© Nowa Audiofonologia® 4(4), 2015: 28–35
DOI: 10.17431/895328

Nowa Audiofonologia®

Prace poglądowe

Ocena akustyczna artykulacji głosek

Acoustic assessment of speech sounds

Małgorzata Urszula Waryszak, Zdzisław Marek Kurkowski

Uniwersytet Marii Curie-Skłodowskiej, Zakład Logopedii i Językoznawstwa Stosowanego, Lublin

Adres autora: Małgorzata Waryszak, Uniwersytet Marii Curie-Skłodowskiej, Zakład Logopedii i Językoznawstwa Stosowanego, ul. Sowińskiego 17, 20-040 Lublin, e-mail: M.U.Waryszak@gmail.com

Streszczenie

Celem artykułu jest ukazanie możliwości oceny artykulacji głosek z zastosowaniem analizy akustycznej. W artykule przedstawiono przegląd literatury polskiej i zagranicznej, dotyczącej badań akustycznych artykulacji prowadzonych na potrzeby logopedii i językoznawstwa. Omówiono metody zastosowane przez cytowanych autorów oraz najważniejsze wnioski z ich badań.

Słowa kluczowe: akustyka mowy • artykulacja • dźwięki mowy

Abstract

This article aims to show various means of assesing articulation of sounds using acoustic analysis. It is the overview of Polish and foreign literature concerning the acoustic research on articulation carried out for the needs of speech therapy and linguistics. Methods used and the most important conclusions of the cited studies are discussed.

Key words: speech acoustics • articulation • speech sounds

Wprowadzenie

Hogden i wsp. [1] przeprowadzili eksperyment, który dowiódł, że na podstawie pomiarów akustycznych można odwzorować ruchy artykulacyjne. Oprócz nagrań audio, które następnie zostały przeanalizowane z zastosowaniem metod akustycznych, jednocześnie rejestrowano ruchy artykulacyjne osoby badanej za pomocą artykulografu elektromagnetycznego EMMA (ang. *electromagnetic midsagittal articulometer*). Znaczniki przymocowano na wargach, szczęce górnej i dolnej, na języku i grzbiecie nosa. Badanie polega na tym, że osobę badaną umieszcza się w polu magnetycznym, stąd ruchy znaczników (wywołane odpowiednimi ruchami artykulacyjnymi) mogą być rejestrowane i analizowane w układzie współrzędnych. Zbadano jedną osobę. Jej zadaniem było wypowiedzenie 90 logatomów utworzonych tak, że w nagłosie i wygłosie umieszczono głoskę [g], zaś w śródgłosie – dwie samogłoski, wybrane z grupy dziewięciu szwedzkich i jednej angielskiej [ɛ]. Do analizy akustycznej zastosowano widmo FFT gładzone cepstralnie z oknem Hamminga 26,5 ms. Widmo zostało posegmentowane, a każdy segment, odpowiadający jednemu położeniu artykulatorów, znormalizowano oraz przypisano mu reprezentację wektorową. Wszystkie wektory poddano kategoryzacji wektorowej (ang. *Vector Quantization,* VQ). Następnie zastosowano algorytm FSCL (ang. *Frequency Sensitive Competitive Learning*) w celu uporządkowania danych. Położenie znaczników, szacowane na podstawie pomiarów akustycznych, było w 94% zgodne z rzeczywistym, z dokładnością do 2 mm. Przytoczone wyniki świadczą o tym, że ocena artykulacji głosek z zastosowaniem analizy akustycznej jest wiarygodna.

Zalety i ograniczenia metod analizy akustycznej w badaniach nad artykulacją

Zaletą akustycznych badań nad artykulacją jest to, że osoba badana podczas eksperymentu mówi swobodnie, a jej wypowiedzi mogą być dowolne (w przeciwieństwie na przykład do badań palatograficznych, kiedy należy wypowiadać głoski w izolacji, mając w ustach „sztuczne podniebienie" – płytkę, na której odwzorowywane są miejsca kontaktu języka z *palatum*). W szczególnych okolicznościach osoba badana może być nieświadoma tego, że jej wypowiedzi są rejestrowane lub będą wykorzystane do oceny wymowy, wtedy takie nagranie w pełni oddaje naturalną artykulację [2].

Ograniczeniem badań akustycznych zawsze były urządzenia techniczne, które powinny zapewniać najwyższą jakość rejestrowanego dźwięku oraz precyzyjnie odwzorowywać przebiegi akustyczne w postaci wykresów. Dawniej używano nagrań z taśm magnetofonowych lub płyt gramofonowych i transponowano ten zapis za pomocą różnych urządzeń elektronicznych na spektrogramy lub oscylogramy. Rejestracja odbywała się za pomocą pisaków umieszczonych na ruchomej taśmie, innym razem ślad na papierze pozostawiały przeskakujące iskry. Szczegóły tych

historycznych już metod badań akustycznych wyczerpująco opisali np. von Essen [2] i Jassem [3]. Współcześnie użycie komputerów znacznie ułatwia i przyspiesza proces analizy akustycznej, ponadto nagrania dźwiękowe są wysokiej jakości oraz w postaci cyfrowej.

Sposoby wizualizacji sygnału akustycznego

Istnieją różne sposoby graficznej wizualizacji sygnału akustycznego. Warto wymienić kilka podstawowych. Oscylogram jest wykresem zależności zmian ciśnienia akustycznego sygnału względem czasu. Widmo zaś ukazuje poziom natężenia dźwięku w funkcji częstotliwości. Z kolei spektrogram to wykres w swej istocie bardzo zbliżony do widma, ale jest reprezentacją trójwymiarową: oprócz amplitudy i częstotliwości uwzględnia również trzeci czynnik – czas. Parametr ten jest zmienną niezależną, częstotliwość – zmienną zależną, natomiast amplituda składowych częstotliwości sygnału jest wyrażona za pomocą stopnia zaciemnienia (proporcjonalnego do wzrostu intensywności) [3].

