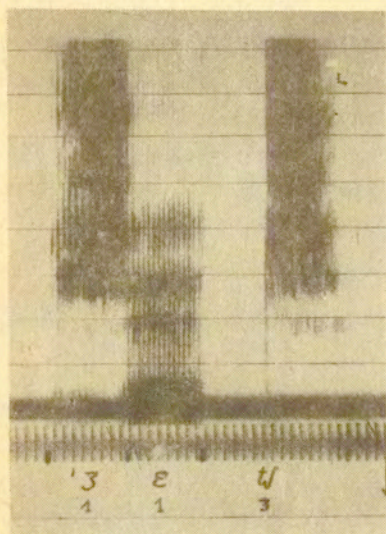
Ryc.1. Angielskie wyrazy *lick*, *lack*, *lock*, *look*.



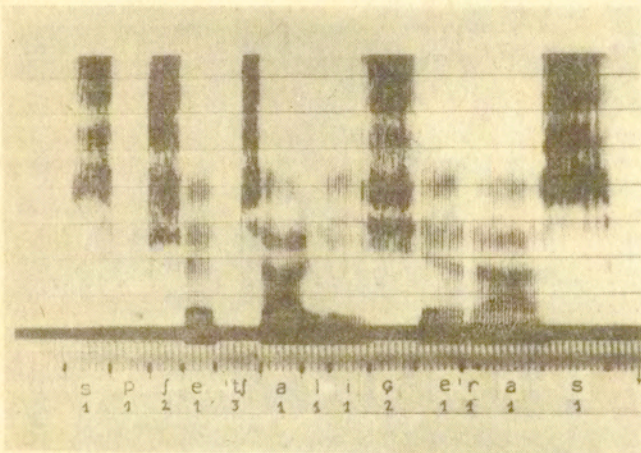
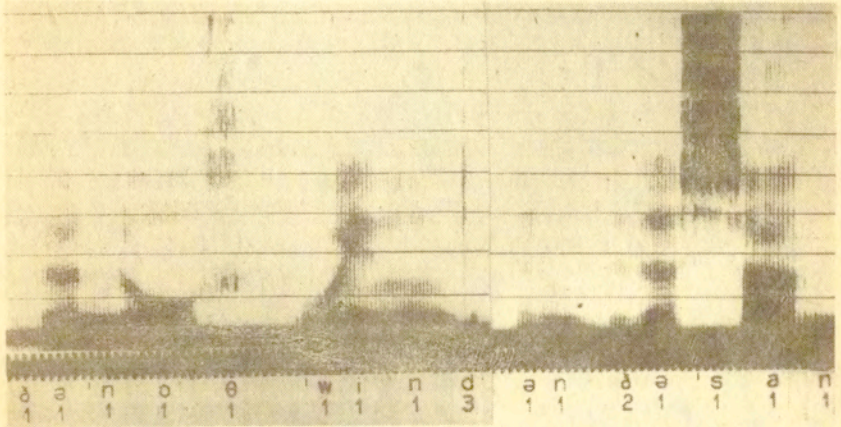
Ryc. 3. Spektrogramy wyrazów francuskich *manger* i *mangeait*.



Ryc. 4. Spektrogramy pol. wyrazu *dbać* oraz ang. *abdomen*.



Ryc. 5. Spektrogramy wyrazów rzecz, wieś (wiesz), sen, dzień.



Ryc. 6. Spektrogramy wycinków bajki Ezopa w wersji angielskiej i polskiej z zaznaczoną segmentacją i transkrypcją.

