Uniwersytet im. Adama Mickiewicza
w Poznaniu

Wydział Fizyki i Astronomii
Katedra Akustytki

Rozprawa doktorska

Wpływ stopnia niedosłuchu i warunków akustycznych na zrozumiałość mowy: walidacja polskiego testu matrix oraz analiza nowych narzędzi wspierających diagnostykę i ocenę percepcji słuchowej

Anna Katarzyna Pastusiak

Promotor:
prof. UAM Jędrzej Kociński
dr Anna Warzybok-Oetjen (Promotor pomocniczy)

POZNAŃ 2024

„...When you speak, speak sincere,
And believe my friend, everyone will hear...”

— Grian Chatten, *A Hero's Death*, 2020

Dziękuję Dziadkom, Rodzicom i Wojtkowi za to, że są,
Prof. Annie Preis i dr Honoracie Hafke-Dys za pomoc i pozytywną energię,
Koleżankom i Kolegom za niezawodną obecność,
Dr Annie Warzybok-Oetjen za wsparcie i życzliwość,
wreszcie – każdemu, kto wziął udział w prowadzonych przeze mnie badaniach.

Szczególnie wdzięczna jestem mojemu Promotorowi, prof. Jędrzejowi Kocińskiemu, za niewyczerpane pokłady cierpliwości, mądrość, pomoc w stawianiu pierwszych kroków na ścieżce naukowej. I wiarę w to, że dam radę (*Don't give up!*).

Wyniki przedstawione w sekcji eskperymentalnej niniejszej rozprawy doktorskiej pochodzą z badań własnych autorki (po uzyskaniu zgody KE UAM – Uchwała nr 19/2023/2024) lub, gdzie zaznaczono, badań realizowanych w szerszej grupie naukowej. Część pomiarów została przeprowadzona ze wsparciem dr Anny Warzybok-Oetjen z Carl von Ossietzky Universität w Oldenburgu, w ramach grantu Deutsche Forschungsgemeinschaft („Multilingual model-based rehabilitative audiology"; nr 25439187). Wszystkie badania, poza opisanymi w części E10, były prowadzone w Katedrze Akustyki Uniwersytetu im. Adama Mickiewicza w Poznaniu.

Summary

Hearing loss of varied etiology and severity affects a significant part of the population, posing a major health issue with broad impact on daily functioning. It can lead to communication difficulties, social isolation, mood decline, depression, and an overall reduction in quality of life. Given these considerations, advancing research to deepen understanding of auditory perception and refine diagnostic tools for its assessment is of critical importance. This approach is especially relevant in the fields of audiology and hearing prosthetics and in the broader psychoacoustic perspective, which integrates room acoustics and speech signal analysis, beyond the functions of peripheral auditory system. For this reason, the scope of this work includes research directly related to hearing impairments while also highlighting entirely different applications of speech intelligibility tests and subjective evaluations of signal perception.

The dissertation focuses on examining speech perception processes in diverse acoustic conditions, combining theoretical foundations with experimental methodologies. The first section explores evolutionary and psychophysiological mechanisms associated with speech production and perception, its physical properties, and a range of cognitive processes, including multimodal ones. Particular attention is given to factors affecting speech intelligibility, such as age-related limitations, hearing loss, and environmental influences like background noise and sound propagation conditions. A literature review discusses speech audiometry as an underutilized diagnostic tool and emphasizes the growing relevance of listening effort as an essential metric for assessing speech signal perception.

The experimental section presents a series of original studies using the Polish sentence matrix test, validating its application in measuring speech intelligibility and listening effort. It was demonstrated that the test, applied and described on such a broad scale for the first time in this dissertation, can be used not only to assess the impact of hearing loss severity on speech intelligibility but also to verify the effectiveness of compensation methods, as evidenced by measurements performed in a group of hearing aid users. In some experiments, objective assessments of speech intelligibility in various masking conditions were complemented by subjective

evaluations. Listening effort, gaining recognition as a critical factor in auditory quality and comfort, was also assessed. This was supported by validating the Polish version of the ACALES scale for listening effort evaluation.

The dissertation also includes findings from research conducted by the author as part of broader projects, which demonstrate the versatility of speech intelligibility studies. This tool is not limited to the precise analysis of hearing loss but can be applied to various aspects of speech signal processing, both in terms of its analysis at higher levels of the central nervous system and in evaluating systems through which speech is transmitted. The first project examined the impact of room acoustic parameters on speech intelligibility and listening effort, while the second involved a pilot multifaceted analysis of auditory system function in patients in the prodromal phase of Alzheimer's disease. The inclusion of these topics in the experimental section significantly enriches the scope of auditory diagnostics research. Despite employing different linguistic materials, the central theme remains speech intelligibility and the evaluation of influencing factors. The results presented argue that speech-based auditory perception assessment can extend beyond diagnosing hearing loss or fitting hearing aids and cochlear implants. It can also validate technical solutions for intentional or incidental speech signal processing (e.g., transmission systems or acoustic environments) and aid in diagnosing other conditions, such as neurodegenerative diseases.

Overall, this dissertation ultimately demonstrates how complex the issue of speech intelligibility is and how many variables influence it, posing a challenge for the development, validation, and practical application of new research tools.

Streszczenie

Niedosłuch o zróżnicowanej etiologii i głębokości dotyka znaczną części populacji, stanowiąc istotny problem zdrowotny o szerokim wpływie na codzienne funkcjonowanie. Może prowadzić do trudności w komunikacji, izolacji społecznej, obniżenia nastroju, depresji, a także ogólnego pogorszenia jakości życia. Mając na uwadze te przesłanki, kluczowe wydaje się prowadzenie badań rozszerzających wiedzę na temat percepcji słuchowej oraz weryfikacji narzędzi diagnostycznych służących jej ocenie. Takie podejście jest szczególnie istotne zarówno w kontekście audiologii i protetyki słuchu, jak i w szerszym ujęciu psychoakustycznym, które integruje akustykę wnętrz oraz analizę komunikatu językowego, wykraczając poza funkcje peryferyjnego układu słuchowego. Z tego względu zakres pracy obejmuje badania związane bezpośrednio z ubytkami słuchu, ale także wskazuje na zupełnie inne zastosowanie testów zrozumiałości mowy i subiektywnej oceny percepcji tego sygnału.

Rozprawa doktorska skupia się na badaniu procesów percepcji mowy w zróżnicowanych warunkach akustycznych, łącząc podstawy teoretyczne z podejściem eksperymentalnym. W pierwszej części pracy przedstawiono ewolucyjne i psychofizjologiczne mechanizmy związane z wytwarzaniem i percepcją mowy, jej fizyczne właściwości oraz szereg procesów poznawczych, w tym multimodalnych. Szczególną uwagę poświęcono czynnikom wpływającym na zrozumiałość mowy, takim jak ograniczenia związane z wiekiem, ubytkiem słuchu, czy czynniki środowiskowe – obecność sygnałów zakłócających oraz warunki propagacji dźwięku. W przeglądzie literatury omówiono również audiometrię mowy jako (wciąż niedoceniane) narzędzie diagnostyczne oraz opisano coraz częściej pojawiające się aspekty związane z wysiłkiem słuchowym jako dodatkowym, niezwykle istotnym narzędziu służącym do oceny percepcji sygnału mowy.

Część eksperymentalna pracy zawiera serię autorskich badań przeprowadzonych z użyciem polskiego testu zdaniowego matrix, walidujących jego zastosowanie w pomiarze zrozumiałości mowy oraz wysiłku słuchowego. Wykazano, że test, który po raz pierwszy został w tak szerokim zakresie zastosowany i opisany w niniejszej rozprawie, może znaleźć zastosowanie nie tylko w ocenie wpływu stopnia

niedosłuchu na zrozumiałość mowy, ale również weryfikacji skuteczności metod jego kompensacji, o czym świadczą pomiary wykonane w grupie użytkowników aparatów słuchowych. W przypadku części badań próbie obiektywizacji oceny zrozumiałości mowy prezentowanej w różnych warunkach maskowania towarzyszyło gromadzenie informacji subiektywnych. Myśląc o ocenie jakościowej oraz komforcie słyszenia, nie można zapomnieć o zyskującym na znaczeniu zjawisku, jakim jest wysiłek słuchowy. W przytoczonych badaniach znaleźć można zapis walidacji polskiej wersji skali ACALES służącej do jego oceny.