Na podstawie otrzymanych wykresów można opisywać i porównywać ze sobą różne parametry fonetyczno-akustyczne, na przykład: iloczas głoski, natężenie, częstotliwość podstawową (liczba drgań w jednostce czasu), strukturę widmową [4]. Jeśli w jakimś odcinku przebiegu akustycznego takie parametry są stałe lub zmieniają się w określonym kierunku, jest to segment fonetyczno-akustyczny. Dzięki zaobserwowaniu tej powtarzalności można było przypisać głoskom charakterystyczne segmenty akustyczne i zorganizować je w określoną hierarchię [3]. Sprawą wymagającą odrębnych rozważań jest rozróżnienie między głoską a fonemem, co zostało wyjaśnione np. przez Jassema [3].

Klasy głosek

Dokonując oceny akustycznej artykulacji, należy przyjąć wybraną klasyfikację głosek. Von Essen [2] proponuje ich klasyfikację z uwzględnieniem sposobu artykulacji i wyróżnia: samogłoski ustne i nosowe, półsamogłoski, spółgłoski: zwarto-wybuchowe, trące, lateralne, drgające, nosowe, ejektywne, iniektywne i mlaski. Niektóre z zaproponowanych nie występują w systemie języka polskiego. Z kolei Jassem [3] proponuje zestawienie dwóch podziałów. Pierwszy, który uwzględnia cechy artykulacyjne, jest następujący: samogłoski (w ich obrębie różne typy), spółgłoski nosowe, boczne i trące, głoski zwarte i inne. W drugim podziale przyjmuje kryterium sposobu pobudzania rezonatorów oraz powtarzalności cech widmowych głosek. Na tej podstawie wyróżnia się: aspiraty (samogłoski bezdźwięczne), dźwięczne bezszumowe (tu również segment zwarcia dźwięcznych głosek zwartych), dźwięczne szumowe, szumowe i udarowe (segment plozji). Natomiast Wierzchowska [5] głoski języka polskiego dzieli na osiem grup na podstawie charakteru przebiegu akustycznego, który jest zależny od udziału fałdów głosowych oraz stopnia zbliżenia narządów mowy. Choć autorka postuluje jednoczesne uwzględnienie artykulacyjnych i akustycznych aspektów głosek, w swojej klasyfikacji skupia się przede wszystkim na opisie akustycznym oraz wymienia przykładowe głoski, pasujące do wybranej grupy, aspekt artykulacyjny traktując domyślnie. Wyróżnia: [5, s. 101]:

- przebiegi quasi-periodyczne ustalone (tu należą samogłoski, zwane przez autorkę głoskami otwartymi ustnymi, spółgłoski nosowe oraz głoski boczne),
- przebiegi quasi-periodyczne nieustalone (samogłoski nosowe i unosowione),
- szumy (w tej grupie autorka wymienia spółgłoski szczelinowe bezdźwięczne),
- impulsy (głoski zwarto-wybuchowe bezdźwięczne),
- kombinacje słabego impulsu i następującego po nim szumu (zwarto-szczelinowe bezdźwięczne),
- kombinacje szumu i przebiegu quasi-periodycznego (głoski szczelinowe dźwięczne),
- kombinacje impulsu i przebiegu quasi-periodycznego (głoski zwarto-wybuchowe dźwięczne i drżąca [r]),
- kombinacje słabego impulsu, szumu i przebiegu quasi-periodycznego (głoski zwarto-szczelinowe dźwięczne).

W niniejszym artykule opisane zostaną głoski według następujących kategorii – samogłoski ustne[1], głoski nosowe[2], spółgłoski zwarte[3], głoski trące[4], głoski półotwarte[5] i głoski drżące[6].

Samogłoski ustne

Wierzchowska [5] w rozważaniach skupia się przede wszystkim na analogii między układem artykulacyjnym a wartościami formantów[7] dźwięku mowy, czyli składowych częstotliwości o wyróżniająco wysokiej energii. Zwraca uwagę, że w największym stopniu na różnice barwy dźwięków mowy wpływają pierwsze dwa formanty. Wysokość formantu pierwszego zmienia się proporcjonalnie do wielkości otworu wargowego, natomiast formant drugi związany jest z układem masy języka. Wierzchowska [5] przekonuje, że można tą metodą badać wszystkie głoski. Jassem [3] jednak stwierdził, że pomiar częstotliwości formantowych pozwala w sposób istotny rozróżnić wyłącznie samogłoski, ale nie wystarcza do wyczerpującego opisania

1. „Samogłoski to takie alofony, które powstają podczas swobodnego przepływu powietrza wzdłuż środkowej linii języka" [3, s. 123].
2. Artykulację głosek nosowych cechuje otwarcie wylotu do jamy nosowej z gardła poprzez opuszczenie podniebienia miękkiego [6].
3. Głoski artykułowane poprzez utworzenie zwarcia (obu warg, zwarcia języka z wałkiem dziąsłowym, podniebieniem twardym lub miękkim albo z języczkiem lub zwarcia fałdów głosowych) i jego uwolnienie pod wpływem wzrastającego ciśnienia zgromadzonego w jamie ustnej powietrza. Wyróżnia się głoski zwarto-wybuchowe (segment zwarcia i plozji) oraz zwarto-szczelinowe (po segmencie zwarcia następuje słaby wybuch i segment szumowy) [6].
4. Wymawiane dzięki utworzeniu wąskiej szczeliny. Przepływające przez nią powietrze wpada w wibracje i wytwarza szum. Miejscem artykulacji jest obszar największego przewężenia [6].
5. Ich artykulacja jest zbliżona do samogłosek, różni ją jednak krótszy iloczas oraz możliwość pełnienia funkcji sylabotwórczej [6].
6. Wytwarzane za pomocą wibracji ponadkrtaniowych [6].
7. Formant dźwięku to pasmo wzmocnionych amplitudowo częstotliwości drgań własnych kanału głosowego. Kolejne formanty oznacza się literą F i przypisaną jej kolejną liczbą, np. F_1, F_2… [5].

innych głosek. Udowodnił doświadczalnie, że czworoboki samogłoskowe, wyznaczane na podstawie położenia masy języka na przekrojach artykulacyjnych samogłosek polskich, odpowiadają pętlom formantowym, a te wykreśla się, umieszczając zmierzone dla każdej wypowiedzianej samogłoski wartości pierwszych dwóch formantów na płaszczyźnie o współrzędnych F_1 i F_2. Ta zgodność w odwzorowaniu układu samogłosek klasyfikowanych według kategorii akustycznych i artykulacyjnych jest dowodem na to, że akustyczne metody oceny głosek są wysoce miarodajne.

