

Janusz Imiolczyk

ZNACZENIE SKŁADU FONOTYCZNEGO  
PRÓBEK MOWY  
W SŁUCHOWEJ IDENTYFIKACJI GŁOSÓW

45/1985

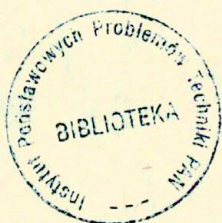
P. 269



WARSZAWA 1985

ISSN 0208-5658

Praca wpłynęła do Redakcji dnia 13 września 1985 r.



56904



Na prawach rękopisu

---

Instytut Podstawowych Problemów Techniki PAN

Nakład 140 egz. Ark.wyd. 0,71 Ark.druk. 1  
Oddano do drukarni w styczniu 1986 r.

---

Warszawska Drukarnia Naukowa, Warszawa,  
ul. Śniadeckich 8  
Zam. 70/86.

Janusz Imiołczyk

Pracownia Fonetyki Akustycznej

IPPT PAN

ZNACZENIE SKŁADU FONETYCZNEGO PRÓBEK MOWY  
W SŁUCHOWEJ IDENTYFIKACJI GŁOSÓW<sup>1)</sup>

Streszczenie

W pracy podjęto jeden z aspektów percepcyjno-słuchowego rozpoznawania głosów, który dotychczas tylko w niewielkim stopniu był przedmiotem zainteresowania eksperymentatorów. Bezpośrednim jej celem było ustalenie, czy poprawność identyfikacji może zależeć od zawartości głoskowej analizowanej próbki mowy, tzn. czy poszczególne głoski niosą taką samą ilość informacji o specyficznych, osobniczych cechach głosu. W omawianym doświadczeniu przedmiotem porównania były samogłoski, spółgłoski /l/ i /r/ oraz spółgłoski nosowe /m, n, ŋ/.

Niezależnie od części doświadczalnej, praca zawiera również krótki przegląd ważniejszych publikacji z dziedziny rozpoznawania głosów.

1. Wstęp.

Każda, nawet pusta znaczeniowo wypowiedź słowna jest nośnikiem informacji o tożsamości osoby mówiącej. Charakterystyczny sposób mówienia, barwa głosu i szereg innych czynników umożliwiają słuchającemu identyfikację osoby, której głos jest mu znany, w sytuacji, gdy osoba ta pozostaje poza polem jego widzenia. Co więcej, potrafi on także rozróżnić (np. przysłuchując się audycji radiowej) głosy kilku osób, których nigdy wcześniej nie słyszał.

Ogół cech decydujących o indywidualnym brzmieniu głosu można podzielić na dwa zasadnicze typy<sup>2)</sup> (por. np. [9], [29], [1]):

1) cechy wrodzone, zależne od budowy anatomicznej toru

1) Praca wykonana w ramach problemu węzłowego O6.9.

2) Mimo zastrzeżeń wysuniętych przez Nolana ([20], str. 19 i dalsze) podział ten można w zasadzie uznać za wystarczający

artykulacyjnego (np. naturalna wysokość głosu)  
oraz

- 2) cechy nabyte (wyuczone), będące efektem właściwych danej jednostce przyzwyczajzeń artykulacyjnych, wykształconych w procesie opanowywania mowy (np. naturalne tempo mówienia, typowe ułożenie ruchomych artykulatorów przy wymawianiu poszczególnych głosek itp.).

