

Daniel Śledziński

Uniwersytet im. Adama Mickiewicza w Poznaniu

AUTOMATYZACJA BADAŃ PERCEPCJI MOWY

SŁOWA KLUCZOWE: percepcja mowy, badania lingwistyczne, test percepcyjny

1. Wstęp

W artykule omówiono najważniejsze etapy badań percepcji mowy oraz przedstawiono koncepcję automatyzacji tych badań. Koncepcja ta jest realizowana w rozwiązaniu technicznym, które obejmuje trzy programy. Program pierwszy służy do projektowania testów percepcyjnych — dysponując zestawem plików dźwiękowych można szybko przygotować test. Program drugi jest przeznaczony do przeprowadzania testów. Zadaniem programu trzeciego jest analiza wyników. W czasie tworzenia niniejszego artykułu istniały prototypowe wersje dwóch programów oraz założenia dotyczące programu analitycznego.

2. Badania wcześniejsze i literatura

Idea oprogramowania ułatwiającego i przyspieszającego badania nad percepcją mowy powstała podczas wcześniejszych badań prowadzonych przez autora artykułu. Jedno z badań polegało na wygenerowaniu syntetycznych połączeń samogłosek oraz spółgłosek (miejsce łączenia pokrywało się z granicą ulokowaną między ośrodkiem sylaby otwartej a nagłosem następnej sylaby). Jakość takich syntetycznych połączeń była oceniana w teście percepcyjnym, a następnie te oceny były porównywane z ocenami analogicznych połączeń sylab, w których nie zastosowano żadnych modyfikacji.

Inne badanie prowadzone przez autora polegało na identyfikacji sylab typu CV (ang. *Consonant-Vowel*), które zostały wycięte z wyrazów nagranych w izolacji. Poza tym wykorzystano sztuczne sieci neuronowe, które były trenowane tak, aby spełniać rolę detektorów reagujących na określone sylaby (zastosowano te same nagrania sylab, które były użyte w teście percepcyjnym). Ostatecznie porównano wyniki identyfikacji sylab otrzymane z badań percepcyjnych oraz wyniki identyfikacji przeprowadzonej przez sztuczne sieci neuronowe.

Wczesne badania nad percepcją mowy dotyczyły wpływu określonych czynników występujących w sygnale mowy na percepcję. Jeden z bardziej znanych eksperymentów został przeprowadzony przez F. Coopera oraz A. Libermana w latach pięćdziesiątych dwudziestego wieku. Wykazał on, że możliwa jest identyfikacja syntetycznych sylab typu CV na podstawie tylko dwóch formantów samogłoskowych oraz ich zagieć (związanych ze zjawiskiem koartykulacji z poprzedzającą spółgłoską) [Cooper i in., 1952]. Percepcją samogłosek zajmowali się również W. Strange [Strange, 1989] oraz K. Stevens [Stevens i in., 1969]. Natomiast percepcją spółgłosek zajmował się między innymi J. M. Picket [Picket, 1980]. Wiele późniejszych badań nad percepcją mowy ukierunkowano na potrzeby techniki oraz medycyny. Na uwagę zasługują badania dotyczące afazji [Baum, 2001; Basso i in., 1977; Gow i in., 1996].

Istnieje wiele publikacji poświęconych percepcji mowy. Usystematyzowany przekrój aktualnych badań z tej dziedziny można odnaleźć w *The Handbook of Speech Perception* [Pisoni i Remez, 2008]. Jest to obszerna publikacja, która ujmuje tę tematykę w sposób bardzo wnikliwy — liczba omawianych eksperymentów i przytaczanych autorów jest imponująca.

Inną wartą polecenia publikacją jest podręcznik *Psycholingwistyka* [Gleason i Rathner 2009]), w którym rozdział trzeci poświęcono percepcji mowy. Przedstawiono tam wiele istotnych zagadnień, takich jak: percepcja samogłosek i spółgłosek, czas rozpoczęcia dźwięczności (ang. *Voice Onset Time* — VOT), wpływ tempa mowy na percepcję czy modele percepcji mowy.

Bardzo wartościową pozycją jest monografia pod redakcją Andrzeja Obrębowskiego *Wybrane zagadnienia z audiometrii mowy* [Obrękowski, 2011]. W publikacji tej przedstawiono między innymi problemy związane z metodologią badań, strukturą testów czy możliwościami wykorzystania audiometrii mowy w orzecznictwie lekarskim.