W latach dziewięćdziesiątych XX wieku rozpowszechniła się już spektrografia komputerowa, ale bazowała na wcześniejszych metodach mechanicznych. Gonet [4] przeprowadził badania spektrograficzne z wykorzystaniem analogowych i cyfrowych spektrogramów dziesięciu jednozdaniowych wypowiedzi, zrealizowanych niezależnie przez czterech lektorów. Z tych przebiegów wyodrębnił i zbadał formanty samogłosek. Udowodnił, że pola samogłoskowe na płaszczyźnie $F_1(F_2)$ układają się w wyraźne czworoboki, jednak głoski [a], [o], [e] cechuje duże rozproszenie. Wskazuje to na możliwość istnienia allofonów tych głosek, a także uwidacznia konieczność uwzględnienia w badaniach efektu koartykulacji. Ponadto obnaża niedoskonałości stosowania wartości średniej formantów. Nieoczekiwaną obserwacją z przedstawionych badań jest częściowe nałożenie się pól samogłosek [y] oraz [e]. Trudno jednak na podstawie cytowanego artykułu wysnuć jednoznaczne wnioski na temat normy wymowy polskiej, ponieważ zbadano małą grupę osób, o których niewiele wiadomo [4].

Częstotliwości formantowe można wyznaczać szacunkowo na podstawie widma z dokładnością ±20 Hz. Istnieją jednak dokładniejsze metody obliczeniowe, wykorzystywane w programach komputerowych. Pierwsza opiera się na wzorze:

$$\log F_m = \frac{\sum A_n \log f_n}{\sum A_n}$$

gdzie:
- F_m – obliczana częstotliwość formantu, m=1,2,3,4,5
- A_n – względna amplituda składowej harmonicznej o numerze n w skali liniowej, odniesiona do maksymalnej amplitudy składowej harmonicznej w obrębie formantu
- f_n – częstotliwość składowej o numerze n

W drugiej metodzie korzysta się z analogicznego wzoru, w którym logarytmy częstotliwości zastąpione są ich bezpośrednimi wartościami. Jest to sposób łatwiejszy, a jednocześnie dla częstotliwości powyżej 500 Hz równie dokładny jak poprzedni. Metoda trzecia jest w zasadzie postulatem, by przed dokonaniem obliczeń zgodnie z wcześniej przedstawionymi metodami wprowadzić poprawkę +6 dB na oktawę [3, s. 199–203].

Badaniem porównawczym monoftongów [i, y, u, e, o, ɤ, ɑ, z, ʐ][8] w dwóch odmianach języka mandaryńskiego zajmowali się Xu oraz Deterding [7]. Badano dorosłe osoby z Brunei i Pekinu. Ich zadaniem było wypowiadanie słów, w których badane głoski tworzą sylabę otwartą. Do analizy wykorzystano program PRAAT. Mierzono formanty F_1 i F_2 na podstawie spektrogramu, na środku samogłoski. Wyniki nanoszono na płaszczyznę $F_1(F_2)$, wcześniej ujednolicając wyniki za pomocą skali barkowej. Następnie liczono odległość Euklidesową między poszczególnymi klastrami głosek. Dzięki temu otrzymano wyniki niezależne od naturalnej wysokości głosu każdego z mówców. Dodatkowo zastosowanie skali barkowej sprawia, że wyciągnięte z doświadczenia wnioski są spójne z doświadczeniem percepcyjnym słuchaczy. Zmierzono także formant F_3 w celu zbadania takich zjawisk jak retrofleksja[9] czy zaokrąglenie warg. Samogłoski labio-welarne cechuje obniżenie wartości F_2 względem głoski neutralnej (por. Jassem [3]). Okazuje się, że dotyczy to również formantu F_3 i jego dodatkowe zastosowanie pozwala na dokładniejsze opisanie różnic pomiędzy głoskami blisko położonymi na płaszczyźnie $F_1(F_2)$. Wszystkie formanty zostały automatycznie wykryte za pomocą programu PRAAT. Badania wykazały, że osoby z Pekinu wyraźnie odróżniają głoski [i, y], zaś w Brunei są to warianty fakultatywne.

Temat zaburzonej wymowy samogłosek podjęto w badaniach Sapira i wsp. [8]. Analizowano mowę osób z chorobą Parkinsona przed leczeniem i po leczeniu w celu ewaluacji skuteczności terapii. Na podstawie nagranych wypowiedzi oceniano samogłoski [i], [u], [a] w śródgłosie wyrazów. Zmierzono wartości formantów F_1 i F_2 na fragmencie spektrogramu głosek (dla samogłoski [u] uwzględniono ostatnie 30 ms, dla pozostałych – środkowe). Dodatkowo wyznaczono parametry:
- VSA (ang. *Vowel Space Area*) – pole samogłosek,
- lnVSA,
- współczynnik F2i/F2u,
- FCR (ang. *Formant Centralization Ratio*) – wskaźnik centralizacji.

Zauważono, że efekty leczenia można monitorować wyłącznie z zastosowaniem parametrów F2i/F2u i FCR. Wysunięto propozycję zastosowania parametru FCR zamiast powszechnie stosowanego VSA. Lansford i Liss [9] również krytycznie odnoszą się do wykorzystywania parametru VSA. Ich badania dotyczyły zrozumiałości wymowy samogłosek u osób z dyzartrią[10]. Porównywano wyniki rozpoznawania głosek na podstawie analizy akustycznej oraz audytywnej. Stwierdzono, że przewidywanie, jakich odpowiedzi udzielą słuchacze, jest trudne na podstawie parametru VSA, ponieważ zależy on m.in. od czynników takich jak: materiał językowy, płeć badanego, choroba będąca przyczyną dyzartrii. Parametr FCR i średni rozrzut samogłosek narożnych względem [ə] (średnie odległości [i], [æ], [a], [u] od [ə] na płaszczyźnie o współrzędnych F_1 i F_2) wykazały najsilniejsze korelacje z wynikami

[8] Autorzy przyznają, że włączenie [z, ʐ] do monoftongów jest kontrowersyjne, ale zdecydowali się na to ze względu na tradycje badań języka mandaryńskiego oraz podobieństwo sposobu artykulacji tych głosek i samogłosek [6].