SPECIAL IPA CHARACTERS

character	ɑ	ʌ	ɛ	ɒ	æ	ɔ	ɔ	ɔ	ɔ	ɔ
Br. Engl.	RP	RP	-	RP	RP	-	-	-	-	RP
Gen. Am.	GA	GA	-	GA	GA	-	-	-	-	GA
French	F	-	-	-	-	-	-	-	-	-
German	G	-	G	-	-	-	-	G	-	-
Russian	-	R	-	-	-	R	-	-	-	-
Swedish	S	-	-	-	-	-	-	S	S	-
Polish	-	-	-	-	-	-	P	-	-	-
Italian	-	-	-	-	-	-	-	-	-	-
character	ɔ	ɔ	ɔ	ɔ	ə	ɛ	ɜ	f	ŋ	l
Br. Engl.	-	-	RP	-	RP	RP	RP	-	-	RP
Gen. Am.	-	-	GA	-	GA	GA	-	-	-	GA
French	-	-	-	-	F	F	-	-	-	-
German	-	-	G	-	G	G	-	-	-	G
Russian	R	-	R	-	R	-	-	R	-	R
Swedish	-	-	-	-	S	S	-	-	S	S
Polish	-	P	P	P	-	-	-	-	-	-
Italian	-	I	I	-	-	I	-	-	-	-
character	i	j	k	l	ʌ	ɪ	l	m	n	ŋ
Br. Engl.	-	-	-	-	-	-	-	-	-	-
Gen. Am.	GA	-	-	-	-	-	-	-	-	-
French	-	-	-	-	-	-	-	-	F	-
German	-	-	-	-	-	-	-	-	-	-
Russian	R	-	R	R	-	-	-	R	-	R
Swedish	-	-	-	-	-	-	S	-	-	-
Polish	P	P	-	-	-	P	-	-	P	-
Italian	-	-	-	-	I	-	-	-	I	-
character	ŋ	ŋ	ʃ	ʃ	ɔ	ɔ	ɔ	ɑ	ɑ	ɪ
Br. Engl.	-	RP	-	RP	-	RP	-	-	-	-
Gen. Am.	-	GA	-	GA	-	GA	-	-	-	-
French	-	-	F	-	-	F	F	-	-	-
German	-	G	G	G	-	G	G	-	-	-
Russian	-	-	-	R	-	-	-	-	R	R
Swedish	S	S	S	S	S	S	S	-	-	-
Polish	-	P	-	-	-	-	-	-	-	-
Italian	-	-	-	-	-	I	-	-	-	-
character	ʃ	ʃ	ʃ	ʃ	ʃ	θ	θ	ʃ	ʃ	ʃ
Br. Engl.	-	-	RP	-	-	RP	-	RP	-	-
Gen. Am.	-	-	GA	-	-	GA	-	GA	-	-
French	-	-	F	-	-	-	-	-	-	-
German	-	-	G	-	-	-	-	G	-	-
Russian	R	-	R	R	-	-	R	R	-	-
Swedish	-	S	-	-	S	-	-	-	-	-
Polish	-	-	P	-	-	-	P	P	P	-
Italian	-	-	I	-	-	-	I	I	-	-

Tablica 1.

Specjalne litery transkrypcyjne

character	Y	W	V	U	Y	Z	Z	Z	3	?
Br. Engl.	-	-	-	-	-	-	-	-	RP	RP
Gen. Am.	-	-	-	-	-	-	-	-	GA	-
French	-	-	-	F	-	-	-	-	F	-
German	-	-	-	-	G	-	-	-	G	G
Russian	-	-	R	-	-	-	R	-	R	-
Swedish	-	-	-	-	S	-	-	S	-	-
Polish	-	-	-	-	-	P	-	-	P	-
Italian	-	-	-	-	-	-	-	-	-	-

character	I	I	I
Br. Engl.	RP	RP	-
Gen. Am.	GA	GA	-
French	F	F	-
German	G	G	-
Russian	R	R	-
Swedish	S	S	S
Polish	P	P	-
Italian	I	I	-

(Tablica 1. - c.d.)

puw.notsni ,wjatr i'swontse / spje'tjalige / ,ktoznix
jesqil'nejfi / ,vwaqne pje.xoqiw ,drogõ jaciq.tjwowjek
ovi,peni ,ftqepwi 'pwaftj / umu,vilige vobets'tego / ze.tenznix
kturi,pjerfi ,zmuqi tego'two'vjeka / zebi,zdjowqgebbe
o'kritqe / berqeuwa,zani zaqilnej'fego / puw.notsni 'wjatr /
zatqowod,razu ,doptq ,stsavej 'qivi / aleim,motsnej 'dow /
tim,wjentsej pod,ruczni otulawqe 'fpwaftj / ,fkoptsu puw.notsni
,wjatr daw'spokuj / ,ften ,sposup puw.notsni ,wjatr ,muqaw
'pfiznatq / ze ,swontse jesqil'nejfe od'nego /