Rozprawa obejmuje również wyniki badań prowadzonych przez autorkę w ramach szerszych projektów badawczych, które pokazują uniwersalność badań zrozumiałości mowy. Narzędzie to nie ogranicza się bowiem do jak najdokładniejszej analizy niedosłuchu, ale może być wykorzystane w wielu innych aspektach, w których sygnał mowy jest przetwarzany, zarówno jeśli chodzi o jego analizę na wyższych piętrach centralnego układu nerwowego, jak i przy ocenie układów, w których mowa jest transmitowana. Pierwszy z projektów dotyczył wpływu parametrów akustycznych pomieszczenia na zrozumiałość mowy i wysiłek słuchowy, drugi zaś – pilotażowej analizy wieloaspektowej czynności układu słuchowego u pacjentów w prodromalnej fazie choroby Alzheimera. Te zagadnienia wydają się niezwykle istotne, a ich umieszczenie w części eksperymentalnej stanowi znaczące uzupełnienie przedstawionych badań dotyczących diagnostyki słuchu. Mimo że obydwa eksperymenty wykorzystywały inny materiał językowy, osią łączącą pozostaje zrozumiałość mowy i ocena czynników mogących mieć nań wpływ. Fakt opisu wyników tych badań w niniejszej rozprawie stanowi także argument przemawiający za tym, że ocena percepcji słuchowej, wykorzystująca sygnał mowy, może być stosowana nie tylko w diagnostyce osób z niedosłuchem czy dopasowywaniu aparatów słuchowych, bądź implantów ślimakowych, ale również szerzej – w weryfikacji rozwiązań technicznych przetwarzających mowę w sposób celowy (np. układy transmisji, poprawy zrozumiałości mowy) lub pośrednio (np. pomieszczenia), a także do oceny innych schorzeń, np. neurodegeneracyjnych.

Całość dysertacji pokazuje więc ostatecznie, jak złożonym zagadnieniem jest zrozumiałość mowy i jak wiele zmiennych ma nań wpływ, co stanowi wyzwanie dla tworzenia, walidacji i wykorzystania w praktyce nowych narzędzi badawczych.

Spis treści

1 Wstęp: podstawy ewolucyjne oraz psychofizjologiczne wytwarzania i percepcji mowy 1

2 Podstawowe fizyczne cechy mowy 11

3 Podstawowe aspekty percepcji mowy 17

3.1 Interakcja wzrokowo-słuchowa 21

4 Czynniki zewnętrzne i wewnętrzne wpływające na percepcję mowy 25

4.1 Obecność sygnałów zakłócających 26

4.2 Maskowanie energetyczne i informacyjne 28

4.3 Szybkość mowy 31

4.4 Warunki propagacji – akustyka pomieszczenia 31

4.5 Wiek 32

4.6 Słyszenie binauralne 33

4.7 *Cocktail-party effect* 36

4.8 Ubytek słuchu 37

4.9 Kontekst 39

4.10 Kompensacja poznawcza 40

5 Audiometria mowy jako metoda diagnostyki słuchu 43

5.1 Zasadność oceny zrozumiałości mowy 44

5.2 Techniczne aspekty oceny zrozumiałości mowy 44

5.2.1 Wyrazistość i zrozumiałość 44

5.2.2 Krzywa psychometryczna 45

5.2.3 SRT (*Speech Recognition Threshold*) 47

5.2.4 TCT (*Time-Compression Threshold*) 48

5.2.5 Procedura adaptacyjna 48

6 Rodzaje testów wykorzystywanych w audiometrii i badaniach zrozumiałości mowy 51

6.1 Testy liczbowe 51

6.2 Testy trypletowe . . . 52
6.3 Testy sylabowe . . . 52
6.4 Testy rymowe . . . 53
6.5 Testy logatomowe . . . 53
6.6 Testy wyrazowe . . . 54
6.7 Testy zdaniowe . . . 55
6.7.1 Polski test zdaniowy (PTZ) . . . 56
6.7.2 Test matrix . . . 56
6.8 Inne rodzaje testów . . . 59
6.9 Wpływ treningu na wyniki . . . 60

7 Wysiłek słuchowy 63
7.1 Czynniki wpływające na wysiłek słuchowy . . . 66
7.2 Przegląd istniejących metod pomiaru wysiłku słuchowego . . . 69
7.2.1 Metody obiektywne . . . 70
7.2.2 Metody subiektywne . . . 74

Część eksperymentalna 79
E1 Walidacja polskiego testu zdaniowego matrix . . . 87
E2 Pomiary efektu treningu . . . 95
E3 Nachylenie krzywych psychometrycznych . . . 99
E4 Ocena wpływu szumów zmodulowanych na zrozumiałość mowy – *masking release* . . . 105
E5 Wykorzystanie polskiego testu zdaniowego matrix do oceny zysku z protezowania słuchu . . . 111
E6 Wykorzystanie kwestionariusza *Speech, Spatial and Qualities of Hearing* (SSQ) do subiektywnej oceny łatwości komunikacji i lokalizacji dźwięku . . . 121
E7 Wykorzystanie testu matrix do oceny wysiłku słuchowego (*listening effort*) 127
E8 Zrozumiałość mowy przyspieszonej . . . 141
E9 Weryfikacja dopasowania uzyskanych wyników do klasyfikacji Bisgaarda . . . 149
E10 Wysiłek słuchowy (*listening effort*) w salach o zmiennych cechach akustycznych . . . 155
E11 Zrozumiałość mowy u osób w fazie prodromalnej (bezobjawowej) choroby Alzheimera – wstępne wyniki z projektu *Alzheimer Prediction Project* . . . 163

Podsumowanie 171

Spis rysunków 177

Spis tabel 181

Bibliografia 183

CZĘŚĆ TEORETYCZNA

Rozdział 1

Wstęp: podstawy ewolucyjne oraz psychofizjologiczne wytwarzania i percepcji mowy

Nie bez powodu wybitny polski neurobiolog prof. Jerzy Vetulani pisał, że „mowa jest najbardziej ludzkim osiągnięciem ewolucyjnym" (2012). Poza swoim podstawowym celem (realizowanym zarówno w formie pierwotnej – dźwiękowej, jak i wtórnej – w systemie pisemnym), którym jest utworzona w toku socjalnej i biologicznej ewolucji funkcja komunikacyjna, mowa stanowi także podstawę ludzkiej świadomości (Bickerton, 1990). Wykracza zatem poza stanowienie systemu wzajemnego przekazywania informacji, jak ma to miejsce w przypadku organizmów zwierzęcych, które wykształciły umiejętności porozumiewania się, również z wykorzystaniem głosu.

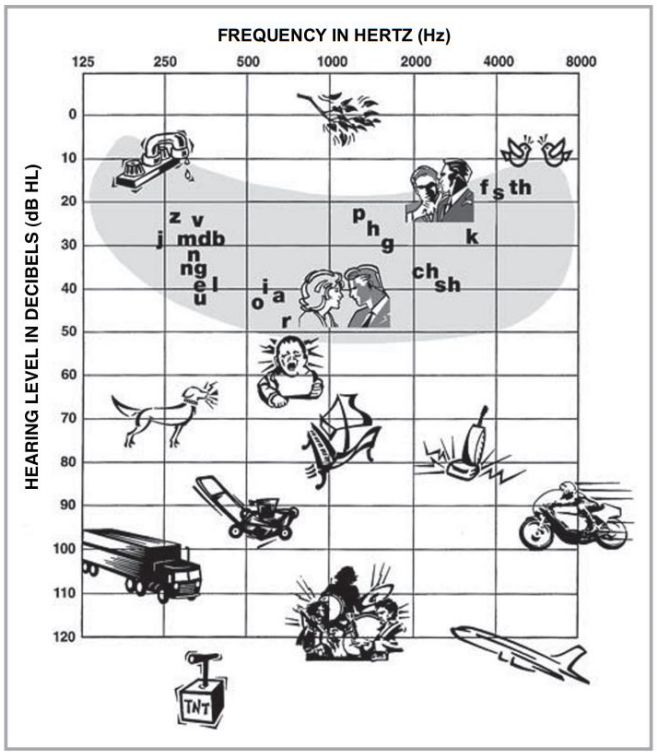
Oczywistym jest, że wytwarzanie mowy i jej percepcja, w toku adaptacji do warunków środowiska życia, musiały się rozwijać i wzajemnie stymulować. Sama fizjologia, zarówno aparatu mowy, jak i narządu słuchu, jest skomplikowana, a jej opis wykracza poza cel niniejszej rozprawy, stąd szczegóły tych zagadnień zostały pominięte. Wydaje się jednak, że kilka podstawowych aspektów ewolucyjno-fizjologicznych związanych z percepcją dźwięków w kontekście sygnału mowy jest wartych choćby wspomnienia.

Dzięki plastyczności, wysokiemu wyspecjalizowaniu i precyzji działania aparatu mowy, składającego się ze struktur kontrolowanych przez skoordynowaną aktywność neuronalną (Kent, 2004), człowiek ma zdolność do generowania nieskończonej liczby dźwięków (stąd mowa jest systemem łączności o charakterze otwartym). Na przestrzeni tysięcy lat niektóre z nich zaczęły nabierać nacechowania semantycznego, a więc stawać się nośnikami użytecznej informacji (Lieberman, 2006). Proces ten, określany mianem semantyzacji, stoi u podstaw ludzkiej komunikacji językowej i polega na przypisywaniu konkretnym grupom dźwięków

określonego znaczenia. Zdaniem wielu neurolingwistów (m.in. Pinker, 1994) to właśnie zdolność wytwarzania tak różnorodnych dźwięków oraz ich semantyczne wykorzystanie były i pozostają kluczowe dla wypracowania tak skomplikowanego i precyzyjnego systemu komunikacyjnego, mającego istotne znaczenie ewolucyjne. Dzięki mowie człowiek miał możliwość, w stopniu przewyższającym inne gatunki, zbudowania, a następnie podtrzymania społecznej organizacji oraz przekazywania wiedzy, często wykraczającej poza potrzebną do realizacji podstawowych potrzeb życiowych (Deacon, 1997).