W sytuacji idealnej zespół cech pierwszego i drugiego typu stanowiłby o charakterystycznym, niezmiennym brzmieniu głosu danej osoby i - przy założeniu, że głosy są niepowtarzalne w tym samym stopniu, co odciski palców - mógłby służyć jako podstawa do odróżnienia go od wszelkich innych głosów. W rzeczywistości tak oczywiście nie jest. Sprawę komplikuje nie tylko fakt, iż głosy mogą być - zarówno w ocenie słuchowej, jak i pod względem cech widmowych - mniej lub bardziej do siebie podobne, ale również stosunkowo znaczna zmienność wewnątrzosobnicza (ang. intratalker variation), spowodowana przez szereg czynników natury anatomicznej (przypadki zmian patologicznych), fizjologicznej (np. przeziębienie, zob. [22]) czy psychologicznej (np. stres, zob. [7], [12]), bądź też wynikająca ze świadomego postępowania "właściciela" głosu (np. celowa zmiana naturalnego brzmienia lub imitacja głosu innej osoby, (zob. np. [7], [8])). Porównanie dokonanych w różnych terminach nagrań tego samego głosu wykazało także, iż zmienność wewnątrzosobnicza wzrasta w miarę wydłużania czasu dzielącego poszczególne nagrania (zob. np. [10], [26]).

Metody rozpoznawania głosów dzieli się tradycyjnie (por. np. [26]) na subiektywne, obejmujące (a) badania słuchowe i (b) analizę (wzrokową) spektrogramów próbek mowy różnych osób oraz "obiektywne", polegające na (a) półautomatycznym bądź (b) automatycznym porównywaniu cech widmowych poszczególnych głosów. Metody zwane obiektywnymi są w istocie jedynie zobiektywizowanymi metodami subiektywnymi, jako że i one nie obywiają się bez ingerencji człowieka-eksperta.

Badania nad rozpoznawaniem głosów zapoczątkowała picnierska praca Mc Gehee [18] z 1937 roku. Szczególnie żywe zainteresowanie tym zagadnieniem zaznaczyło się od połowy lat sześćdzie-

siątych w związku z ukazaniem się serii prac Kersty ( m. in. [14] i [15] ) oraz dopuszczeniem nagrań mowy jako dowodów pomocniczych w sądach amerykańskich. W ostrej polemice, jaka wywiązała się w latach sześćdziesiątych i siedemdziesiątych (zob. np. [2] i [16] ) poddawano głównie krytyce metodę porównywania spektrogramów, zarzucając jej przede wszystkim, że ze względu na niewystarczająco wysoką wiarygodność wyników nie może ona stanowić podstawy do orzekania o czyjejs winie bądź niewinności. Zwracano uwagę na brak analogii między odciskami palców a "odciskami" głosów (od ang. "voiceprint", terminu wprowadzonego przez Kerstę), które charakteryzuje duża zmienność, będąca efektem wpływu szeregu czynników fizjologicznych, psychologicznych, akustycznych i in. na brzmienie mowy.

W pracach powstałych po tym okresie, a dotyczących głównie półautomatycznego i automatycznego rozpoznawania głosów, położono nacisk na znalezienie jasnych, bardziej efektywnych kryteriów identyfikacji, a więc na ustalenie, jakie parametry akustyczne niosą najwięcej informacji o charakterystycznych, osobniczych cechach głosu. Poszukiwania te nie zostały jednak zakończone pełnym sukcesem: nie udało się znaleźć inwariantnych cech<sup>1)</sup>, które umożliwiłyby skuteczne rozróżnianie głosów. Rozpoznawanie słuchowe pozostaje w związku z tym metodą nadal najbardziej efektywną i najpewniejszą (por. np. [3], [4], [24] i [30]).

## 2. Zasadnicze problemy słuchowego rozpoznawania głosów.

Jak zauważa Tosi [26], mimo występującej współcześnie dominacji metod automatycznych, badania nad percepcyjno-słuchowym rozpoznawaniem głosów są nadal warte prowadzenia (str.63). Wciąż jeszcze pozostaje niewyjaśnione, jakie informacje wykorzystuje słuchacz w celu zidentyfikowania głosu oraz jakie mechanizmy odgrywają zasadniczą rolę w stosowanych przy tym przez niego strategiach decyzyjnych.