Północny wiatr i słońce sprzeczały się, kto z nich jest silniejszy. Właśnie przechodził drogą jakiś człowiek owinięty w ciepły płaszcz. Umówili się wobec tego, że ten z nich, który pierwszy zmusi tego człowieka, żeby zdjął z siebie okrycie, będzie uważany za silniejszego. Północny wiatr zaczął od razu dąć z całej siły. Ale im mocniej dął, tym więcej podróżny otulał się w płaszcz. Wreszcie północny wiatr dał spokój. Wtedy słońce zaczęło przygrzewać i w chwilę później podróżny zdjął płaszcz. W ten sposób północny wiatr musiał przyznać, że słońce jest silniejsze od niego.

Tablica 2.

Bajka Ezopa *Północny wiatr i Słońce* w transkrypcji fonologicznej (fonematycznej) i w wersji ortograficznej.

English: RP (1)

ðə 'nɔ:θ 'wɪnd ən ðə 'sʌn wə dɪ'spju:tɪŋ 'wɪf wəz ðə
'strɒŋgə, wen ə 'trævlə ,keɪm ə'lɒŋ 'ræpt ɪn ə 'wɔ:m 'klɒk.
ðeɪ ə'grɪ:d ðæt ðə ,wʌn hu:(ɪ) 'fə:st sək'sɪ:dɪd ɪn'meɪkɪŋ ðə
'trævlə 'teɪk ɪz ,klɒk 'ɔf ʃəd bi:(ɪ) kən'sɪdəd 'strɒŋgə
ðən ðɪ ,ʌðə. 'ðen ðə 'nɔ:θ 'wɪnd 'blu: əz 'hɑ:d əz i:(ɪ) 'kɒd,
bət ðə 'mɔ:r i:(ɪ) 'blu: ðə,mɔ: 'klɒsɪl dɪd ðə 'trævlə
'fəwɪd ɪz 'klɒk ə'rəʊnd ɪm; ənd ət 'lɑ:st ðə 'nɔ:θ 'wɪnd
'geɪv ,ʌp ðɪ ə'teɪt. ðen ðə 'sʌn 'ʃɒn 'ɔwt 'wɔ:mli, ənd
ɪ'mɪ:dʒətli ðə 'trævlə 'tɒk ɒf ɪz 'klɒk. ən 'səʊ ðə 'nɔ:θ
'wɪnd wəz ə'blaɪdʒd tə kən'fes ðæt ðə 'sʌn wəz ðə 'strɒŋgə
əv ðə ,tu:.

English: RP (2)

ðə'nɒθ ,wɪnd ənðə'sʌn / wədi'spju:tɪŋ / 'wɪf wəz
ðə'strɒŋgə / wenə'trævlə ,keɪm ə'lɒŋ / 'ræpt ɪnə'wɒm
'klɒk / ðeɪə'grɪd / ðætðə'wʌn hu'fəst sək,sɪdɪd ɪn,meɪkɪŋ
ðə,trevlə 'teɪk ɪz,klɒk 'ɔf / ʃədbɪkən,sɪdəd 'strɒŋgə
ðənðɪ,əðə / 'ðen / ðə'nɒθ ,wɪnd 'blu: əz'hɑd əzi'kʊd /
bətðə'mɔr ɪj'blu: / ðə,mɔ 'klɒsli dɪdðə,trevlə ,fəwɪd
ɪz,klɒk ə'rəʊndɪm / əndət'lɑst / ðə'nɒθ 'wɪnd / 'geɪvəp
ðə'teɪt / 'ðen / ðə'sʌn 'ʃɒn ,ɔwt 'wɒmli / əndɪ'mɪdʒətli /
ðə'trævlə 'tʊkɔf ɪz'klɒk / ən'səʊ / ðə'nɒθ 'wɪnd /
wəzə'blaɪdʒd təkən'fes / ðætðə'sʌn wəzðə'strɒŋgə əvðə,tu: /

Tablica 3.