W analizie aspektów ewolucyjno-fizjologicznych związanych z wytwarzaniem i percepcją mowy szczególnie istotne wydaje się być założenie Hocketta (1979) o jej dwukanałowości. W swoich pracach zaznacza on, że wykorzystywanie w komunikacji kanału głosowo-słuchowego stanowi podstawową cechą określającą możliwości generacji bądź odbioru komunikatów. Utrzymanie prawidłowej czynności zarówno jednego, jak i drugiego elementu oraz, co ważne, ich efektywnego współdziałania, warunkuje prawidłowe i pełne przetwarzanie informacji językowej.

Również z perspektywy psychoakustyki można wskazać kilka składowych gwarantujących sukces adaptacyjny rodzaju *Homo*. Po pierwsze zakresy progów słyszenia pokrywają się z zakresem dźwięków mowy (Stevens, 2000). W sposób bardzo uproszczony obrazuje to chętnie wykorzystywany m.in. przez logopedów tzw. „banan mowy" (pole Fanta) złożony z punktów określających poziom (oś pionowa) oraz częstotliwość (oś pozioma) fonemów a układających się w kształt właśnie tego owocu (Ross, 2004; Hillis, Uchanski i Davidson, 2023). Schemat wywodzi się z analiz prowadzonych w latach 60. przez szwedzkiego badacza mowy i jej syntezy Fanta (Fant, 1960) w firmie technologiczno-telekomunikacyjnej Ericsson. Przykładowy schemat dla języka angielskiego przedstawiono na Rys. 1 – na osi odciętych znajduje się częstotliwość (w Hz), a rzędnych poziom (w dB HL).



Rys. 1: *„Banan mowy" dla języka angielskiego* (źr.: Northern i Downs, 2014)

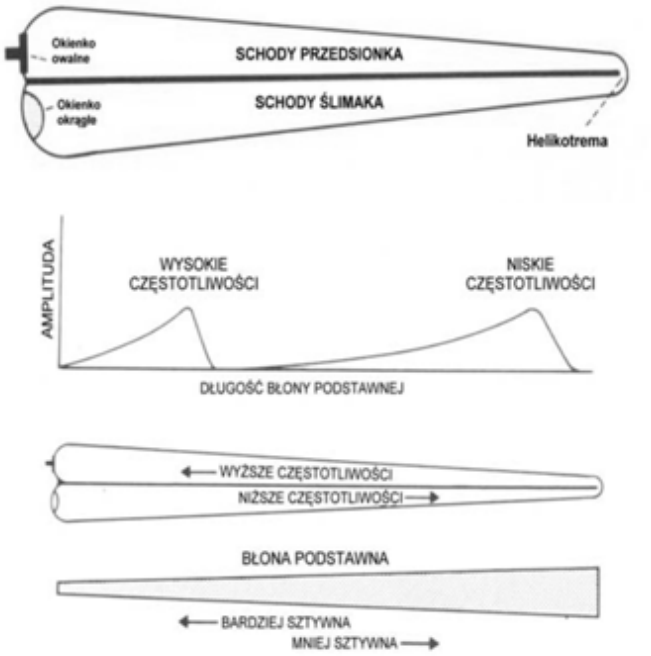
Ludzki narząd słuchu ma możliwość różnicowania dźwięków o szerokiej dynamice intensywności – od najcichszych, rzędu kilku, kilkunastu decybeli do głośnych, na granicy progu bólu. Było to istotne z punktu widzenia przetrwania ewolucyjnego (polowanie lub ucieczka przed drapieżnikiem), ale i pozwoliło na efektywną percepcję, także w niekorzystnych warunkach środowiskowych (Yost, 2007). Ze zrozumieniem mowy, nawet w hałaśliwym otoczeniu, związane jest również przetwarzanie kontekstualne (Liberman i Mattingly, 1985) polegające na zdolności ludzkiego mózgu do przetwarzania dźwięków mowy w kontekście, przy uwzględnieniu nakładania się dźwięków mowy (koartykulacja). Niemniej ciekawym zagadnieniem jest teoria sugerująca, że istnieje wrodzony, wyspecjalizowany mechanizm neurologiczny, który działa na zasadzie detektora cech fonetycznych, wrażliwego na mowę, ale nie bodźce, które nią nie są. Określa się go „modem mowy" i zgodnie z przyjętymi założeniami, przetwarzanie języka, zarówno w zakresie generacji, jak i percepcji mowy, może przebiegać dzięki niemu bardziej efektywnie (Mattingly i wsp., 1971; Remez i in., 1981).

Na percepcję mowy ma wpływ wiele czynników. Zostały one szczegółowo opisane w dalszych częściach pracy. Wydaje się jednak, że, z ewolucyjnego punktu widzenia, to składowe, które pozwalają na skuteczną ekstrakcję informacji ze zniekształconego lub zakłóconego sygnału są najważniejsze.

W ogólności można przypisać te cechy do informacji o charakterze:

- spektralnym (widmowym),
- czasowym,
- przestrzennym.

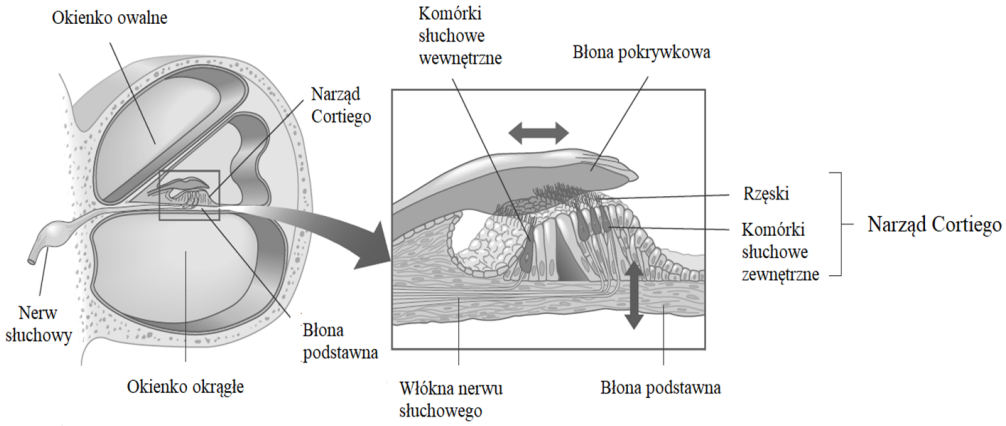
Pierwsza z nich dotyczy zawartości częstotliwościowej sygnału i jest poddawana analizie na poziomie ucha wewnętrznego. To właśnie tam, w ślimaku, określone częstotliwości powodują powstawanie drgań w odpowiadających im miejscach na błonie podstawnej. Zależność ta, określana mianem tonotopowości, została po raz pierwszy opisana przez Helmholtza w drugiej połowie XIX w., jednak dopiero badania Békésego, w których opisał propagację fali mechanicznej wzdłuż ślimaka osiągającą wartość szczytową (ang. *peak*) w miejscu zależnym od częstotliwości (Békésy, 1956 oraz 1960) pozwoliły lepiej zrozumieć zjawisko. Tonotopową strukturę ślimaka wraz z odpowiadającym kolejnym fragmentom częstotliwościom przedstawiono na Rys. 2.



Rys. 2: *Schematyczne przedstawienie tonotopowej struktury ślimaka* (źr.: Pruszewicz, 2010)

Niezwykle istotne jest to, że tonotopowość przenosi się również na wyższe piętra drogi słuchowej. Na poziomie peryferyjnym drgania błony podstawnej wywołane pobudzeniem falą akustyczną (przekształconą już na etapie odkształceń błony bębenkowej w bodziec mechaniczny) poruszają komórki słuchowe zlokalizowane

w narządzie Cortiego w ślimaku sprawiając, że zaczynają one pocierać swoimi rzęskami błonę nakrywową (Rys. 3).



Rys. 3: *Budowa narządu Cortiego* (oprac. własne na podst.: Goldstein, 2013)

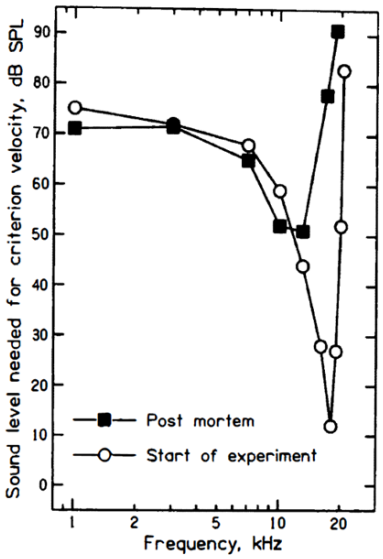
Wskutek powstałego w ten sposób podrażnienia generowane są impulsy przekazywane do znajdujących się u podstawy każdej komórki słuchowej dendrytów nerwu słuchowego. Zgodnie z „zasadą miejsca", zaproponowaną przez Helmholtza, a rozszerzoną do formy współczesnej „teorii fal wędrownych", której autorem jest G. von Bekesy[1] (1959), dźwięk o określonej częstotliwości prowadzi do odkształcenia błony podstawnej w konkretnym miejscu. Liczba pobudzanych włókien nerwu słuchowego zależy od natężenia sygnału akustycznego, a przedziały czasowe pomiędzy kolejnymi impulsami nerwowymi odpowiadają okresowi bodźca lub jego całkowitej wielokrotności. Częstotliwość fali dźwiękowej jest odzwierciedlona w częstotliwości wyładowań w nerwie słuchowym. Co ciekawe, wrażenie wysokości może powstać nawet wtedy, gdy odpowiadająca jej częstotliwość nie znajduje odwzorowania w miejscu maksymalnego pobudzenia komórek rzęskowych („teoria czasu"; Moore, 1999). Informacja o częstotliwości drgań (wysokości dźwięku) jest wtedy przenoszona do mózgu z różnych grup komórek rzęskowych w postaci czasu okresu drgania.