3. Etapy badań percepcji mowy

Badania percepcji audytywnej wymagają dużego nakładu pracy. Poniżej przedstawiono propozycję etapów pracy uwzględniających użycie omawianego w pracy oprogramowania:

- planowanie badań,
- przygotowanie materiału badawczego,
- przygotowanie testu percepcyjnego,
- przeprowadzenie testów percepcyjnych,
- analiza wyników oraz wnioski.

W dalszym ciągu artykułu omówiono każdy z wymienionych etapów oraz przedstawiono problemy, które można napotkać na tych etapach. Omówiono też rolę, jaką pełnią poszczególne programy w kolejnych etapach badań.

3.1. Planowanie badań

Podjęcie właściwych decyzji na tym etapie ma zasadniczy wpływ na powodzenie eksperymentu. Z kolei decyzje błędne mogą sprawić, że eksperyment będzie nieudany, a uzyskane wyniki nie będą miały istotnej wartości naukowej. Planując badania należy przeanalizować aktualny stan wiedzy z zakresu podejmowanego zagadnienia badawczego. Na obecnym etapie rozwoju nauki trudno wskazać na problemy zupełnie nowe. Najczęściej poszukuje się nowych aspektów problemów już wcześniej analizowanych. Poza tym w badaniach lingwistycznych często mamy do czynienia z sytuacją, w której określony problem był analizowany tylko dla jednego lub dla kilku języków. Dlatego ważna jest odpowiedź na pytanie, czy istnieją przesłanki uzasadniające podjęcie danego tematu badawczego dla języka, dla którego to zagadnienie nie było jeszcze analizowane. W większości przypadków odpowiedź na to pytanie będzie pozytywna — ze względu na różnice występujące pomiędzy językami — na przykład różnice na poziomach: leksykalnym, morfologicznym czy fonetycznym — z wyróżnieniem fonetyki artykulacyjnej, fonetyki akustycznej czy fonetyki audytywnej, której rola w omawianych zagadnieniach wydaje się być najbardziej istotna. Wymienione różnice sprawiają, że można odkrywać nieznane wcześniej aspekty określonych zjawisk — aspekty, które mogą uwidaczniać się w przypadku niektórych języków. Poza tym prowadzenie podobnych badań dla różnych języków umożliwia porównywanie tych języków pod kątem określonych zagadnień badawczych.

Kolejny problem, który trzeba koniecznie uwzględnić już na etapie planowania badań, jest związany z możliwościami technicznymi, jakimi dysponuje autor badań lub zespół badawczy. W niniejszym artykule omawiane jest rozwiązanie, które z założenia ma ułatwić prowadzenie badań percepcji audytywnej w szczególności osobom, które nie dysponują zaawansowanymi możliwościami i umiejętnościami technicznymi (głównie umiejętnością tworzenia programów komputerowych).

Prawdopodobnie najbardziej istotny problem techniczny w badaniach percepcji audytywnej związany jest z tym, w jaki sposób ma być przeprowadzony test percepcyjny — rozwiązanie, które pozwoli na odtwarzanie kolejnych nagrań oraz rejestrowane odpowiedzi. Przy najprostszym rozwiązaniu nie trzeba angażować zaawansowanych środków technicznych — wystarczy dysponować zbiorem plików dźwiękowych zapisanych w folderze i ułożonych w kolejności zgodnej z wykazem, do którego wgląd ma osoba badana. Takie rozwiązanie ma jednak wadę — nie jest ono wygodne dla osoby badanej, zatem mogą nastąpić trudności z koncentracją na merytorycznej stronie testu. Inne często stosowane rozwiązanie polega na utworzeniu specjalnej aplikacji przeznaczonej dla realizacji konkretnej koncepcji badawczej. Takie podejście umożliwia utworzenie wygodnego interfejsu dostosowanego do specyfiki określonego zadania badawczego, jednak wymaga ono zatrudnienia programisty. Celem omawianych w artykule narzędzi jest połączenie zalet dwóch wymienionych rozwiązań — łatwej dostępności oraz wysokiej jakości badań.

Na etapie projektowania badań trzeba też zastanowić się, czy realizacja danego zadania badawczego jest w ogóle możliwa. Problem ten dotyczy w szczególności badań nad percepcją mowy, ponieważ procesy zachodzące w ludzkim mózgu (w czasie percepcji) są bardzo słabo dostępne dla badacza. Uzyskane wyniki można porównywać z dostępnymi teoriami dotyczącymi percepcji mowy [Elman i in., 1984; Massaro, 1989].