RP PHONETIC ALPHABET (1)

i:	ɪ	ɛ	æ	ʌ	ɑ:	ɒ	ɔ:	
key	kid	bet	back	cut	mark	dock	lord	
ki:	kɪd	bɛt	bæk	kʌt	mɑ:k	dɒk	lɔ:d	
	ʊ	u:	ɜ:	ə				
	look	food	bird	ago				
	lʊk	fu:d	bɜ:d	ə'gəʊ				
eɪ	aɪ	ɔɪ	əʊ	ɑʊ	ɪə	ɛə	oə	
day	bite	boy	boat	foul	fear	fair	tour	
deɪ	baɪt	bɔɪ	bəʊt	fɑʊl	fɪə	fɛə	tʊə	
p	b	t	d	k	g	(?)	tʃ	dʒ
pay	bone	tea	door	cat	go	catch	chip	just
peɪ	bəʊn	ti:	dɔ:	kæt	gəʊ	kætʃ	tʃɪp	dʒʌst
f	v	θ	ð	s	z	ʃ	ʒ	h
fee	vine	thin	that	sea	zone	shut	rouge	hat
fi:	vaɪn	θɪn	ðæt	si:	zəʊn	ʃʌt	ru:ʒ	hæt
m	n	ŋ	l	r	j	w		
mud	nut	king	load	run	yes	win		
mʌd	nʌt	kɪŋ	ləʊd	rʌn	jes	wɪn		
m	n	l	r					
rhythm	button	bottle	honorary					
'rɪðm	'bʌtn	'bɒtl	'ɒnərɪ					

Tablica 4.

Russian

і	ı	ı	e	a	o	o
пить	идти	дым	нет	дар	дом	урок
'pit	'ı'ti	'dim	'net	'dar	'dom	u'rok
		u	ə	ʌ		
		ухо	рано	они		
		'uxə	'ranə	ʌ'ni		
р	ɹ	b	b	t	t̪	d
путь	петь	брат	бить	так	тихо	дама
'put̪	'pɛt̪	'brat	'bʲit̪	'tak	'tʲixə	'damə
д	k	k̪	g	f	f̪	v
день	как	руки	гора	ров	кровь	вода
'dɛn̪	'kak	ro'ki	gɔ'ra	'rof	'krof̪	vɔ'da
ѵ	s	ʂ	z	z̪	ʃ	ʒ
вес	сад	семь	зуб	зима	шаг	жить
'ves	'sat	'ʂɛm̪	'zup	zɪ'ma	'ʃak	'ʒit̪
х	t̪	tʃ	m	m̪	n	
хата	цвет	час	мать	мясо	ноч	
'xata	't̪vɛt	'tʃas	'mat̪	'mʲasə	'not̪	
п	l	l̪	r	r̪	j	
небо	луна	лес	рыба	ряд	еж	
'nɛbə	lɔ'na	'les	'riba	'rat̪	'joʃ	

Tablica 5.

Znaki transkrypcji kwazialofonicznej języka rosyjskiego z kluczem.

POL. IPA

C	1	2	3	Y	G	L	J	N
q	α	ᾱ	α̅	i	j	†	p	ŋ
S	4	5	6	Z	\$	%		
f	ts	tʃ	tɕ	3	z	?		

FRENCH. IPA

A	@	E	N	X
a	ə	ε	ŋ	ø
O	&	S	Y	Z
o	œ	f	u	3

RUSS. IPA

A	B	D	2	@	F	I	Y	K	L	M	
Λ	b	d	ᾱ	ə	f	l	i	k	l	m	
N	O	P	R	S	\$	T	5	4	V	%	Z
ŋ	o	ʀ	r	s	j	t	tʃ	ts	v	3	z

ITAL. IPA

1	2	E	L	N	O	S	4	5
ᾱ	α̅	ε	λ	ŋ	o	j	ts	tʃ

Tablica 6.1.]