Od czasu klasycznej pracy Mc Gehee [18] w doświadczeniach z zakresu percepcyjnego rozpoznawania głosów starano się

1) Por. wymagania stawiane takim cechom w: Nolan [20], str.8

głównie uzyskać odpowiedź na pytanie, jak dobrze (tzn z jakim procentem poprawności) człowiek jest w stanie rozróżniać bądź identyfikować głosy w różnych typach warunków. Badano m. in. stopień poprawności rozpoznawania głosów:

- znanych bądź nie znanych słuchaczom
- przetworzonych akustycznie (np. poprzez filtrowanie w różnych zakresach częstotliwości czy maskowanie szumem)
- nagranych w różnych odstępach czasu
- o celowo zmienionym brzmieniu (w tym także imitacji głosów znanych osobistości przez profesjonalnych naśladowców)
- w warunkach stresu czy
- w mowie szeptanej.

W innych pracach (zob. np. [13] i [28]) podjęto próbę określenia istotnych atrybutów percepcyjnych głosów w celu opracowania odpowiednich skal klasyfikacyjnych, które stanowiłyby trwałą podstawę dla rozróżniania głosów. Zaproponowane w oparciu o wyniki doświadczeń "parametry" percepcyjne (np. "wysokość", "ożywienie", "czystość") okazały się jednak niewystarczające: przykładowo, w pracy Clarke'a i Beckera [4] procent poprawnych odpowiedzi wynosił zaledwie ok. 68 (dla porównania: w bezpośrednim teście odsłuchowym uzyskano 90 % prawidłowych rozpoznań).

Stosunkowo niewiele prób podjęto w celu ustalenia, jakie źródła informacji osobniczych są najistotniejsze z punktu widzenia oceny słuchowej. Skoncentrowano się w nich głównie na najbardziej ogólnych aspektach rozpoznawania. Najczęściej podejmowanym zagadnieniem był wpływ częstotliwości podstawowej ( $F_0$ ) oraz funkcji przenoszenia toru głosowego na uzyskiwane w testach odsłuchowych rezultaty. Mimo swojej niezaprzeczalnie istotnej roli (zob. np. [17]),  $F_0$  okazała się być mniej zasobnym nośnikiem informacji osobniczych niż drugi z wymienionych czynników (por. [5], [19] i [23]). Przykładowo, wyniki eksperymentu Colemana ([5]) wskazują na to, że źródło tonu krtaniowego dostarcza nie więcej niż ok. 10 % ogółu informacji wykorzystywanych do identyfikacji głosu.

Nawet ze skąpych danych istniejących prac eksperymentalnych wynika, iż stopień poprawności rozpoznawania głosów

przez słuchaczy zależy od typu użytego w nagraniach materiału językowego. Bricker i Pruzansky [3] zauważyli m. in., że pewną rolę w tym zakresie odgrywa np. rodzaj samogłoski. Konkluzję tę potwierdziły rezultaty uzyskane przez Stevensa i in. [24], z których wynika, iż akcentowane samogłoski przede wszystkim zawierają więcej informacji osobniczych niż akcentowane samogłoski tylne<sup>1)</sup>. Problem istotności składu fonetycznego materiałów testowych nie doczekał się jednak szerszej zakrojonego opracowania.

### 3. Słuchowe rozpoznawanie głosek. Opis doświadczenia.

#### 3.1. Materiał.

Celem eksperymentu było stwierdzenie, czy - i jak dalece - poprawność rozpoznawania głosek zależy od zawartości fonetycznej materiału odsłuchowego.

Lista użytych w doświadczeniu bodźców obejmowała 27 następujących logatomów, utworzonych ze wszystkich dopuszczalnych fonotaktycznie połączeń (typu CV) spółgłosek /l, r, m, n, ŋ/ z samogłoskami (/i, ɪ, e, a, o, u/):

lilili, lelele, lalala, lololo, lululu,  
ririri, ryriry, rerere, rarara, rororo, rururu,  
mimimi, mymymy, mememe, mamama, momomo, mumumu,  
nynyny, nenene, nanana, nonono, nununu,  
ninini, nienienie, nianiania, niñionio, niuniuniu.