Mówiąc o możliwościach realizacji zadań badawczych związanych z percepcją mowy należy pamiętać o dużej liczbie niektórych jednostek lub trudnościach związanych z ich wyznaczaniem. Jako przykład można podać sylaby — badacz napotyka problem dużej liczby tych jednostek (jest ich wiele tysięcy) oraz problem z wyznaczeniem ich granic (problem ten w szczególności dotyczy języka polskiego ze względu na występujące powszechnie wieloelementowe grupy spółgłosek o niespotykanej w innych językach strukturze). Te powody sprawiają, że niektóre problemy badawcze mogą być rozpatrywane tylko fragmentarycznie.

3.2. Przygotowanie materiału badawczego

Po sformułowaniu problemu badawczego można przystąpić do przygotowania materiału badawczego. Badania percepcji audytywnej polegają na prezentowaniu kolejnych bodźców dźwiękowych oraz rejestrowaniu reakcji osoby badanej. Materiał badawczy obejmuje zatem:

- zapis tekstowy bodźców wraz z danymi towarzyszącymi (tekstowa lista lub arkusz danych),
- zapis dźwiękowy bodźców (zbiór plików dźwiękowych).

Opracowanie listy bodźców (zawierającej treści, które będą odtwarzane w teście percepcyjnym) nie jest zadaniem łatwym. Przygotowany materiał badawczy powinien być adekwatny do przyjętych założeń badawczych. Jeżeli potencjalna liczba bodźców jest zbyt duża, to badacz zmuszony jest ograniczyć zakres badań do rozmiaru możliwego do zrealizowania. Dla każdego bodźca trzeba opracować zestaw danych niezbędnych dla realizacji testu — na przykład informacje, które będą pojawiały się na ekranie komputera w czasie odtwarzania plików dźwiękowych lub etykiety przycisków pojawiających się po odtworzeniu bodźca.

Dysponując gotową listą bodźców można przystąpić do nagrywania mowy. Czasami generuje się mowę syntetyczną dla badań percepcyjnych (zagadnienia związane z syntezą mowy nie są poruszane w tym artykule). Oto pytania, na które trzeba odpowiedzieć przed rozpoczęciem nagrań mowy:

- czy nagrywana będzie mowa ciągła (czytany tekst) czy wyrazy izolowane lub mniejsze jednostki?
- ile osób będzie uczestniczyło w nagraniach?
- jaka będzie płeć osób nagrywanych?

- jaki będzie wiek osób nagrywanych?
- czy istnieją dodatkowe wymagania dotyczące nagrań?

Często nagrywane są fragmenty, które obejmują wiele wyrazów, dlatego wymagają one dalszej obróbki — muszą być pocięte (rozdzielone na pojedyncze bodźce) i umieszczone w oddzielnych plikach dźwiękowych. Przed rozdzielaniem na mniejsze pliki można wykonać inne czynności — na przykład przeprowadzić normalizację natężenia dźwięku lub przeprowadzić konwersję podstawowych parametrów dźwięku cyfrowego — liczby kanałów, częstotliwości próbkowania lub rozmiaru próbek dźwięku.

3.3. Przygotowanie testu percepcyjnego

Dysponując zbiorem plików dźwiękowych oraz niezbędnymi danymi można przystąpić do tworzenia testu. Omawiany w artykule pakiet programów obejmuje program służący do przygotowywania testów percepcyjnych. Posługując się tym programem można wygenerować plik zawierający wszystkie niezbędne dane dotyczące przebiegu testu. Oto lista najważniejszych funkcji programu:

- wczytywanie zbiorów plików dźwiękowych,
- ustalanie struktury bodźca — liczby powtórzeń nagrania oraz przerwy pomiędzy kolejnymi odtworzeniami nagrania (liczonej w milisekundach),
- ustalanie struktury odpowiedzi na bodźce — wprowadzono podział na odpowiedzi jawne (wyświetlane w przyciskach) oraz odpowiedzi niejawne (wpisywane do pól tekstowych przy użyciu klawiatury),
- możliwość ustalania dodatkowych pól z danymi — do każdego bodźca można dołączyć dane, które nie są wykorzystywane w teście, jednak mogą być one użyte na etapie analizy wyników,
- randomizacja bodźców,
- zapisywanie do pliku oraz odczytywanie z pliku wersji roboczej testu.

3.4. Przeprowadzenie testu percepcyjnego

Kolejny etap polega na przeprowadzeniu właściwego testu. Liczba osób badanych jest uzależniona od przyjętych założeń, jednak zaangażowanie zbyt małej liczby osób może sprawić, że wyniki będą niewiarygodne. Inny problem związany jest z tym, że pojedynczy test percepcyjny nie powinien być zbyt obszerny (zawierać zbyt dużo bodźców) — może to spowodować znużenie lub zmęczenie osoby badanej oraz wpłynąć negatywnie na jakość wyników. Istotne jest również miejsce, w którym wykonywany jest test — powinno być ono ciche — tak żeby osoba badana mogła się skupić na zadaniu.