[9] Sposób artykulacji głosek z takim uformowaniem masy języka, że jego czubek jest zagięty ku górze lub do tyłu w kierunku podniebienia twardego [6].

[10] Dyzartria jest definiowana jako „zaburzenie oddechowo-fonacyjno-artykulacyjno-prozodyczne spowodowane uszkodzeniem ośrodków podkorowych i dróg unerwiających aparat mowy" [10].

zrozumiałości. Jednak według autorek parametr FCR nie nadaje się do automatycznego odróżniania osób z dyzartrią od osób bez zaburzeń mowy (zbyt niska swoistość).

W badaniach nad parami samogłosek języka angielskiego [11] ([i]: [ɪ], [ɑ]: [ʌ], [u]: [ʊ]) również mierzono wartość formantów F_1–F_3 (zastosowano skalę barkową), dodatkowo wyznaczono czas trwania samogłosek oraz parametr VISC (*ang. Vowel Inherent Spectral Change*). Pomiarów dokonano za pomocą programu PRAAT, a opracowano z użyciem programu MATLAB. Osoby badane wypowiadały wyrazy o strukturze [kVd], gdzie V to badana samogłoska. Na uwagę zasługuje parametr VISC, który pozwala opisać dynamikę przebiegów formantowych. Zmianę wartości formantów opisuje parametr Λ (ang. *spectral change*), zaś zmianę kierunku przebiegu Ω (ang. *spectral angle*). Wartości te wyznacza się na podstawie równań, uwzględniających wartości formanów F_1–F_3, zmierzonych przy początku i przy końcu samogłoski. Ten sam parametr stosowali Jin i Liu [12]. Wniosek z ich badań jest następujący: wartość Λ, w przeciwieństwie do Ω, różnicuje wymowę osób, dla których badany język jest ojczystym, i tych, dla których nie jest.

Głoski nosowe

Odrębną grupą są samogłoski nosowe. Ciekawy sposób ich analizy proponuje Chen [13]. Cytowane badania dotyczyły porównania samogłosek ustnych oraz unosowionych (umieszczonych w otoczeniu dwóch spółgłosek nosowych) w języku angielskim oraz porównania dwóch momentów przebiegu samogłosek nosowych języka francuskiego[11] – o najsilniejszym i najsłabszym nasileniu nosowości. Chen postuluje wykorzystanie widma – odnotowuje w nim obecność dodatkowych wierzchołków: jednego pomiędzy pierwszymi dwoma formantami o amplitudzie P1 oraz drugiego w niskich częstotliwościach, często poniżej pierwszego formantu, o amplitudzie P0. Amplituda pierwszego formantu (A1) również obniża się wraz z nasileniem nosowości. Zmierzenie tych wartości i obliczenie stosownych korelatów (A1-P1 oraz A1-P0) pozwala na opisanie nasilenia nosowości. Wspomniane wartości mierzono na początku samogłoski, 20 ms od początku, w środku i na końcu samogłoski. Ważnym postulatem cytowanych badań jest wykorzystanie wszechkierunkowego mikrofonu, umieszczonego tak (ok. 15 cm od twarzy), by zminimalizować efekt pogłosu, występujący ze względu na różnicę odległości mikrofonu od ust i od nosa mówiącego. Do analizy akustycznej wykorzystano program KLSPEC93 z pakietu Klatt.

Vallino-Napoli i Montgomery [14] do badań nad stopniem nosowania u osób z rozszczepem podniebienia wykorzystali nasometr (Kay Elemetrics), który tworzą dwa mikrofony – jeden zbierający sygnał z ust, drugi z nosa. Wynikiem badania jest obliczona wartość nosowania, czyli iloraz energii akustycznej zebranej z górnego mikrofonu do sumy energii zebranej z obu mikrofonów, podzielony przez 100.

Lorenc, Święciński i Król [15] przedstawiają 16-kanałowe urządzenie (MARP-16, *ang. 16-channel microphone-array recorder/processor*), rejestrujące energię akustyczną płynącą zarówno ustami, jak i nosem (dodatkowo połączone z kamerą). Sygnał akustyczny jest zapisywany w bezstratnym formacie WAV w postaci 16 plików. Choć przedmiot opisanych w artykule badań był inny, przedstawione urządzenie z powodzeniem może być wykorzystane do oceny głosek nosowych.

Spółgłoski zwarte

Sprawą wartą uwagi jest dźwięczność spółgłosek zwartych. Opis akustyczny kontrastu fonologicznego w tym zakresie jest złożony. Dla wielu języków, w tym języka polskiego, wystarczająca jest analiza segmentu zwarcia na spektrogramach (w głoskach dźwięcznych obserwuje się niskoczęstotliwościowe składowe harmoniczne, w bezdźwięcznych – ciszę akustyczną). Jednak w innych językach lub w zaburzeniach mowy występują również dodatkowe różnicujące atrybuty fonetyczne, na przykład segment szumowy, zwany aspiracją (przydechem), lub siła plozji. Parametrem łączącym wymienione cechy jest VOT (ang. *Voice Onset Time*), czyli czas rozpoczęcia dźwięczności) [16–19]. „Stopień dźwięczności głosek zwarto-wybuchowych możemy określić zatem jako iloczas drgań periodycznych, które mogą wyprzedzać zwolnienie zwarcia lub następować po nim" [17, s. 99]. Przykładowo w języku gaelickim szkockim jest to jedyny parametr różnicujący głoski [d] i [t] w nagłosie [17]. Trochymiuk (później Lorenc) [17,18] przyjęła następujące zasady pomiarów:

- wykorzystano oscylogram, pomocniczo posługując się spektrogramem,
- punktem odniesienia i czasem zerowym była realizacja plozji spółgłoski zwartej,
- punktem pomiarowym momentu uwolnienia był pierwszy impuls, a konkretnie jego pierwsze przecięcie osi w chwili wzrostu,
- punktem pomiarowym momentu rozpoczęcia drgań fałdów głosowych było przecięcie osi oscylogramu podczas wzrostu wartości w pierwszym okresie serii periodycznych drgań,
- VOT zerowy – gdy moment plozji pokrywa się z momentem rozpoczęcia drgań fałdów głosowych,
- VOT dodatni – gdy moment plozji wyprzedza pojawienie się drgań quasi-periodycznych,
- VOT ujemny – gdy moment plozji pojawia się później niż drgania quasi-periodyczne,
- dodatkowo mierzono iloczas impulsów plozji, a jeśli występowały wielokrotne, traktowano je łącznie.

Opisaną metodologię wykorzystano w badaniach nad mową osób z niedosłuchem, które posługują się fonogestami [16,17]. Dokonano akustycznej oceny kontrastu fonologicznego w zakresie dźwięczności spółgłosek zwarto-wybuchowych w grupie dzieci w normie biologicznej oraz dzieci z obustronnym, głębokim uszkodzeniem słuchu. Ustalono, że w mowie dzieci niesłyszących występują przeważnie ubezdźwięczniania głosek dźwięcznych, jednak nie jest to zjawisko jednorodne – zaobserwowano siedem sposobów realizacji tych głosek [16, s. 82–83]:

- całkowicie dźwięczne,
- częściowo dźwięczne,
- częściowo bezdźwięczne,

11. W języku francuskim nosowość jest cechą dystynktywną, ale samogłoski nosowe nie mają swoich ustnych odpowiedników.

- częściowo bezdźwięczne w połączeniu z długą afrykacją,
- bezdźwięczne,
- bezdźwięczne w połączeniu z krótką afrykacją,
- bezdźwięczne w połączeniu z długą afrykacją.

Należy zwrócić uwagę na rozróżnienia między dwoma segmentami szumowymi: afrykacją i aspiracją. Wierzchowska [19] oba te zjawiska nazywa ogólnie „przydechem". Natomiast inni [3,16] rozgraniczają afrykację i aspirację. Pierwsza może się pojawić po plozji spółgłosek zwarto-wybuchowych dźwięcznych i bezdźwięcznych i plasuje się w wysokich częstotliwościach, natomiast druga fakultatywnie występuje wyłącznie po segmencie plozji spółgłosek bezdźwięcznych, a jej charakterystyka widmowa jest wyrównana (rozciąga się we wszystkich częstotliwościach).

Omówienia wymaga status tzw. spółgłosek wargowych miękkich. Spór dotyczy tego, czy są to warianty pozycyjne odpowiednich fonemów twardych czy połączenia dwufonemowe. Łobacz [16] opisuje przykład pomiaru iloczasu wyrazów *bez, bies, bis, jest*. Wyniki świadczą za bifonematyczną strukturą omawianych głosek [por. 3].

Kant i wsp. przeprowadzili badania akustyczne nad mową 15 dzieci z głębokim, prelingwalnym niedosłuchem użytkujących implanty ślimakowe [20]. Grupą kontrolną było 15 dzieci normalnie słyszących. Do analizy akustycznej wykorzystano program PRAAT. Oceny dokonano na podstawie spektrogramów. Dla głosek zwarto-wybuchowych zmierzono środkową częstotliwość i natężenie wybuchu oraz VOT. Badano różne głoski, ale istotne statystycznie odchylenia od normy w wymowie osób z implantem ślimakowym zauważono wyłącznie w dwóch przypadkach. Były to: podwyższona środkowa częstotliwość wybuchu dla głoski [p] oraz wydłużony względem normy VOT głoski [b].

Głoski trące

Jassem [3], w odróżnieniu od Wierzchowskiej [5], twierdził, że pomiar częstotliwości formantowych nie jest wystarczający, by rozróżniać głoski inne niż samogłoski. Na poparcie swojej tezy podaje, że do badania segmentów szumowych należy dodatkowo uwzględnić amplitudy mierzonych formantów, częstotliwości antyformantów[12] oraz zakres szumów, które mierzy się 20 i 30 dB poniżej wartości szczytowej poziomu amplitudy w widmie. Ponadto Jassem [3] zwraca uwagę, że otoczenie samogłoskowe wywiera silny wpływ na wartości formantów spółgłosek.

Temat cech identyfikacyjnych spółgłosek trących jest poruszany w pracy Łobacz i Dobrzańskiej [22]. Autorki badały dzieci przedszkolne w różnym wieku w celu odkrycia norm rozwojowych wymowy głosek trących i zwarto-trących. Materiał badawczy stanowiły nagrania 28 specjalnie dobranych jednostek (wyrazów i wyrażeń przyimkowych), wypowiadanych przez 19 dzieci. Nagrania poddano analizie akustycznej z zastosowaniem tzw. spektrogramów szerokowstęgowych całych wyrazów. Podstawa pomiarów to widmo gładzone cepstralnie oraz (pomocniczo) widmo LPC i widmo uśrednione długiego transientu. Kategorię dźwięczności oceniano na podstawie oscylogramów. Przyjęto, że „o naturze sybilantności świadczy względnie silna energia w wysokich rejonach widma" [22, s. 13], natomiast w partiach niższych (zakres F_2 sąsiadujących samogłosek) zaznacza się wpływ kontekstu. Wyznaczano cztery najwyraźniejsze szczyty w obwiedni widma głosek trących (powyżej 2 kHz) oraz trzy w przypadku zwarto-trących. Wyniki poddano analizie statystycznej. Wyznaczone wartości szczytów widmowych cechował duży rozrzut, zależnie od wieku dzieci. Autorki zwracają uwagę na dwa charakterystyczne zjawiska, które umożliwiają dystynkcję głosek trących: segmenty szumowe i ugięcia formantowe sąsiednich samogłosek. Porównanie kształtu widm sybilantów w wymowie badanych dzieci oraz u dorosłych wykazało wyższą amplitudę niskich częstotliwości u przedszkolaków. Mimo to stwierdzono brak silnego wpływu otaczającego kontekstu na artykulację głosek trących. Nie potwierdzono również przekonania o tym, że charakterystyka akustyczna segmentu szumowego wykazuje ścisłą odpowiedniość do tych grup spółgłosek trących, które posiadają to samo miejsce artykulacji, co dowodzi szybszego przyswajania przez dzieci artykulacji głosek trących niż zwarto-trących.