Znaki transkrypcji fonematycznej dla języków: polskiego, francuskiego, rosyjskiego i włoskiego z kodem komputerowym,

ENGL. IPA

A	V	\$	&	D	2	@	E]	I	Y
a	ʌ	ɒ	æ	ɔ	ɔʃ	ə	e	ə	l	i
N	U	O	S	T	5	Z	%			
ŋ	o	ɔ	f	θ	ʃ	ʒ	ʔ			

GER. IPA

A	&	C	2	@	E	I	N	X	U
a	e	ç	ɔʃ	ə	e	l	ŋ	ø	ω
O	\$	S	4	[Z	%			
o	œ	f	4	[ʒ	ʔ			

SWED. IPA

A	C	D	@	E	H	I	N	G	
a	ç	d	ə	ɛ	h	l	ŋ	ŋ	
X	U	%	O	\$	S	T	Y	Z	!
ø	ω	e	ɔ	ø	s	t	v	z	*

Tablica 6.2.

Znaki transkrypcji fonetycznej dla języków: angielskiego, niemieckiego i szwedzkiego z kodem komputerowym.

a ʌ e ɒ æ b, ɸ ɣ
d ð ɔ ɔz ɔʒ ɔʒ ə ɛ
z f ʃ l i j k, l
ʎ t l m, n ɳ ɲ ɳ
ø ω ɵ ɔ æ œ ɹ ɣ
s ʂ ʃ t t ɵ ts tʃ
tʃ ɥ ʏ w ʋ ɥ ʏ ʒ
z z ʒ ʔ ʔ ʔ x

a ʌ e ɒ æ b, ɸ ɣ
d ð ɔ ɔz ɔʒ ɔʒ ə ɛ
z f ʃ l i j k, l
ʎ t l m, n ɳ ɲ ɳ
ø ω ɵ ɔ æ œ ɹ ɣ
s ʂ ʃ t t ɵ ts tʃ
tʃ ɥ ʏ w ʋ ɥ ʏ ʒ
z z ʒ ʔ ʔ ʔ x

Tablica 7.
Specjalne litery transkrypcyjne powiększone w antykwie
i kursywie.

¹ W przypadku dzieci przedszkolnych, analfabetów oraz tych rzadkich już ludów, które nie posiadają pisma, w pamięci językowej przechowywane są oczywiście jedynie wzorce foniczne. W przypadkach szczególnych istnieją inne formy i związane z nimi kody, np. kod komunikacji migowej u ludzi głuchych.

² Grafem jest klasa liter określonego alfabetu, które są ekwiwalentne, jako że różnice między nimi są zależne jedynie od sposobu zapisu (ręczny, drukowany) lub kroju albo cech osobniczych pisma.

³ Jedynym z nielicznych wyjątków jest bardzo dawny artykuł D. Jonesa (1937).

⁴ Nie istniejące w rzeczywistości twory quasi-językowe oznacza się gwiazdka.

⁵ Tutaj i w dalszym ciągu posługujemy się przykładami w formie wyrazów. Jednakże wyraz nie jest jednostką fonetyczną. W każdym z przykładów implikujemy, że chodzi o całkowity makrosegment fonetyczny, tj. wyższego rzędu jednostkę fonologiczną, mianowicie fragment sygnału mowy zawarty między kolejnymi przerwami (pauzami, czyli odpowiednio długimi (ponad 500 ms) segmentami zerowymi. Makrosegmentem może być oddzielnie wymówiony wyraz.