Omawiany w artykule pakiet programów obejmuje program przeznaczony do przeprowadzania testów percepcyjnych — jest to jedyny program, z którym

kontakt mają osoby biorące udział w eksperymencie. Program przeprowadza test zgodny z informacjami zawartymi w pliku wygenerowanym przez program opisany w podrozdziale 3.3. Przed rozpoczęciem testu należy wskazać folder przechowujący wszystkie niezbędne pliki dźwiękowe. Po zakończeniu testu wszystkie odpowiedzi zapisywane są do pliku tekstowego w postaci tabelarycznej (jako separator kolumn używany jest znak tabulacji). Jeżeli zaprojektowany test obejmuje dodatkowe pola z danymi, to te dane również zostają zapisane do pliku wynikowego.

3.5. Analiza wyników badań

Analiza wyników to ostatni etap badania. Powstała koncepcja programu ułatwiającego i przyspieszającego analizę wyników uzyskanych z testów percepcyjnych. Program będzie częścią opisywanego w artykule pakietu. Poniżej przedstawiono główne założenia dotyczące programu:

- jednocześnie odczytywanie wielu wyników (pochodzących z badań przeprowadzanych na różnych osobach),
- grupowanie odpowiedzi na podstawie dodatkowych informacji zapisanych w polach z danymi,
- tworzenie rozkładów pojedynczych odpowiedzi lub rozkładów grup odpowiedzi,
- zapisywanie wyników do pliku.

4. Podsumowanie

W artykule omówiono najważniejsze zadania związane z badaniami nad percepcją mowy. Przedstawione informacje zostały oparte przede wszystkim na doświadczeniu autora, które wynika z jego badań wcześniejszych. Zatem informacje te nie mają kształtu norm czy reguł i mogą one podlegać dalszej dyskusji oraz modyfikacjom. Celem podjętych działań jest uproszczenie i przyspieszenie procedury badawczej.

W artykule omówiono także szereg potencjalnych problemów występujących na kolejnych etapach badań. Wiedza ta może być pomocna przy projektowaniu badań, a także może ułatwić uniknięcie błędów pogarszających jakość wyników.

W czasie pisania artykułu istniały prototypowe wersje dwóch programów — programu służącego do projektowania badań oraz programu przeznaczonego do przeprowadzania testów percepcyjnych. Powstały również założenia dotyczące programu służącego do analizy wyników badań. Kolejne prace będą związane z tworzeniem tego programu oraz z testowaniem działania całego pakietu programów.

BIBLIOGRAFIA

- BASSO A., CASATI G., VIGNOLO L., 1977, *Phonemic identification defect in aphasia*, „Cortex”, t. 13, s. 85–95.
- BAUM S., 2001, *Contextual influence on phonetic identification in aphasia: The effect of speaking rate and semantic bias* [w:] *Brain and Language*, t. 76, s. 266–281.
- COOPER R. S. (red.), 1952, *Some experiments on the perception of synthetic speech sound*, „Journal of the Acoustical Society of America”, 24 (6), s. 597–606.
- ELMAN J., MCCLELLAND J. L., 1984, *The interactive activation model of speech percetion. Language and speech*, Lass N., New York.
- GLEASON J. B., RATHNER N. B., 2005, *Psycholingwistyka*, GWP, Gdańsk, s. 121–171.
- GOW D., CAPLAN D., 1996, *An examination of impaired acoustic-phonetic proccsing in aphasia* [w:] *Brain and Language*, t. 52, s. 386–407.
- MASSARO D., *Testing between TRACE model and fuzzy logical model of speech perception* [w:] *Cognitive Psychology*, 1989, t. 21, s. 398–421.
- OBREBOWSKI A., 2011, *Wybrane zagadnienia z audiometrii mowy*, Wydawnictwo Naukowe UM w Poznaniu, Poznań.
- PICKET J. M., 1980, *A primer of acoustic phonetics and speech perception. The sounds of speech communication*, University Park Press, Baltimore.
- PISONI D. B., REMEZ R. E., 2009, *The Handbook of Speech Perception*, Blackwell Publishing.
- STEVENS K. N., OHMAN S. E. G., 1969, *Crosslanguage study of vowel perception* [w:] *Language and Speech*, t. 12, s. 1–23.
- STRANGE W., 1989, *Evolving theories of vowel perception*, „Journal of Acoustical Society of America”, 85, s. 2081–2087.