Podsumowując, główną rolą przetwarzania ślimakowego jest mapowanie częstotliwości docierającego sygnału akustycznego w określonych lokalizacjach błony podstawnej (częstotliwość dźwięku wywołującego maksymalne wychylenie punktu błony nazywana jest częstotliwością charakterystyczną). W takim podejściu można przyjąć, że ślimak zachowuje się jak swoisty analizator dźwięku: bodziec akustyczny

[1]Georg (György) von Békésy (1899–1972) był węgierskim lekarzem i fizjologiem, którego odkrycia miały niebagatelne znaczenie dla opracowania nowoczesnej teorii słyszenia. von Békésy pozostaje jedynym laureatem nagrody Nobla z zakresu akustyki (wyróżnienie otrzymał w 1961 r.); źr: https://psychology.fas.harvard.edu/people/georg-von-b%C3%A9k%C3%A9sy [dostęp: 30.09.2024 r.]

złożony z tonów o różnych częstotliwościach i amplitudach pobudza w różnym stopniu różne obszary błony podstawnej, a każdy z jej punktów można traktować jako filtr pasmowoprzepustowy. Właściwość ślimaka, pozwalająca na percepcyjne rozseparowanie dwóch tonów o różnej częstotliwości (usłyszenie dwóch różnych wysokości w dźwięku złożonym), nazywa się selektywnością częstotliwościową (Moore, 1999). Warto jednak pamiętać, że analiza dźwięku zachodząca na błonie podstawnej nie jest doskonała: jeśli częstotliwości dwóch tonów są nieznacznie różne to mechanizm analizy dźwięku nie pozwala na ich percepcyjne rozseparowanie i słyszymy wówczas jeden dźwięk.

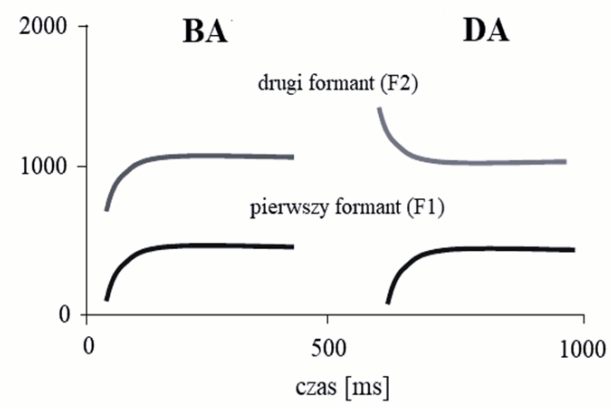
Selektywność częstotliwościową przedstawia się z reguły za pomocą krzywych strojenia (Pickles, 1988), które wyznacza się obserwując wychylenie określonego punktu błony podstawnej w odpowiedzi na pobudzenie tonem o zmiennej częstotliwości. W trakcie wyznaczania krzywych natężenie tonu dobiera się w taki sposób, aby wychylenie badanej lokalizacji zawsze miało taką samą amplitudę. Jeżeli układ słuchowych funkcjonuje właściwie, krzywe strojenia są „ostre" (stromo nachylone zbocza i ostro zarysowane minimum). Badania prowadzone na świnkach morskich (gatunek często wykorzystywany w badania dot. modelowania ludzkiego układu słyszenia) wykazały, że po śmierci zwierzęcia krzywe strojenia stają się mniej strome, co można przypisać wyłącznie mechanicznym funkcjom błony podstawnej (Rys. 4).



Rys. 4: *Zależność poziomu wejściowego (oś y) potrzebnego do wytworzenia jednakowej prędkości drgań błony podstawnej w zależności od częstotliwości stymulacji (oś x). Krzywa oznaczona kółkami uzyskana przy prawidłowym stanie fizjologicznym świnki morskiej, krzywa strojenia oznaczona czarnymi kwadratami pochodzi z pomiarów po śmierci zwierzęcia – post-mortem*
(źr.: Selick i in. 1982)

Te obserwacje, jak i wyniki eksperymentów Bekesy'ego oraz jego następców (Neely i Kim, 1986) wykazały, że selektywność częstotliwości drgań mechanicznych błony podstawnej jest mniejsza niż selektywność mierzona na podstawie odpowiedzi nerwu słuchowego. Wyjaśnianiem stała się hipoteza o „drugim filtrze", który miałby się znajdować pomiędzy wymienionymi fragmentami drogi słuchowej. Zgodnie z nią szczególna „współpraca" błony podstawnej i komórek włoskowatych tworzy złożony i precyzyjny system, w którym aktywność nerwowa może kontrolować wzór drgań błony podstawnej. Komórki nie tylko wykrywają drgania i kodują przenoszone przez nie informację do nerwu słuchowego, ale także mogą wzmacniać ruch mechaniczny błony (w szczególności dotyczy to komórek słuchowych zewnętrznych). Taka aktywna rola ślimaka i komórek włoskowatych (nazywanych wzmacniaczem ślimakowym) opiera się na nieliniowym dodatnim sprzężeniu zwrotnym, w którym bardzo słabe sygnały ulegają wzmocnieniu. Jak piszą Sęk i Wicher (2005), wzmacniacz ślimakowy ma zasadnicze znaczenie zwłaszcza w przypadku najcichszych dźwięków, bowiem to właściwie dzięki temu mechanizmowi można je w ogóle usłyszeć.

Analizując złożony sygnał jakim jest mowa, nie sposób pominąć informacji czasowej, którą z sobą przenosi. Pisząc najprościej, mowa to coś więcej niż tony. Równie istotne, co poziom i amplituda, są przebiegi czasowe. Przykład dobrze ilustrujący tę prawidłowość przedstawiono na schematycznym Rys. 5. Sylaby „ba" i „da", których wykorzystanie może warunkować zupełnie odmienną semantykę wyrazu, różnią się wyłącznie przebiegiem drugiego formantu. Niewychwycenie tej subtelnej rozbieżności na skutek niewystarczająco dokładnego śledzenia zmian w domenie czasu może prowadzić do błędnego rozpoznania wyrazu, w którym występuje któraś z ww. zgłosek.



Rys. 5: *Schematycznie przedstawiony przebieg F1 i F2 zgłosek „ba" i „da"*

Informacja czasowa, a więc również modulacje amplitudy, analizowane są na różnych

poziomach układu słuchowego aż po korę słuchową (Schreiner i wsp., 1986). Dzięki czasowej obwiedni amplitudy mowa może być rozpoznawana nawet w sytuacji niepełnej informacji spektralnej (Shannon i in., 1995). Mechanizm ten został bardziej szczegółowo opisany w dalszej części pracy.

Ostatnie, lecz nie mniej ważne informacje przestrzenne pozwalają ocenić lokalizację źródła dźwięku i bazują na fakcie posiadania przez człowieka pary uszu (np. Marrone i in. 2008; Martin i in. 2012). Porównanie informacji dochodzących do obu z nich pozwala, na podstawie różnic czasowych i natężenia, określić kierunek, z którego dochodzi sygnał akustyczny, często z bardzo dużą rozdzielczością. Tak zwane międzyuszne różnice czasu i poziomu (opisane bardziej szczegółowo w rozdziale 4.6.) pozwalają także przestrzennie rozdzielić sygnał użyteczny (z reguły mowę) od dźwięków zakłócających (m.in. Bronkhorst, 2000). Efekt ten nazywany jest przestrzennym uwolnieniem od maskowania (ang. spatial release from masking), a siła jego oddziaływania osiąga wartości rzędu kilku decybeli u osób ze słuchem prawidłowym.

Poza tonotopowością, pozwalającą na segregację częstotliwościową i analogicznymi filtrami w dziedzinie modulacji, to właśnie przestrzenne odmaskowanie jest jednym z mechanizmów najefektywniej wspierających ekstrakcję sygnału informacyjnego z zakłóceń. Wyniki badań przeprowadzonych przez autorkę pracy, a dotyczących właśnie tego zjawiska przedstawiono w rozdziale 4.6. oraz publikacji w nim opisanej.

Oprócz fizycznych mechanizmów oraz cech samego języka (redundancji, a więc pewnej nadmiarowości informacji), w procesach prowadzących do zrozumienia mowy, zwłaszcza przez oddzielenie jej od sygnałów zakłócających, zaangażowane są również procesy wyższego rzędu, takie jak np. pamięć czy koncentracja (Tchorz, 2022). O ile bowiem słyszenie jest procesem biernego odbierania dźwięku, o tyle słuchanie oznacza zwracanie uwagi na dźwięk z uwagą i zrozumieniem (Kiessling i in., 2003). Powyżej w syntetyczny sposób przedstawiono ewolucyjne aspekty związane z powstawaniem i odbiorem sygnału mowy. Jak zaznaczono, rozwój tych kluczowych elementów przyczynił się do osiągnięcia przez człowieka sukcesu ewolucyjnego.