Logatomom tym nadano postać trzysylabową ze względu na fakt, iż założony czas trwania każdego z nich miał wynosić ok. 1 s. Z obserwacji Pollacka i in. [21], potwierdzonych przez Comptoną [6] wynika, że przedłużanie próbki ponad 1 s nie wpływa istotnie na podwyższenie rezultatów rozpoznawania.

Przy wyborze spółgłoskowych "składników" logatomów kierowano się danymi prac, w których podstawę dla identyfikacji głosek stanowiły parametry widmowe. W świetle tych danych (np. [9], [22] i [25]) spółgłoski nosowe okazują się być szczególnie istotnym nośnikiem informacji osobniczych, /l/ i /r/ natomiast zawierają tych informacji znacznie mniej ([20])

<sup>1)</sup> Odmianą tendencję zaobserwował La Riviere [17], w którego doświadczeniu rozpoznawanie było skuteczniejsze w przypadku spółgłosek otwartych

### 3.2. Nagrania.

Nagrania przeprowadzono w komorze bezdechowej przy wykorzystaniu 8 głósów męskich o zbliżonej częstotliwości podstawowej<sup>1)</sup>. Poszczególne osoby dokonywały nagrań w ramach jednej sesji. Ich zadanie polegało na dwukrotnym powtórzeniu każdego logatomu za słyszaniem w słuchawce głosem "wzorcowym" (zabieg ten miał na celu wyeliminowanie ewentualnych różnic w zakresie iloczasu i przebiegu intonacji). Wszystkie logatomy wymawiane były z intonacją pełną rosnącą i akcentowane na drugiej sylabie. Szczególną uwagę zwrócono na to, aby poziom nagrań był jednakowy. Nagrania poprzedzano każdorazowo szczegółową instrukcją i krótkim treningiem.

### 3.3. Zasady montażu taśmy odsłuchowej.

Na taśmie magnetofonowej przeznaczonej do odsłuchów występowało bezpośrednio po sobie pięć wymówień każdego logatomu, przy czym głos pierwszy w kolejności traktowany był jako "nieznany" i - w ramach poszczególnych piątek - powtarzał się na pozycji trzeciej, czwartej lub piątej. We wszystkich przypadkach na pozycji drugiej występował ten sam, "nie liczący się" głos, którego wmontowanie miało na celu podwyższenie stopnia trudności rozpoznawania.

Odstęp między głosem 1 i 2 wynosił 4 s, zaś między pozostałymi głosami (tj. między 2 i 3, 3 i 4 oraz 4 i 5) - 2 s. Kolejne piątki oddzielała od siebie 5-sekundowa przerwa.

Zasady montażu taśmy ilustruje poniższa tabela:

Lp.	logatom	głos 1	głos 2	głos 3	głos 4	głos 5
1	lilili	<u>AB</u>	CD	EF	<u>AB</u>	GH
2	nianiania	<u>IJ</u>	CD	KL	<u>MN</u>	<u>IJ</u>

Poszczególne logatomy występowały na taśmie w kolejności losowej. Osiem pozostających do dyspozycji głósów przypisano bodźcom w taki sposób, aby:

<sup>1)</sup> Wykorzystano głosy 8 członków Chóru Akademickiego w Poznaniu, reprezentujących bas



- a) dany głos występował na pozycji pierwszej nie częściej niż 4 razy, zaś na pozostałych pozycjach łącznie nie częściej niż 7 razy
- b) odległość między wystąpieniami tego samego głosu na pozycji pierwszej była nie mniejsza niż 16 wymówień
- c) odległość między wystąpieniami tego samego głosu na dowolnej pozycji (poza pierwszą) była nie mniejsza niż 7 wymówień
- d) dany głos nie występował dwukrotnie na pozycji pierwszej w logatomach zaczynających się na tę samą spółgłoskę.

Przy sporządzaniu nagrania wykorzystano tylko jedno wymówienie każdego logatomu przez dany głos, a mianowicie to, które było lepszą imitacją (w zakresie iloczasu i przebiegu intonacji) odpowiedniego wymówienia "wzorcowego".