W akustycznych badaniach amerykańskich [23] podjęto próbę wyznaczenia charakterystycznych cech akustycznych głosek trących. Badaniem objęto grupę dwudziestu dorosłych osób, które wypowiadały logatomy o strukturze VCV, utworzone z kombinacji ośmiu angielskich głosek trących [f], [θ], [s], [ʃ], [v], [ð], [z], [ʒ] z samogłoską neutralną [ɑ], za pierwszym razem w sposób dla siebie naturalny (ang. *conversational speech*), a za drugim tak, jakby rozmawiały z osobą z uszkodzonym słuchem (ang. *clear speech*, co można przetłumaczyć jako „wymowa staranna"). Do oceny akustycznej nagrań używano programu PRAAT, a do obliczeń – programu MATLAB. Wykorzystano dyskretną transformatę Fouriera (DFT). Zmierzone parametry akustyczne:

1) częstotliwość szczytowa (peak),
2) pierwsze cztery momenty widmowe (M1–M4),
3) wartość formantu drugiego wszystkich głosek (wyznaczona metodą liniowej predykcji na podstawie początkowej i końcowej częstotliwości szumu głosek trących oraz środkowej częstotliwości samogłoski),
4) średnia częstotliwość podstawowa samogłosek sąsiadujących z głoskami szumowymi (F_0),
5) nachylenie charakterystyki widmowej poniżej i powyżej częstotliwości maksymalnych (ang. *slope below, slope above*),
6) amplituda skuteczna sygnału (amp rms),
7) natężenie sygnału poniżej 500 Hz (amp500),
8) stosunek sygnału do szumu (HNR),
9) iloczas spółgłosek trących,
10) względna amplituda częstotliwości formantu trzeciego (w sybilantach) lub (w pozostałych głoskach) piątego (na tej podstawie wyznaczono względną amplitudę (FSRA), obliczaną jako różnica wartości natężeń zmierzonych dla głoski trącej i samogłoski).

Badania były bardzo rozbudowane. W niniejszym artykule zostaną przytoczone wyłącznie wnioski, które są najistotniejsze w dyskusji nad akustyczną oceną artykulacji głosek [23]:

12. Antyformant definiowany jest jako wyróżniające się minimum energii, jakie można zaobserwować w charakterystycznych dla poszczególnych głosek nosowych pasmach częstotliwości [20].

- Zanotowano systematyczne różnice w parametrach akustycznych głosek wypowiadanych w sposób naturalny oraz „staranny"; dotyczy to zarówno głosek różniących się miejscem, jak i sposobem artykulacji. Te parametry to głównie drugi formant i amplituda rms. Istotna jest też kurtoza wierzchołka (ang. *peak*), obliczana według wzoru: $(M4/M2^2)-3$.
- Znaczące różnice osobnicze były widoczne we wszystkich parametrach oprócz kurtozy i HNR.
- Mówiący mają tendencję do powtarzania głosek tak, aby maksymalnie różniły się one od sąsiednich dźwięków, a szczególnie tych, z którymi pierwotnie słuchacz je pomylił.
- Energia większości głosek trących rozkłada się w wysokich częstotliwościach, wyjątkiem są palatalno-alweolarne [ʃ], [ʒ].
- W wymowie „starannej" obniżył się parametr FSRA (wynik niezgodny z dotychczas opisanymi badaniami).

Analizą polskich spółgłosek trących bezdźwięcznych, realizowanych przez niesłyszących uczniów szkoły podstawowej zajmował się Kleśta [24]. Materiał językowy stanowiło 40 wyrazów z obrazkowego testu nazywania dla słyszących dzieci żłobkowych. Celem badań było wyznaczenie wzorców akustycznych badanych głosek, aby w przyszłości umożliwić ich automatyczną identyfikację. Analiza akustyczna nagrań obejmowała następujące pomiary: częstotliwości F_1–F_5 w zakresie do ok. 8000 Hz, wyznaczone z zastosowaniem LPC-14 (ang. *Linear Predictive Coding*), autokorelacji, preemfazy o wartości 0,9 i okna Hamminga 20 ms. Trudności sprawiało wyznaczenie pierwszego formantu, toteż szczegółowej analizie poddano jedynie pozostałe 4. Na podstawie uśrednionego widma FFT obliczono również pierwszy moment centralny widma, czyli średnią ważoną częstotliwości mierzonych na początku, w środku i na końcu głoski (unikając ugięć formantów) [por. 23]. Poprawność wyznaczenia polskich głosek trących bezdźwięcznych była największa przy jednoczesnym zastosowaniu formantów i momentu centralnego (o 25% lepsze wyniki w stosunku do analizy z wykorzystaniem samych formantów i prawie o 50% lepsze niż przy wykorzystaniu wyłącznie parametru M1). Za każdym razem jednak identyfikacja głosek realizowanych przez osoby niesłyszące wynosiła średnio o około 30% mniej niż w przypadku mowy niezaburzonej. Świadczy to o niestabilnej artykulacji osób z niedosłuchem, najprawdopodobniej spowodowanej osłabioną autokontrolą słuchową. Najlepiej identyfikowany był fonem [x], najgorzej zaś [s]. Zastosowana metoda pozwala na skuteczne rozróżnianie bezdźwięcznych sybilantów od głosek [f] i [x], a także niskoszumowych między sobą. Istotnym mankamentem jest niedostateczna skuteczność w dyskryminacji sybilantów.