Każdy z wytwarzanych mechanizmów ma jednak swoje naturalne ograniczenia lub może też, na przykład na skutek choroby, częściowo stracić swoją efektywność, prowadząc, w kluczowym dla nas obszarze, do zaburzeń percepcji mowy. Właśnie z tego powodu koniecznie było i pozostaje opracowywanie oraz walidacja metod pozwalających na ocenę działania zarówno poszczególnych elementów układu słuchowego, jak i jego całości. Z uwagi na to, że mowa stanowi immanentną cechę funkcjonowania człowieka w środowisku, ocena taka może być traktowana szerzej, również w obszarze chorób neurologicznych oraz systemów, które tę mowę w sposób

bezpośredni (intencjonalny) lub pośredni (nieintencjonalny) przetwarzają.

O sygnale mowy można pisać z różnych punktów widzenia – ewolucyjnego. psychofizycznego, psychologicznego, intelektualnego, semantycznego itd. Jednak ze względu na cel samej pracy, jakim jest ocena zrozumiałości mowy i scharakteryzowanie czynników mających na nią wpływ, te zagadnienia nie będą w szczegółach opisane. W kolejnych rozdziałach natomiast przedstawione zostaną kwestie, które wydają się ważne z poziomu analizy i wniosków płynących z przeprowadzonych i opisanych tutaj eksperymentów oraz ich wyników.

Struktura pracy jest następująca: pierwsza jej część wprowadza czytelnika w zagadnienia związane z problematyką dotyczącą sygnału mowy – jego fizyczne cechy, czynniki wpływające na zrozumiałość mowy, a także metody jej pomiaru. Rozważania teoretyczne uzupełniono opisem niezwykle ciekawego zjawiska jakim jest wysiłek słuchowy. W drugiej części pracy zawarto wyniki badań eksperymentalnych (E1–E11) prowadzonych przez autorkę niniejszej rozprawy, a także wskazano możliwości i kierunki dalszego rozwoju badań w tym zakresie.

Rozdział 2

Podstawowe fizyczne cechy mowy

Mowa jest sygnałem składającym się ze zmiennych w czasie przebiegów akustycznych, zarówno w dziedzinie częstotliwości, jak i natężenia. Poniżej wymieniono najważniejsze fizyczne parametry pozwalające opisać sygnał mowy, jednostki go budujące, a także scharakteryzowano najczęściej wykorzystywaną metodę jego graficznej reprezentacji – spektrogram.

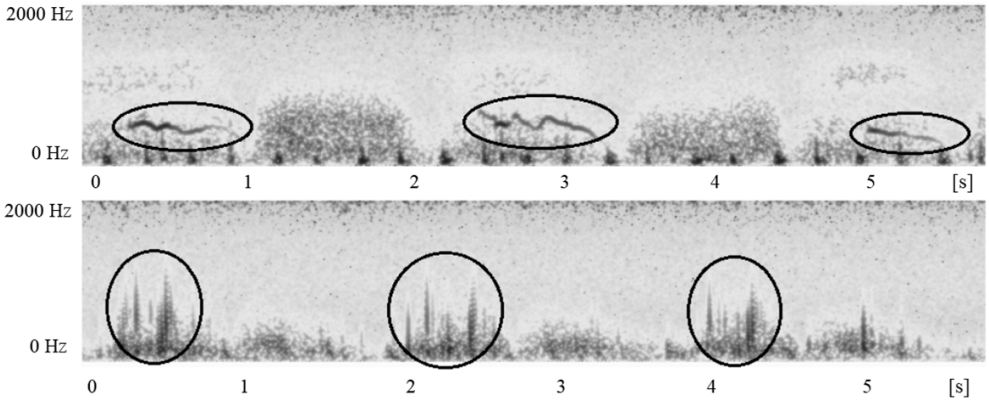
Dźwięki mowy charakteryzują się fluktuacjami amplitudy o niskiej częstotliwości (Drullman, 1994), które są krytyczne nie tylko dla zrozumiałości mowy w ciszy, ale stanowią również wskazówką pozwalającą oddzielić sygnał użyteczny od dźwięków zakłócających obecnych w tle (Festen i in., 1990). Średni zakres poziomów ciśnienia akustycznego dźwięków mowy zawiera się pomiędzy 50 a 80 dB SPL, w przypadku szeptu jest to ok. 30 dB (w odległości 1m), natomiast głosu podniesionego lub krzyku do nawet 100 dB. Widmo sygnału mowy w ogólności rozciąga się na przedział od 80 Hz do 8–10 kHz, zdarza się jednak, że istotna energia akustyczna znajduje się również w zakresach rzędu kilkunastu tysięcy herców. Górna granica tego przedziału zależy od języka, a w szczególności od składających się nań fonemów[2] – najmniejszych cząstek dźwięku, które w danym języku różnicują słowa (Moore, 1999). Może być ich kilkanaście (język hawajski), blisko 40 (w przypadku m.in. angielskiego, polskiego czy niemieckiego), a nawet ponad 60, jak w niektórych dialektach afrykańskich i kaukaskich (Goldstein, 2013). Biorąc pod uwagę możliwości ludzkiego układu słuchu jest to dość szerokie pasmo. Z punktu widzenia zrozumiałości

[2]Fonemy nie mają same w sobie żadnego znaczenia ani nie symbolizują żadnego obiektu, ale, w połączeniu z innymi, pozwalają odróżnić jedno słowo od drugiego. Kompleksową definicję fonemu przedstawia w swojej książce „Nieobecna struktura" ceniony semiotyk i filozof Umberto Eco: „Fonem jest najmniejszą jednostką posiadającą dystynktywne cechy dźwiękowe; jego wartość określa jego pozycja i jego odmienność od innych elementów. Fonem może wykazywać warianty fakultatywne, odmienne u różnych mówiących, które jednak nie zacierają różnic decydujących o znaczeniu. System fonemów jest systemem różnic, który może być analogiczny w różnych językach, nawet jeśli wartości fonetyczne (fizyczna jakość dźwięków) są odmienne".

mowy najbardziej istotny jest jednak zakres 200–5600 Hz (O'Shaughnessy, 2000). Odfiltrowanie częstotliwości leżących poza tym pasmem nie wpływa w sposób znaczący na jej zrozumiałość. Potwierdzają to wyniki badań Fletchera (1950) wskazujące na to, że połowa energii akustycznej mowy zawarta jest w paśmie 100–350 Hz, natomiast druga od 350 do 10 000 Hz. Przy zawężeniu odbioru do zakresu 350–10 000 Hz (odfiltrowanie najniższych częstotliwości) nastąpi utrata około 50% głośności i jedynie 2% wyrazistości (procent prawidłowo powtórzonych elementów testu odsłuchowego). W przypadku ograniczenia pasma mowy do najniższego zakresu (<350 Hz) podobnie traci się 50% głośności, ale mowa staje się niezrozumiała (98% utraty wyrazistości; Demenko, 2011), co częściowo można zaobserwować także na schematycznym rozkładzie częstotliwości dźwięków mowy przedstawionym na Rys. 1.

Mając na względzie to, że mowa stanowi bodziec akustyczny, który zmienia się wielowymiarowo (Jorasz, 1998), przedstawienie jej wyłącznie w postaci chwilowych wahań ciśnienia akustycznego (funkcji obrazującej zmiany ciśnienia akustycznego w czasie) albo widma amplitudowo-częstotliwościowego nie jest działaniem wystarczającym. Aby lepiej opisać zmiany energii w sygnale powinno się równocześnie przedstawiać jego częstotliwość oraz czas, a narzędziem pozwalającym na taki zabieg jest spektrogram. Sygnał dzieli się na krótkie odcinki czasowe i wyznacza ich widma chwilowe. Czas odłożony jest na osi odciętych (x), częstotliwość na osi rzędnych (y), a stopień zaczernienia (lub nasycenie kolorów) odzwierciedla poziom natężenia dźwięku. Ten rodzaj graficznej reprezentacji bardzo często wykorzystywany jest w sejsmologii, biologii (zwłaszcza w zakresie werbalnej komunikacji między zwierzętami, w tym rozpoznawania gatunków ptaków) oraz analizie dźwięków muzycznych. Innym ciekawym obszarem zastosowania jest diagnostyka i monitorowanie przebiegu chorób układu oddechowego.W jednym ze swoich badań autorka niniejszej pracy wraz z zespołem współpracowników dokonali oceny użyteczności spektrogramów w detekcji oraz właściwej klasyfikacji nieprawidłowych dźwięków osłuchowych. Eksperyment, w którym lekarze oraz studenci medycyny opisywali nagrania osłuchowe zaprezentowane w 3 konfiguracjach (próbki audio, audio ze spektrogramem oraz sam spektrogram) wykazał, iż uzupełnienie klasycznego badania osłuchowego jego wizualizacją w postaci spektrogramu może wspomóc decyzje diagnostyczne – czułość (*sensitivity*) w wykrywaniu fenomenów osłuchowych po dodaniu do próbki dźwiękowej spektrogramu wzrasta o 4 p.p. w przypadku lekarzy (z 79% do 83%) oraz 2 p.p. wśród studentów medycyny (z 73% do 75%) nawet bez uprzedniego treningu. Poczynione obserwacje pozwalają uznać spektrogram za narzędzie z ogromnym potencjałem do wspierania diagnostyki i monitorowania pacjentów z chorobami

układu oddechowego, a także edukacji adeptów medycyny (Pastusiak i in., 2024). Poniżej przedstawiono przykładowe spektrogramy nagrań zawierających nieprawidłowe dźwięki oddechowe – powstające na skutek nadmiernego zwężenia (obturacji) dróg oddechowych świsty (w tym przypadku wydechowe; panel górny) oraz nagłego otwierania się wcześniej zamkniętych małych dróg oddechowych – rzężenia[3].