#### 3.4. Badania odsłuchowe.

W badaniach odsłuchowych wzięło udział 25 osób. Zadanie ich polegało na ustaleniu, który z głosów występujących - w obrębie danej piątki - na pozycji trzeciej, czwartej lub piątej jest identyczny z głosem 1-szym. Respondenci byli informowani, iż głosu drugiego nie należy brać pod uwagę.

W związku z tym, że test miał charakter zamknięty (głos 1 zawsze powtarzał się na którejś z trzech końcowych pozycji), jedynym typem błędu, jaki mógł wystąpić była niepoprawna identyfikacja.

#### 3.5. Wyniki.

Na ogólną liczbę 675 odpowiedzi zanotowano 159 odpowiedzi błędnych (ok. 24 %), co daje ok. 6,4 błędów na jednego słuchacza. Liczby niepoprawnych rozpoznań w odniesieniu do poszczególnych logatomów podane są w Tabeli 1. Dowodzą one, że poprawność rozpoznawania była niższa w przypadku logatomów zawierających spółgłoski nosowe. Maksymalna liczba błędnych rozpoznań wynosiła dla lVlVlV i rVrVrV zaledwie 6, podczas gdy dla logatomów z nosowymi aż 16. Bodźce pierwszego typu stanowiły najwidoczniej lepszy materiał do identyfikacji głosów. Ilustruje to Tabela 2. Mimo iż różnice w liczbie błędnych rozpoznań wydają się być dla poszczególnych

grup logatomów dość znaczne, nie osiągają one poziomu istotności statystycznej: wartość krytyczna testu sumy kolejności  $H$  (Kruskala-Wallisa) jest wyższa od uzyskanej

$$\chi^2_{0,05} = 9,49 > \chi^2 = 1,64 \quad 1).$$

Tabela 1

ranga	liczba błędów	logatom	ranga	liczba błędów	logatom
1	0	mymymy	14	5	rururu
2	1	rarara		5	ryryry
3	2	mememe		5	rororo
	2	nanana		5	momomo
	2	nununu	18	6	lololo
	2	niuniuniu		6	rerere
	2	ririri	20	7	nenene
8	3	nionionio		7	nianiania
	3	lelele	22	8	mumumu
	3	lalala	23	12	mimimi
	3	lululu	24	14	nienienie
12	4	nonono	25	16	mamama
	4	lilili		16	nynyny
				16	ninini

Tabela 2

Lp.	typ logatomu	liczba błędów	średnia błędów na 1 logatom
1	lVlVlV	19	3,8
2	rVrVrV	24	4
3	nVnVnV	31	6,2
4	mVmVmV	43	7,2
5	V V V	42	8,4

1) Różnice te okazały się nieistotne statystycznie także w przypadku zastosowania jednoczynnikowej analizy wariancji ( $F_{0,05} = 5,79 > F = 1,09$ ) oraz testu  $\chi^2$  (dla rozkładów lVlVlV

i nVnVnV:  $\chi^2_{0,05} = 9,49 > \chi^2 = 6,42$ )

Wśród samogłosek najwięcej informacji osobniczych zawierają samogłoski tylne (Tabela 3), jednakże i w tym przypadku różnice w liczbie niepoprawnych identyfikacji okazały się nieistotne ( $F_{0,05} = 4,55 > F = 2,07$ ).

Tabela 3

Lp.	typ logatomu	liczba błędów	średnia błędów na 1 logatom
1	CuCuCu	20	4
2	CoCoCo	23	4,6
3	CaCaCa	29	5,8
4	CeCeCe	32	6,4
5	CyCyCy	21	7
6	CiCiCi	34	8,5

Czynnikami mającym pewien wpływ na wyniki rozpoznawania była pozycja, na której powtarzał się głos 1. Niewątpliwie zadanie respondentów było najłatwiejsze, jeżeli głos 1 był identyczny z głosem 3 (por. np. [4], str.754). Rzeczywiście, uzyskane wyniki były w tym przypadku lepsze, przy czym różnice okazały się być istotne statystycznie:  $\chi^2_{0,05} = 5,99 < \chi^2 = 9,4$ . Należy więc wnioskować, iż poprawność identyfikacji tych głosów, które w teście odsłuchowym powtórzyły się na pozycji trzeciej byłaby niższa, gdyby wystąpiły one na pozycji czwartej lub piątej. Tym samym, liczba błędów dla niektórych logatomów uległaby podwyższeniu. Logatomy te to:

mymymy (0 błędów), rarara (1 bł.), nanana (2 bł.), niuniuniu (2 bł.), lelele (3 bł.) i nionionio (3 bł.).