BIBLIOGRAFIA

- Albright, R.W. (1958) *The International Phonetic Alphabet. Its Backgrounds and Development*, Indiana University Research Center in Anthropology, Folklore and Linguistics, vol.24 No 1.
- Beyer, F. (1929) *Frazoesische Phonetik*, Quelle & Meyer, Leipzig.
- Cohen, J.R. (1981) Segmenting Speech Using Dynamic Programming, *JASA* vol.69, 1430-1444.
- Dixon, N.R. & Martin, T.B., eds. (1978) *Automatic Speech and Speaker Recognition*, IEEE Press, New York.
- Domagała, P. (1984) *Automatyzacja procesu segmentacji sygnału mowy w układzie analogowo-cyfrowym*, Prace IPPT 5/1984, Warszawa.
- Elert C.C. (1964) *Phonological Studies of Quantity in Swedish*, Stockholms Kommunal forvaltning, Uppsala.
- Fant, B. & Lindblom, B. (1961) Studies in Minimal Speech Sound Units, *STL QPSR*, 2/1961, 1-11.
- Fowler, C.A. (1983) Realism and Unrealism: A Reply, *Journ. Phon.* vol. 11, 303-322.
- Frackowiak-Richter, L. (1973) The Duration of Polish Vowels, *Speech Analysis and Synthesis*, vol.3 (W. Jassem, ed.) Państwowe Wyd. Nauk., Warszawa, 87-115.
- Fry, D.B. (1950) Experiments in Mechanical Speech Recognition *Information Theory* (C. Cherry, ed.) Academic Press, New York, 205-212.
- Gimson, A.C. (1962) *An Introduction to the Pronunciation of English*, Arnold Publ. Ltd., London.
- Hammarberg, R. (1976) The Metaphysics of Coarticulation, *Journ. Phon.* vol.11, 353-115.
- (1982) On Defining Coarticulation, *Journ. Phon* vol. 4, 353-363.
- Haton, J.P. (1982) ed. *Automatic Speech Analysis and Recognition*, Reidel Publ. Co. Dordrecht.
- Henton, C.G. (1987) The IPA Consonant Chart: Mugwumps, Holes and Therapeutic Suggestions, *JIPA* vol. 17 No1, 15-25.

- Jassem, W. (1949) Indication of Rhythm in the Transcription of Educated Southern English, *Le Maître Phonétique* vol. III No. 92, 22-24.
- (1950) Phonemic Transcription of the Vowel Sounds of Educated Southern English, *Le Maître Phonétique* vol. III No 93, 10-12.
- (1978) On the Distributional Analysis of Pitch Phenomena *Language and Speech*, vol. 21 No. 4, 362-372.
- (1984) Automatic Segmentation of the Speech Signal into Phone-Length Elements, *Proc. of the 10th Intern. Congr. Phon. Sci.*, Foris Publ. Dordrecht, 318-321.
- (1986) Vowel Formant Frequencies as Linguistic and Speaker-Specific Features of the Speech Signal, *Language Global Perspectives* (B.F. Elson, ed), The Summer Institute of Linguistics, Dallas, Texas, 303-312.
- (1987) *The Phonology of Modern English*, PWN, Warszawa.
- Jassem, W., Demenko, G. (1986) Extracting Linguistic Information from F_0 Traces, *Intonation in Discourse* (C. John-Lewis, ed) Croom Helm, London, 1-17.
- Jassem, W., Domagała P. (1986) Phonetic Segmentation in Bottom-Up Automatic Speech Analysis, *Intern. Conference on Speech Input/Output; Techniques and Applications*, London, 59-62.
- Jassem, W., Hill, D.R. and Witten I.H. (1984) Isochrony in English Speech: Its Statistical Validity and Linguistic Relevance: *Intonation, Accent and Rhythm* (D. Gibbon and H. Richter, eds.), W de Gruyter, Berlin, 203-225.
- Jassem, W., Krzyśko, M., Stolarski, P. (1984) *Częstotliwości formantowe jako cechy fonetyczne i osobnicze w świetle statystycznej analizy dyskryminacyjnej*, Prace IPPT 27/1984, Warszawa.
- Jassem, W., Kubzdela H., Domagała P. (1983) Automatic Acoustic-Phonetic Segmentation of the Speech Signal, *From Sounds to Words* (R. Hedquist, ed.), Almqvist and Wiksell, Umeå, 159-167.
- Jones, D. (1938) Concrete and Abstract Sounds, *Proceedings of the 3rd Intern. Congr. Phon. Sci.*, Ghent
- Jones, D. (1950) *The Phoneme: Its Nature and Use*, Heffner &

Sons, Cambridge.