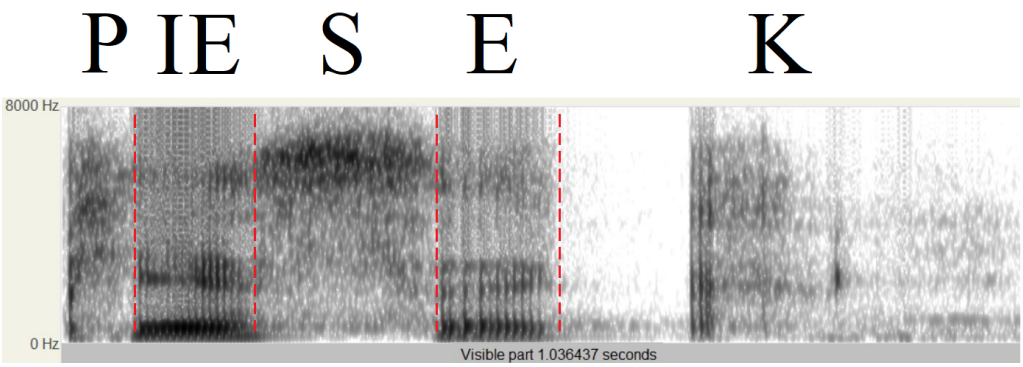


Rys. 6: *Przykładowe spektrogramy nagrań zawierających nieprawidłowe dźwięki oddechowe – świsty wydechowe (panel górny) oraz rzężenia wdechowe (panel dolny)*

Zaprezentowany przykład stanowi dobre, choć nieco uproszczone, zobrazowanie sposobu wykrywania i wstępnej analizy na najniższych poziomach układu słuchowego. Reprezentacja graficzna dźwięków oddechowych pokazuje, na jakiej podstawie można dokonać detekcji sygnału użytecznego w dziedzinie czasu (w przypadku świstów) i częstotliwości (rzężenia), gdy występują one w towarzyszeniu zakłóceń. Nie można zapomnieć, że istotą niniejszej pracy jest jednak analiza mowy, która jest o wiele bardziej skomplikowanym sygnałem i poza samą detekcją w obu dziedzinach wymaga także odpowiedniej interpretacji na wyższych piętrach układu słuchowego, w obszarach mózgowych za to odpowiedzialnych. Mówiąc o spektrogramie trzeba mieć na uwadze to, że nie jest możliwe jednoczesne osiągnięcie wysokiej rozdzielczości w dziedzinach czasu i częstotliwości. Utrzymanie wysokiej rozdzielczości częstotliwości odbywa się kosztem zmniejszenia rozdzielczości w dziedzinie czasu i *vice versa*. W zależności od celu badania cech akustycznych sygnału wybiera się pasma analizujące o różnej szerokości. Tzw. „szerokopasmowe" spektrogramy (o szerokości pasma rzędu 100 Hz) pozwalają na precyzyjną ocenę przebiegu czasowego i kształtu

[3]Świsty to dźwięki ciągłe, wysokoczęstotliwościowe, powstające wskutek turbulentnego przepływu powietrza przez zwężone drogi oddechowe typowe dla przebiegu m.in. astmy.
Rzężenia (w tym przypadku drobnobańkowe) to impulsowe dźwięki przypominające trzaski wywoływane przez nagłe wyrównanie ciśnienia gazów pomiędzy dwoma obszarami płucnymi, występujące np. w zapaleniu oskrzeli.

formantów. Z kolei spektrogramy „wąskopasmowe” (szerokość pasma zazwyczaj 45 Hz) cechują się gorszą rozdzielczością czasową, ale umożliwiają dokładniejszą niż w przypadku szerokopasmowych obserwację zmienności poszczególnych składowych harmonicznych (Moore, 2003). Przykładowy spektrogram (zakres 0–8000 Hz, długość okna 0.005 s, zakres dynamiki 70 dB) stanowiący graficzną reprezentację wypowiedzi „piesek” przedstawiono na Rys. 7.



Rys. 7: *Spektrogram graficznie reprezentujący wypowiedź „piesek”*

Początek nagrania to plozja (ciemna, pionowa linia) inicjująca artykulację głoski „p”. W kolejnym fragmencie wyraźnie zarysowany jest ton krtaniowy (poprzeczna, czarna linia w zakresie niskich częstotliwości) związany z dźwięcznością samogłosek „i’ oraz „e”. Podobnie wygląda część spektrogramu ilustrująca drugą głoskę „e” występującą w połowie nagrania. „S” to pasmo szumu zobrazowane jako zagęszczenie energii w zakresie wysokich częstotliwości przekraczających 5000 Hz. Ostatni fragment obejmuje fazę zwarcia (interwał ciszy), po którym następuje plozja zawierająca energię w szerokim zakresie częstotliwości (szerokopasmowy impuls). Wypowiedź kończy segment szumowy (tzw. afrykacja).

Spektrogram może być wystarczającym narzędziem służącym do identyfikacji samogłosek, co odbywa się na podstawie określania wartości kolejnych formantów[4] (lokalnych maksimów energii). Do identyfikacji wszystkich, funkcjonujących w języku polskim (z wyjątkiem „y”) wystarczy znajomość pierwszego i drugiego formantu – F1 i F2 (Jassem, 1973). Z tego względu spektrogram jest używany jako podstawowy i wystarczający element modelowania percepcji mowy w przypadku tych elementów mowy.

[4]Dźwięki mowy wytwarzane są w tzw. aparacie głosowym, na który składają się płuca, tchawica, krtań (w której znajdują się więzadła głosowe) oraz kanał głosowy – gardło, nos, jama nosowa oraz usta (Moore, 1999). Elementy znajdujące się powyżej krtani stanowią układ filtrów o określonych częstotliwościach rezonansowych (określanych mianem formantów). Częstotliwość środkowa każdego z formantów jest inna (rozpoczynając od najniższej częstotliwości – F1) i jest bezpośrednio związana z kształtem kanału głosowego.

Myśląc o rozpoznawaniu elementów mowy nie sposób pominąć transformacji cepstralnej. Cepstrum (nazwa stanowi anagram ang. *spectrum* – widmo) to odwrotna transformata Fouriera widma sygnału po raz pierwszy opisana przez Bogerta w 1963 roku. Przekształcenie to polega na obliczeniu logarytmicznej reprezentacji widma mocy sygnału, która może pomóc w wyizolowaniu cech charakterystycznych dźwięków mowy. W przypadku rozpoznawania spółgłosek, oprócz oceny obecności tonu krtaniowego (z reguły łatwo identyfikowalnego na spektrogramie, jak na Rys. 7), to właśnie cepstrum może pomóc w identyfikacji cech akustycznych.

Analizując fizyczne cechy mowy warto wziąć pod uwagę jeszcze jeden aspekt. W trakcie wypowiadania określonych wyrazów lub zdań nadawana jest im pewnego rodzaju melodia, wynikająca z modulacji głosu, co określa się mianem intonacji. Z reguły zmiana intonacji ma na celu zaznaczenie pewnych fragmentów wypowiedzi i sprzyja wyrażaniu emocji. Istnieją jednak języki, w których pełni ona również funkcję dystynktywną, wpływającą na znaczenie użytych wyrazów. Taka sytuacja ma miejsce chociażby w przypadku języka tajskiego – wyrazy o identycznym zapisie graficznym zmieniają swoją semantykę w zależności od modulacji głosu.