Analiza spektrograficzna może być obiektywnym narzędziem do badania tzw. ukrytych kontrastów, czyli bardzo subtelnych, ale istotnych różnic między porównywanymi segmentami [16]. Przykładem są badania nad artykulacją polskiej głoski [x], wymawianej przez mówców arabskich, którzy płynnie posługują się językiem polskim. Dzięki spektrogramom można było zdiagnozować mechanizm deformacji badanej głoski. W wymowie normatywnej [x] jest głoską bezdźwięczną, trącą o welarnym miejscu artykulacji. W wymowie badanych osób wykryto następujące zjawiska: dodatkową artykulację krtaniową, faryngalizację[13] i udźwięcznienie, tłumaczone przeniesieniem nawyków artykulacyjnych z rodzimego języka. Realizowane w ten sposób głoski w ocenie audytywnej przypominały [x] z dodatkowym segmentem szumowym, dźwięczne [γ] lub nawet [r]. Dziwić może fakt, że nie klasyfikuje się tych zjawisk jako zaburzeń mowy, choć ich obrazy akustyczne są takie, jak uzyskane w badaniach nad mową dyzartryczną [16].

Głoski półotwarte

Charakterystyki akustycznej dwóch allofonów głoski [l], występujących w nowofundlandzkiej odmianie języka angielskiego, realizowanych jako [l] lub [ɬ], zależnie od dialektu, dokonali Mackenzie, De Decker i Pierson [25] Próbowali dociec, jakie parametry akustyczne odpowiadają za jasne brzmienie głoski [l] i ciemne [ɬ]. Do badań użyto programu PRAAT. Zmierzono wartości pierwszego i drugiego formantu, w połowie czasu artykulacji badanych głosek. Wartością opisującą „jasność" brzmienia jest różnica F_2 i F_1. Porównywano te wartości również dla głosek w pozycji nagłosowej i wygłosowej.

Cytowane wcześniej przy okazji omawiania urządzeń do rejestrowania sygnału akustycznego badania Lorenc, Święcińskiego i Króla [15] dotyczyły oceny bocznej artykulacji głoski [l], należącej do polskiego systemu językowego. Badane osoby wypowiadały wyrazy z tą głoską występującą w śródgłosie. Wynikiem analizy tych realizacji są trójwymiarowe rozkłady pola akustycznego, które w połączeniu z obrazem z kamery pozwalają określić źródło energii akustycznej. Autorzy przedstawiają także wykresy dwuwymiarowe, obrazujące pionowy lub poziomy rozkład energii akustycznej w czasie. Pierwszy umożliwia sprawdzenie, czy artykułowana głoska jest ustna czy nosowa w każdym momencie artykulacji, natomiast drugi pozwala zobaczyć, jak rozkłada się ciśnienie akustyczne w przestrzeni od jednego do drugiego kącika ust. Oprócz omówionych grafów posiłkowano się również oscylogramami i spektrogramami (do segmentacji sygnału). Procentowy udział energii, pochodzącej z obszaru lewego i prawego kącika ust oraz centrum, także przedstawiono na wykresie w funkcji czasu.

Głoski drżące

Problematykę wpływu otoczenia samogłoskowego na wymowę głosek drżących podjęli Dhananjaya, Yegnanarayana i Bhaskararao [26]. Zbadano między innymi właściwości akustyczyczno-fonetyczne dźwięków wytwarzanych przez okresowe drgania wierzchołka języka (łac. *apex*). Materiałem badawczym były nagrania wypowiadanych sylab o strukturze CV. Spółgłoskę drżącą łączono kolejno z trzema wybranymi samogłoskami: [a], [i], [u]. Wykorzystano metodę predykcji liniowej. Wykrywanie w sygnale miejsca przejść przez zero (ang. *zero-crossing*), a dokładnie *zero-time liftering* oraz *zero-frequency filtering*. Dźwięczne głoski drżące charakteryzuje występowanie na wykresach powtarzających się periodycznie określonych cech, które są wynikiem nałożenia się sygnałów, odzwierciedlających jednocześnie aktywność fałdów głosowych oraz jam

[13] Dźwięki wymawiane gardłowo. Obsada języka zbliżona jest do tylnej ściany jamy gardłowej [6].

rezonacyjnych. Okres takiego sygnału (nazywany w cytowanym artykule *epoch*, czyli „epoką") można wyznaczyć, mierząc na wykresie odległość między kolejnymi przejściami przez zero osi rzędnych. Każde takie przejście świadczy o zamknięciu fałdów głosowych. Dokładne oszacowanie „epoki" pozwala na wyznaczenie parametru F_0 sygnału. Przykładowe wnioski z tych badań są następujące: głoski drżące są mniej podatne na wpływ sąsiadujących samogłosek w porównaniu z głoskami zwarto-wybuchowymi o tym samym miejscu artykulacji. Prawidłowa wymowa głosek drżących wymaga spełnienia określonych warunków atykulacyjno-aerodynamcznych, jednak niestabilne wartości formantu drugiego mogą świadczyć o tym, że można realizować tę samą głoskę drżącą z nieco odmiennym położeniem języka.

Podsumowanie

Istnieją różne sposoby i aspekty wizualizacji sygnału mowy, pozwalające na jego wszechstronną ocenę. Obecnie dzięki programom komputerowym jest to proces łatwiejszy, dokładniejszy i szybszy niż dawniej.

Analizę akustyczną można stosować do oceny artykulacji głosek, gdyż budowa anatomiczna aparatu artykulacyjnego oraz zespół czynności artykulacyjnych mają odbicie w sygnale mowy; wymowa zaburzona zwykle powoduje jego specyficzne zniekształcenia, możliwe do zbadania i opisania. Według Łobacz „usprawnianie mowy zaburzonej powinno się odbywać z silniejszym uwzględnieniem czynników akustyczno-audytywnych niż artykulacyjnych, zgodnie z percepcyjną teorią produkcji mowy" [16, s. 194, por. 18].