Jak widać, w grupie tej występują cztery logatomy zawierające spółgłoski nosowe (w tym dwa zawierające /ŋ/). Gdyby poprawność identyfikacji była w ich przypadku niższa, różnice w liczbie błędnych odpowiedzi dla logatomów typu lVlVlV, mVmVmV itd. mogłyby okazać się istotne statystycznie.

Spośród ośmiu głosów wykorzystanych w nagraniach dwa (WZ i AW) okazały się łatwiejsze do rozpoznania niż pozostałe, przy czym rozpoznanie jednego głosu (AG) stanowiło najwyraźniej dla respondentów zasadniczy problem.

Poszczególne głosy zostały rozpoznane z następującą poprawnością:

- WZ - 93 %, AW - 91 %, SI - 80 %, AS - 78 %, PO - 72 %, FG - 70 %, RM - 69 % i AG - 61 %.

Analiza uzyskanych wyników przy zastosowaniu metody skalowania wielowymiarowego wykazała, iż do skutecznego rozróżnienia poszczególnych głosów niezbędnych jest co najmniej 5 wymiarów. Najbardziej podobne do siebie okazały się być głosy AW i PO, AG i AW oraz FG i PO, zaś najmniej podobne - głosy RM i SI, AW i AS, FG i WZ oraz PO i WZ. Interesujący jest przy tym fakt, iż ta relacja podobieństwa nie jest bynajmniej zwrotna. O ile, przykładowo, głos AG został rozpoznany jako AW 15 razy o tyle substytucja odwrotna nie wystąpiła ani razu. Analogiczny przypadek zanotowano w odniesieniu do głosów FG i PO oraz PO i AW.

Ciekawa jest jeszcze jedna obserwacja: mimo względnie wysokiej rozróżnialności (91 % poprawnych identyfikacji) głos AW był często wykorzystywany w błędnych odpowiedziach (łącznie: 39 razy). Z trudnych do ustalenia powodów głos ten był odbierany jako podobny do pozostałych (z wyjątkiem AS i SI), mimo że sam stanowił "obiekt" stosunkowo łatwy do rozpoznania.

### 3.6. Podsumowanie.

Uzyskany w niniejszym eksperymencie procent poprawnych odpowiedzi (ok. 76,4) nie da się bezpośrednio porównać z rezultatami innych prac ze względu na rozbieżności występujące w zakresie metodologii i użytego materiału językowego. Niewątpliwie, na wynik ten wpłynęło szereg czynników powszechnie uznawanych za istotne w tego typu doświadczeniach. Przykładowo, stopień poprawności rozpoznawania zależał w pewnej mierze od "odległości" jaka dzieliła - w ramach tego samego logatomu - dwa kolejne wystąpienia głosu 1 (tj. głosu identyfikowanego): stopień ten ulegał obniżeniu wraz ze wzrostem liczby dzielących oba te wystąpienia nieprawidłowych alternatyw (por. [4] i [11]). Zaznaczyło się także typowe zróżnicowanie rezultatów uzyskanych przez poszczególnych

respondentów (por. [11] i [26]): w przypadku najzdolniejszych słuchaczy błędne rozpoznania stanowiły zaledwie 8 % ogółu odpowiedzi (2 błędy), zaś w przypadku najmniej zdolnych - aż 44 % (11 błędów)<sup>1)</sup>.