- Karaś, M., Madejowa, M. (1984) *Słownik wymowy polskiej* PWN Warszawa
- Labov, W. (1986) Sources of Inherent Variation in the Speech Process, *Invariance and Variability in Speech Processes* (J.S. Perkell & D.H. Klatt, eds.) Erlbaum Ass. Publ., Hillside, N.J. 402-423.
- Ladefoged, P. (1975) *A Course in Phonetics*, Harcourt, Brace Jovanovich, Incl., New York.
- (1987a) Macintosh computer programs for phoneticians, *Working Papers in Phonetics* Univ. of California, Los Angeles, No 66. 98-103.
- (1987b) Updating the Theory, *JIPA* vol.17 No.1, 10-14.
- Lea, W.A. ed. (1980) *Trends in Speech Recognition*, Prentice Hall, Inc., Englewood Cliffs, N.J.
- Lobanov, B.M. (1971) Classification of Russian Vowels Spoken by Different Speakers, *JASA* vol. 49, 606-608.
- Lobacz, . (1971) *Wpływ tempa mowy na przebiegi formantów samogłosek polskich*, Prace IPPT 67/1971, Warszawa.
- (1973) Non-Unique Phonemic Interpretation of the Polish Speech Sounds, *Speech Analysis and Synthesis* vol.3 (W. Jassem, ed.) PWN Warszawa 75-83.
- (1985) *Processing and Decoding the Signal in Speech Perception*, *Hamburger Phonetische Beitrage*, B. 44. H. Buske Verl., Hamburg, 95-101.
- Otten, K.W. (1964) *Simulation and Evaluation of Phonetic Speech Recognition Techniques*, vol. 1: Segmentation of Continuous Speech into Phonemes, Technical Report No. RTD-TOR-1964. 63-4005 Avionics Laboratory, Ohio.
- Papcun, G. (1980) *How do Different Speakers Say the Same Vowels?* UCLA Working Papers in Phonetics No.8.
- Peterson, G.E. (1951) The Phonetic Value of Vowels, *Language* vol. 27, 541-553.
- Peterson, G.E. & Barney H.L. (1952) Control Methods Used in a Study of the Vowels, *JASA* vol. 24, 175-184.
- Peterson, G.E. & Lehiste, I. (1960) *Duration of Syllabic Nuclei in English*, 2. The Univ. of Michigan Speech Research Laboratory Report No 4, Ann Arbor, Michigan.

- Pols, L.C. (1977) *Spectral Analysis and Identification of Vowels in Monosyllabic Words*, Akademische Pers. B.V., Amsterdam.
- The Principles of the International Phonetic Association* (1949) IPA London.
- Richter, L. (1987) Modelling of the Rhythmic Structure of Utterances in Polish, *Studia Phonetica Posnaniensia* vol. 1. 91-125.
- Roach, P. (1983) *English Phonetics and Phonology. A Practical Course*. Cambridge Univ. Press, Cambridge.
- Słownik wymowy polskiej (red. M. Karaś i M. Madejowa) (1985) PWN Warszawa.
- Steffen-Batogowa M. (1975) *Automatyzacja transkrypcji fonematycznej tekstów polskich*, PWN, Warszawa.
- Stieber, Z. (1960) *Historyczna i współczesna fonetyka języka polskiego*, PWN, Warszawa.
- Summerfield, A.C. & Haggard, M.P. (1975) Vocal Tract Normalization as Demonstrated by Reaction Times, *Auditory Analysis and Perception of Speech* (G.Fant & M.A.A. Tatham, eds.) Academic Press, London. 115-141.
- Surmanowicz-Demenko, G. (1987) Mathematical Aspects of the Classification of Basic Pitch Patterns, *European Conference on Speech Technology* vol.1. Edinburgh 301-304.
- Trubetzkoj, N.S. (1949) *Principles de phonologie*, Librairie C.Klincksieck, Paris.
- Thudicum, G. (1926) *Manuel pratique de diction française a l'usage des entrangers*, Librairie Kundig, Geneve.
- Wells, J.C. (1987) *Computer Coded Phonetic Transcription* (Report) London 1987.



56751