Istnieją opracowania na temat zastosowań analizy akustycznej w opisie artykulacji normatywnej i zaburzonej, zarówno w literaturze polskiej, jak i obcej. Badacze wykorzystują różne metody transkrypcji fonetycznej mowy. W niniejszym artykule posługiwano się transkrypcją oryginalną z cytowanych prac. Do wypowiedzi normatywnych realizowanych w języku polskim wystarcza transkrypcja slawistyczna, mimo to warto rozważyć stosowanie symboli IPA w celu ułatwienia dialogu badawczego między przedstawicielami różnych języków oraz do dokładnego opisu wypowiedzi zaburzonych.

Artykuł powstał w związku z realizacją projektu „Zintegrowany system narzędzi do diagnostyki i telerehabilitacji schorzeń narządów zmysłów (słuchu, wzroku, mowy, równowagi, smaku, powonienia)", współfinansowanego przez Narodowe Centrum Badań i Rozwoju w ramach Programu STRATEGMED.

Piśmiennictwo:

1. Hogden J, Lofqvist A, Gracco V, Zlokarnik I, Rubin P, Saltzman E. Accurate recovery of articulator positions from acoustics: New conclusions based on human data. J Acoust Soc Am, 1996; 100(3): 1819–34.
2. von Essen O. Fonetyka ogólna i stosowana. Warszawa: PWN; 1967.
3. Jassem W. Podstawy fonetyki akustycznej. Warszawa: PWN; 1973.
4. Gonet W. Próba określenia normy wymowy polskich samogłosek ustnych. W: Bartmiński J i wsp., red. Opuscula Logopaedica: in honorem Leonis Kaczmarek. Lublin: Uniwersytet Marii Curie-Skłodowskiej; 1993, s. 232.
5. Wierzchowska B. Struktura akustyczna dźwięków języka polskiego w świetle wyników współczesnych badań fonetycznych. Logopedia, 1967; 7: 88–104.
6. Trochymiuk A, Święciński R. Symbole podstawowej transkrypcji Międzynarodowego Towarzystwa Fonetycznego (IPA) i jej rozszerzenia (ExtIPA). Audiofonologia, 2004; 25.
7. Xu S, Deterding D. An acoustic study of monophthongs in Brunei mandarin. Materiały konferencji The International Congress of Phonetic Sciences, Glasgow, 10–14.08.2015.
8. Sapir S, Ramig LO, Spielman JL, Fox C. Formant Centralization Ratio: A proposal for a new acoustic measure of dysarthric speech. J Speech Lang Hear Res, 2010; 53: 114–25.
9. Lansford KL, Liss JM. Vowel acoustics in dysarthria: mapping to perception. J Speech Lang Hear Res, 2014; 57: 68–80.
10. Mirecka U, Gustaw K. Dyzartria w mózgowym porażeniu dziecięcym. Eksperymentalna Skala Dyzartrii jako technika diagnostyczna pomocna w określaniu specyfiki zaburzeń mowy w mpd. Logopedia, 2005; 34: 281.
11. King K, Leung W, Jongman A, Wang Y, Sereno JA. Acoustic characteristics of clearly spoken English – tense and lax vowels. Materiały z konferencji The International Congress of Phonetic Sciences, Glasgow, 10–14.08.2015.
12. Jin SH, Liu C. The vowel inherent spectral change of English vowels spoken by native and non-native speakers. J Acoust Soc Am, 2013; 133(5): 363–69.
13. Chen MY. Acoustic correlates of English and French nasalized vowels. J Acoust Soc Am, 1997; 102(4): 2360–70.
14. Vallino-Napoli LD, Montgomery AA. Examination of the standard deviation of mean nasalance scores in subjects with cleft palate: Implications for clinical use. Cleft Palate Craniofac J, 1997; 34(6): 512–19.
15. Lorenc A, Święciński R, Król D. Assessment of sound laterality with the use of a multi-channel recorder. Materiały z konferencji The International Congress of Phonetic Sciences, Glasgow, 10–14.08.2015.
16. Łobacz P. Wymowa patologiczna a norma fonetyczna w świetle analizy akustycznej. W: Grabias S, Domagała A, Muzyka E, red. Zaburzenia mowy. Mowa – Teoria – Praktyka. Lublin: Uniwersytet Marii Curie-Skłodowskiej; 2001, s. 189–215.
17. Trochymiuk A. Wymowa dzieci niesłyszących. Analiza akustyczna i audytywna. Lublin: Uniwersytet Marii Curie-Skłodowskiej; 2008.
18. Lorenc A. Zaburzenia dźwięczności. Analiza akustyczna i audytywna. Logopedia, 2012; 41: 71.
19. Wierzchowska B. Fonetyka i fonologia języka polskiego. Wrocław: Ossolineum; 1980.
20. Kant AR, Patadia R, Govale P, Rangasayee R, Kirtane M. Acoustic analysis of speech of cochlear implanters and its implications. Clinical and Experimental Otorhinolaryngology, 2012; 5: 14-18.
21. Dukiewicz L, Sawicka I. Fonetyka i fonologia. W: Wróbel H, red. Gramatyka współczesnego języka polskiego. Kraków: Instytut Języka Polskiego PAN; 1995.
22. Łobacz P, Dobrzańska K. Opis akustyczny głosek sybilantnych w wymowie dzieci przedszkolnych. Audiofonologia, 1999; 14: 5–26.

23. Maniwaa K, Jongman A, Wade T. Acoustic characteristics of clearly spoken English fricatives. J Acoust Soc Am, 2009; 125(6): 3962–73.
24. Kleśta J. Analiza akustyczna polskich spółgłosek trących bezdźwięcznych, realizowanych przez dzieci niesłyszące. Audiofonologia, 2004; 26: 105–18.
25. Mackenzie S, De Decker P, Pierson R. An acoustic and articulatory study of [l] allophony in Newfoundland English. Materiały z konferencji The International Congress of Phonetic Sciences, Glasgow, 10–14.08.2015.
26. Dhananjaya N, Yegnanarayana B, Bhaskararao P. Acoustic analysis of trill sounds. J Acoust Soc Am, 2012; 131(4): 3141–52.