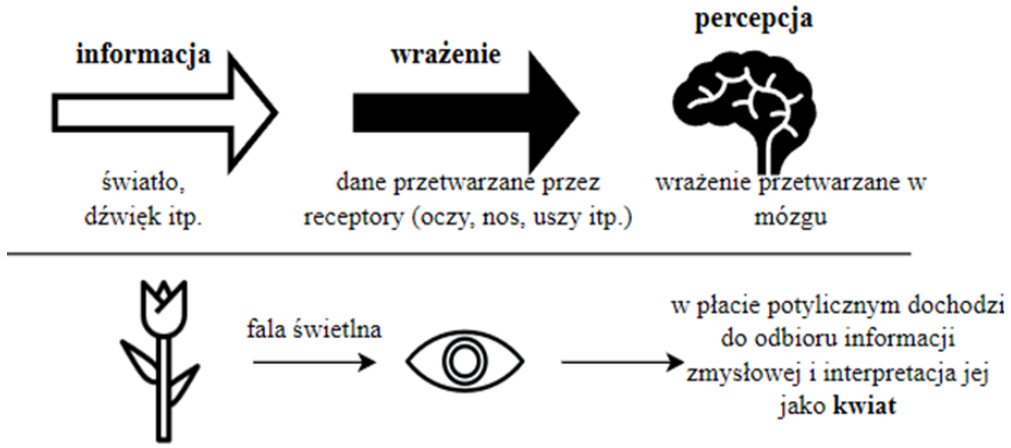
Rozdział 3

Podstawowe aspekty percepcji mowy

U podstaw funkcjonowania ludzi w szeroko rozumianym środowisku leży odbiór jego zmiany (bodźca), odpowiednie rozpoznanie i klasyfikacja. Dopiero po przejściu przez każdy z wymienionych etapów i przy zachowaniu wysokiej efektywności aparatu poznawczego możliwe jest właściwe zinterpretowanie docierających pobudzeń oraz właściwa reakcja nań. Pojęciem, które spaja powyższe procesy jest „percepcja". W poniższym rozdziale zostało ono pokrótce zdefiniowane ze szczególnym uwzględnieniem domeny akustycznej. Część 3.1. zawiera szereg dowodów potwierdzających złożoność percepcji i jej labilną naturę pozwalającą na równoczesne korzystanie z informacji kierowanych do różnych modalności – w tym konkretnym przypadku, wzroku i słuchu.

Aby zrozumieć na czym polega proces słyszenia koniecznie jest własciwie sprecyzowanie pojęcia „dźwięk". Jest to niezwykle istotne, a dylemat ten świetnie obrazuje przykład z pogranicza psychoakustyki i filozofii: *Czy jeśli w lesie przewróci się drzewo, ale nie będzie nikogo, kto to usłyszy, to czy można mówić o dźwięku?* Goldstein (2013) proponuje rozdzielenie pojęcia na dwie równolegle funkcjonujące idee. Po pierwsze, dźwięk można rozumieć jako pobudzenie fizyczne. Zgodnie z definicją (m.in. Makarewicz, 2023), falą akustyczną określa się przeniesienie zaburzenia ośrodka, a więc zmiany ciśnienia akustycznego. Można je opisać w dwóch głównych domenach – częstotliwości i amplitudy, co zostało wyjaśnione w rozdziale 2. O wrażeniu dźwięku (słyszeniu) mówi się jednak dopiero, i jest to druga ścieżka w koncepcji Goldsteina, w sytuacji odbioru informacji przenoszonej przez falę akustyczną. Mogą być one wykorzystane po właściwej identyfikacji i przetworzeniu na kolejnych etapach szeroko rozumianej drogi słuchowej. Myśląc o tym, jak sygnał akustyczny jest odbierany i przekazywany od części peryferyjnej do ośrodków centralnych konieczne jest prawidłowe zdefiniowanie podstawowych pojęć. Są one tożsame dla wszystkich procesów poznawczych, również dotyczących innych

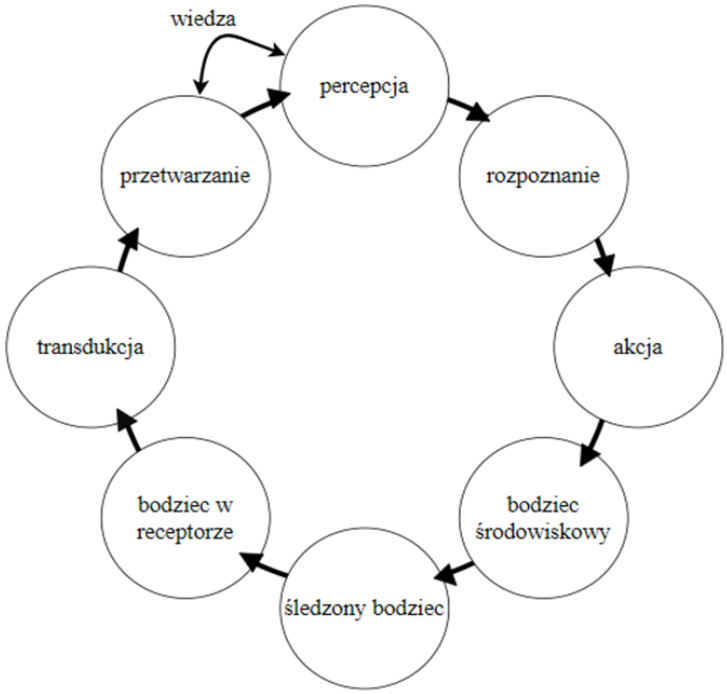
modalności. Szczególnie dobrze opisane są zagadnienia dotyczące zmysłu wzroku. Na początku procesu percepcyjnego zawsze pojawia się bodziec, który można rozumieć jako fizyczną zmianę zachodzącą w środowisku. Jeśli jej wielkość przekracza progi pobudliwości receptorów dochodzi do jej odbioru, a następnie przetwarzania do formy, która może poddana być interpretacji. W przypadku wzroku sygnałem pobudzającym jej fala elektromagnetyczna w postaci światła. Po jego absorpcji, pigmenty w białkach fotoreceptorów znajdujących się na siatkówce albo zmieniają swoją konformację, albo ulegają fotoredukcji. Uruchamia to kaskadę transdukcji sygnału – dochodzi do zmian w natężeniu prądu naładowanych cząstek przez błonę komórki fotoreceptora i dalej do zmiany jego potencjału, który jest przekazywany w postacie pobudzenia nerowego do ośrodków centralnych (Rys. 8).



Rys. 8: *Wrażenie i percepcja na przykładzie wzroku* (oprac. własne na podst.: Goldstein, 2013)

Dość podobnie procesy te wyglądają w przypadku zmysłu słuchu, również w odniesieniu do sygnału mowy. Na pierwszą, recepcyjną część procesu percepcji mowy przypadają aspekty dotyczące detekcji sygnału akustycznego oraz jego transmisji w obrębie peryferyjnego układu słuchowego, na który składają się ucho zewnętrzne, środkowe i wewnętrzne. Podobnie jak w przypadku innych sygnałów z zakresu słyszalnego przez człowieka, energia akustyczna sygnału mowy przenoszona jest do receptora słuchowego na drodze przewodnictwa powietrznego i kostnego, spośród których najistotniejszą rolę odgrywa pierwszy z nich. W kolejnych krokach dochodzi do konwersji bodźca mechanicznego w ciąg impulsów nerwowych przechowujących dane dotyczące częstotliwości czy natężenia (amplitudy) sygnału (przetwarzanie neurofizjologiczne). Przekazana do ośrodków centralnych informacja zostaje przetworzona (rozkodowana) i właściwie rozpoznana. Procesy te zachodzą przy wykorzystaniu wiedzy i porównaniu z przechowywanymi w pamięci wzorcami. Dopiero na tym etapie można mówić o świadomej percepcji, która prowadzi do określonego zachowania. Prostym przykładem realizującym

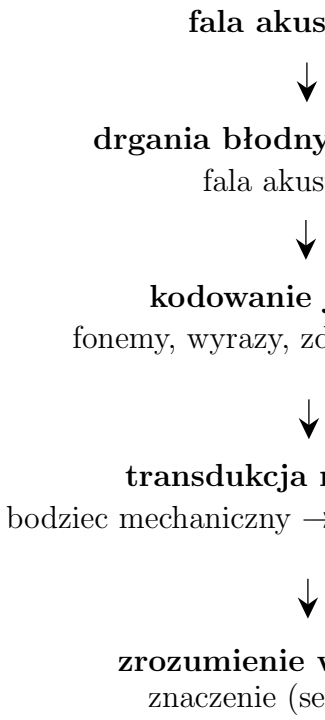
ten ciąg zależności (przedstawionych również na Rys. 9) może być np. drgający metalowy dzwonek lub syrena generująca propagującą się falę akustyczną. Natężenie bodźca przekracza wartość progową, dlatego może być odebrane przez układ recepcyjny, a na poziomie ślimaka przekształcony w pobudzenie elektryczne. Po dotarciu do ośrodków mózgowych dochodzi do klasyfikacji bodźca. Na podstawie wiedzy i wcześniej zdobytych doświadczeń interpretujemy go jako dźwięk sygnału alarmowego i możemy podjąć odpowiednie działanie (ucieczkę z pomieszczenia).