Jak wspomniano na str. 8, przy doborze głosów dla celów niniejszego doświadczenia położono nacisk na homogeniczność grupy pod względem typowej częstotliwości podstawowej głosu (trudno byłoby oczywiście dokonać tego typu "normalizacji" w odniesieniu do funkcji przenoszenia toru głosowego poszczególnych mówców). Mimo to w ocenie słuchowej głosu różniły się oczywiście między sobą, przy czym - co typowe (por. np. [24]) - niektóre z nich były łatwiejsze (WZ), niektóre zaś trudniejsze do rozpoznania (AG).

Nie notowane w literaturze jest natomiast zjawisko polegające na wykorzystywaniu stosunkowo dobrze rozróżnialnych głosów (szczególnie uderzające jest to w przypadku głosu AW) jako uniwersalnych błędnych odpowiedzi. W oparciu o niniejszy materiał trudno jest ustalić, jakie czynniki spowodowały niezwrótność owej relacji podobieństwa.

Główną kwestię, którą starano się rozstrzygnąć w opisanym wyżej doświadczeniu stanowił wpływ składu fonetycznego próbek mowy na wyniki identyfikacji słuchowej głosów. W związku z niemożnością całkowitego wyeliminowania niektórych czynników interferujących, takich, jak pozycja, na której powtarzał się głos 1 czy nie w pełni zadowalająca homogeniczność grupy mówców, odpowiedź na to pytanie nie może mieć charakteru kategoriycznego. Ze względu na brak adekwatnych metod statystycznych, dokładne ustalenie jaki jest wpływ poszczególnych czynników oraz czy/jakie zachodzą między nimi interakcje nie jest w odniesieniu do użytego materiału możliwe. Z zaobserwowanych tendencji wynika (przy zastrzeżeniu, iż poziom porównywanych różnic nie osiąga istotności statystycznej), że (a) przeciwnie niż w rozpoznawaniu wzrokowym

1) Respondenci stanowili grupę dość jednorodną pod względem wieku (25 - 35 lat) i wszyscy uczestniczyli w tego typu eksperymencie po raz pierwszy

na podstawie analizy spektrogramów, w rozpoznawaniu słuchowym spółgłoski nosowe ustępują /l/ i /r/ pod względem zawartości (wykorzystywanych) informacji osobniczych oraz że (b) samogłoski tylne niosą tych informacji więcej niż przednie.

Właściwym komentarzem do pierwszej z powyższych obserwacji byłby wniosek wypływający z pracy Stevensa i in. [24], a mianowicie, iż materiał językowy stanowiący dobrą podstawę dla identyfikacji słuchowej niekoniecznie stanowi dobrą podstawę dla identyfikacji wzrokowej. Ponadto, w przypadku posłużenia się identycznym materiałem, przy ocenie tożsamości nietożsamości głosów "...w każdej z metod wykorzystujemy odmienne informacje..." (Clarke i Becker [4], str.760).

Wobec rozbieżności danych literaturowych na temat wpływu samogłosek na poprawność rozpoznawania głosów, konieczne wydaje się przeprowadzenie szerszej zakrojonych prac poświęconych tej kwestii. Konieczność taka występuje zresztą w odniesieniu do szeregu problemów związanych z rozpoznawaniem głosów. Mimo bowiem znacznego zainteresowania tym zagadnieniem w ostatnich latach, niektóre jego aspekty - zarówno w ujęciu akustycznym, jak i percepcyjnym - pozostają nadal nie wyjaśnione.

BIBLIOGRAFIA

- [1] ATAL, B.S., Automatic recognition of speakers from their voices, Proceedings of the IEEE 64, 1976, 460-75.
- [2] BOLT, R.H., COOPER, F.S., DAVID, E.E. Jr., DENES, P.B., PICKETT, J.M., STEVENS, K.N., Speaker identification by speech spectrograms: A scientists' view of its reliability for legal purposes, JASA 47, 1970, 597-612.
- [3] BRICKER, P.D., PRUZANSKY, S., Effects of stimulus content and duration on talker identification, JASA 40, 1966, 1441-49.
- [4] CLARKE, F.R., BECKER, R.W., Comparison of techniques for discrimination among talkers, Journal of Speech and Hearing Research 12, 1969, 747-61.
- [5] COLEMAN, R.O., Speaker identification in the absence of intersubject differences in glottal source characteristics, JASA 53, 1973, 1741-43.
- [6] COMPTON, A.J., Effects of filtering and vocal duration upon the identification of speakers aurally, JASA 35, 1963, 1748-52.
- [7] DOHERTY, E.T., HOLLIEN, H., Multiple-factor speaker identification of normal and distorted speech, Journal of Phonetics 6, 1978, 611-18.
- [8] ENDRES, W., BAMBACH, W., FLOSSER, G., Voice spectrograms as a function of age, voice disguise and voice imitation, JASA 49, 1971, 1842-48.
- [9] GLENN, J.W., KLEINER, N., Speaker identification based on nasal phonation, JASA 43, 1968, 368-72.
- [10] HARGREAVES, W.A., STARKWEATHER, J.A., Recognition of speaker identity, Language and Speech 6, 1963, 65-67.
- [11] HECKER, M.H.L., Speaker Recognition: An Interpretative Survey of the Literature, American Speech and Hearing Association Monograph 16, 1971.
- [12] HOLLIEN, H., MAJEWSKI, W., DOHERTY, E.T., Perceptual identification of voices under normal, stress and disguise speaking conditions, J. of Phonetics 10, 1982, 139-48.
- [13] HOLMGREN, G.L., Physical and psychological correlates of speaker recognition, J. Speech Hear. Res. 19, 1967, 57-66.
- [14] KERSTA, L.G., Voiceprint identification, Nature 196, 1962, 1253-57.

- [15] KERSTA, L.S., Voiceprint identification infallibility, JASA 34, 1962, 1978.
- [16] LADEFOGED, P., VANDERSLICE, R., The "voiceprint" mystique, UCLA Working Papers in Phonetics, 1967, 126-42.
- [17] LA RIVIERE, C.L., Contributions of fundamental frequency and formant frequencies to speaker identification, Phonetica 31, 1975, 185-97.
- [18] MC GEHEE, F., The reliability of the identification of the human voice, Journal of General Psychology 17, 1937, 249-71.
- [19] MILLER, J.E., Decapitation and recapitation: a study of voice quality, JASA 36, 1964, 2002 (A).
- [20] NOLAN, F.J.D., The Phonetic Bases of Speaker Recognition, PhD thesis, Cambridge University, 1980.
- [21] POLLACK, I., PICKETT, J.M., SUMBY, W.H., On the identification of speakers by voice, JASA 26, 1954, 403-12.
- [22] SAMBUR, H.R., Selection of acoustic features for speaker identification, IEEE Transactions on Acoustics, Speech and Signal Processing, ASSP-23, 1975, 176-82.
- [23] SHEARNE, J.N., HOLMES, J.N., An experiment concerning the recognition of voices, Language and Speech 2, 1959, 123-31.
- [24] STEVENS, E.N., WILLIAMS, C.E., CARBONELL, J.R., WOODS, B., Speaker authentication and identification: a comparison of spectrographic and auditory presentation of speech material, JASA 44, 1968, 1596-1607.
- [25] SU, L.-S., LI, K.-P., FU, K.S., Identification of speakers by use of nasal coarticulation, JASA 56, 1974, 1876-82.
- [26] TOSI, C., Voice Identification: Theory and Legal Applications, University Park Press, Baltimore 1979.
- [27] TOSI, C., OYER, E., LASHBROOK, W., PEDREY, C., NICCI, L., NASH, E., Experiment on voice identification, JASA 51, 1972, 2030-45.
- [28] VOIERS, W.D., Perceptual basis of speaker identity, JASA 36, 1964, 1065-73.
- [29] WOLF, J., Efficient acoustic parameters for speaker recognition, JASA 51, 1972, 2044-56.
- [30] YOUNG, K.A., CAMPBELL, R.A., Effects of context on talker identification, JASA 42, 1967, 1250-54.