Rys. 9: *Schematyczny przebieg procesu percepcyjnego (oprac. własne na podst.: Goldstein, 2013)*

W przypadku mowy i rozpoznawania elementów językowych wymienione procesy wymagają jeszcze większej precyzji i udziału zasobów poznawczych. Dzięki aktywacji mózgowych odpowiedników fonologicznych i leksykalnych bodźców możliwa jest ekstrakcja dystynktywnych (różnicujących fonemy), segmentalnych oraz supresegmentalnych (akcent, intonacja i iloczas) cech fonetycznych. W tym kontekście szczególności istotne jest przywołanie ośrodka Brocki związanego z oddziaływaniem krótkotrwałej pamięci roboczej oraz ośrodka Wernickego odgrywającego znaczącą rolę w kojarzeniu formy określonego słowa z jego znaczeniem na zasadzie porównania docierającej jednostki językowej z wzorcami przechowywanym w osobniczym katalogu (leksykonie). W ogólności można więc powiedzieć, że prawidłowe zrozumienie komunikatu słownego wymaga odwzorowania (zmapowania) dynamicznego, zmiennego w dziedzinie czasu, częstotliwości i natężenia ciągłego sygnału akustycznego na postać dyskretnych reprezentacji językowych w mózgu (Davis i Johnsrude, 2007). Procesy te przebiegają

w nieustannej komparacji z wzorcami przechowywanymi w pamięci długotrwałej.
Pierwszych wskazówek dotyczących tego w jaki sposób mózg, albo, szerzej, układ nerwowy, są zaangażowane w percepcję informacji językowej dostarczyli badaczom pacjenci mający problem z mową – jej wytwarzaniem lub rozumieniem. Już starożytni Grecy opisywali przypadki utraty zdolności mówienia, określonego przez nich mianem „afazji". Jednak dopiero badania prowadzone pod kierownictwem Marca Daxa w pierwszej połowie XIX wieku pozwoliły powiązać utratę możliwości artykulacji z uszkodzeniem lewej półkuli mózgu (Manning i in., 2010). Ćwierć wieku później związek ten znalazł potwierdzenie w pracach Broki (Dronkers i wsp., 2007). Analizując post-mortem strukturę mózgu jednego ze swoich pacjentów dotkniętych afazją zaobserwował uszkodzenie lewej części kory czołowej. Od nazwiska badacza zaczęto nazywać tę część mózgu obszarem Broki. On sam ściśle wiązał funkcje językowe właśnie z tą lokalizacją twierdząc: „mówimy lewą półkulą" (Kandel i in., 2012). Nie minęły dwie dekady gdy Wernicke, niemiecki psychiatra i neurolog, odkrył, iż problemy językowe mogą być spowodowane uszkodzeniami również innej części mózgu, znajdującej się w tylnej części płata skroniowego (nazywanej później obszarem Wernickego). Jest ona połączona z korą czołową wiązką włókien nerwowych (tzw. powięzią łukowatą), której uszkodzenie skutkuje wystąpieniem afazji przewodzeniowej – osoby nią dotknięte rozumieją komunikaty językowe, natomiast nie są w stanie powtarzać słów, a ich artykulacja jest pozbawiona sensu. Wernicke dokonał jeszcze jednego istotnego spostrzeżenia – posługiwanie się językiem, podobnie jak inne złożone funkcje psychiczne nie jest zlokalizowane w pojedynczym, ściśle zlokalizowanym obszarze mózgu (Tremblay i in., 2016). Wręcz przeciwnie – do pracy angażowane są liczne rejony wzajemnie ze sobą wiązane w postaci sieci. Odbiór mowy odbywa się oddzielnie od jej wytwarzania, ale domeny te nie są zupełnie od siebie niezależne.
Dziś wiadomo, że połączenia mózgowe związane z językiem są dużo bardziej skomplikowane niż zakładali Broca i Wernicke. Niezaprzeczalnie jednak ich ustalenia, a zwłaszcza wyniki badań, które pozwoliły zlokalizować najistotniejsze dla wytwarzania i odbioru języka struktury, stanowią podstawę współczesnego spojrzenia na neurologiczne aspekty mowy.
Istnieje szereg czynników mogących mieć wpływ na to, jak efektywnie przebiega percepcja dźwięków mowy, ich przetwarzanie oraz, finalnie, interpretacja przedstawione schematycznie na Rys 10.



Rys. 10: *Kolejne etapy percepcji słuchowej prowadzące do zrozumienia sygnału mowy*

Wybrane z nich (słyszenie binauralne, interakcja wzrokowo-słuchowa, ubytek słuchu, warunki odsłuchu, choroby towarzyszące) przedstawiono w kolejnych podrozdziałach części 4. niniejszej rozprawy.

3.1 Interakcja wzrokowo-słuchowa

W percepcji dźwięków mowy istotną rolę pełnią inne aspekty pozasłuchowe np. ruchy twarzy i warg (Šabic i in., 2020). Wymagające akustycznie warunki (obecność sygnałów zakłócających, nakładające się głosy rozmówców, pogłos itp.) często wymuszają mimowolne spoglądanie słuchaczy na usta osoby mówiącej, a próba odczytania komunikatu z ruchu warg ma być niejako kompensacją trudności w akustycznej percepcji mowy (Sumby i Pollack, 1954; Aparicio i wsp., 2017). Szereg badań (m.in. Liberman, 2007) dowodzi istnieniu sprzężenia pomiędzy informacją akustyczną i wizualną docierającą do mózgu. W tym kontekście książkowym przykładem wydaje się być tzw. efekt McGurka, który doskonale ilustruje tę zależność. Gdy wskazówki wzrokowe i słuchowe docierające do odbiorcy są sprzeczne, mózg podejmuje próbę połączenia rozbieżne informacji w jedną, tworząc w ten sposób iluzję. W klasycznej formie eksperymentu, badani zgłaszają słyszenie zgłoski „ta", gdy na materiał wizualny prezentujący artykulację sylaby „ka" zostanie nałożona ścieżka dźwiękowa zawierająca sylabę „pa". Słuchacze

nie opierają swojego wrażenia zmysłowego tylko na tym, co usłyszeli lub zobaczyli, ale raczej, podświadomie, na bazie swoistej fuzji percepcyjnej bazującej na danych pochodzących z dwóch modalności (McGurk i MacDonald, 1976). Innym zjawiskiem psychoakustycznym wykorzystującym wzajemne uzupełnianie się informacji multimodalnej jest tzw. „efekt brzuchomówcy" (ang. *ventriloquism effect*). Występuje on, gdy bodźce wzrokowe i słuchowe prezentowane są równocześnie, ale każdy z nich ma inną lokalizację. Wrażenie percepcyjne wskazuje natomiast wyłącznie na kierunek, z którego pochodzi informacja wizualna, z pominięciem lokalizacji źródła wskazówki akustycznej (Woods i Recanzone 2004). Zarówno efekt McGurka, jak i brzuchomówcy stanowią niezaprzeczalne dowody na dominację bodźców wzrokowych nad słuchowymi w sytuacji jednoczesnej prezentacji.

Istnieją jednak eksperymenty, w których obserwuje się odwrotną zależność. Przykładem jest *double-flash illusion*. Badanym prezentowany jest pojedynczy, krótki i intensywny bodziec świetlny (błysk, z ang. *flash*), któremu towarzyszą dwa impulsowe sygnały tonalne. Wyniki pomiarów przeprowadzonych m.in. przez Shamsa i wsp. (2000) wskazują, że w takich przypadkach uczestnicy pomiaru deklarują percepcję dwóch błysków – pobudzenie akustyczne modyfikuje zatem odbiór informacji wizualnej.

Bez względu na kierunek zależności, przytoczone zjawiska, a także wiele innych eksperymentów wskazuje na nierozerwalne powiązanie zmysłu wzroku i słuchu oraz wzajemny wpływ docierających do nich informacji (Cytowic i in., 2009).

Naturalnym wydaje się być postawienie pytania, jaki jest mechanizm fizjologiczny takiego oddziaływania. Odpowiedź na nie stała się możliwa dzięki rozwojowi narzędzi pozwalających na neuroobrazowanie funkcjonalne, tzn. obserwację nie tylko struktur anatomicznych, ale ich aktywność w trakcie realizacji konkretnych czynności. Okazuje się, że wielozmysłowa natura ludzkiego doświadczenia poznawczego jest odzwierciedlona we wzajemnych połączeniach różnych obszarów sensorycznych mózgu.

Olbrzymim wyzwaniem komunikacyjnym, które postawiła przed społeczeństwem sytuacja epidemiczna związana z globalnym rozprzestrzenianiem się SARS–CoV-2 (zwłaszcza w latach 2020–2021) było powszechne noszenie maseczek zakrywających usta i nos, mających zapobiegać transmisji patogenów. Szczególnie rozpowszechnione były maseczki chirurgiczne, zwykle wykonane z trzywarstwowej włókniny polipropylenowej (Chua i in., 2020) oraz maseczki typu KN95 (maseczki filtrujące, spełniające międzynarodowy standard jakości w zakresie skuteczności ochrony przed małymi cząstkami pyłu, dymu oraz bakterii i wirusów o wielkości do 0.6 µm; EN 149:2001+A1:2009). O ile te środki ochrony osobistej, prawidłowo

używane spełniały swoje podstawowe zadanie (m.in. Tabatabaeizadeh, 2021), korzystanie z nich mogło być istotną przeszkodą w werbalnej wymianie informacji (Bottalico i in., 2020; Hampton i in., 2020; Saunders i in., 2021). To zjawisko opisywane było w literaturze już wcześniej (m.in. w kontekście komunikacji chirurgów na sali operacyjnej, m.in. Atcherson i in., 2017), jednak znacznie większy wydźwięk zyskało właśnie w czasie pandemii. Na zagadnienie to składają się dwa aspekty – fizyczny (związany z charakterystyką materiałową) oraz wizualny (związany z percepcją wzrokową). Po pierwsze, tworzywa, z których wykonane są maseczki stanowią fizyczną barierę dla propagującej się z ust mówcy fali akustycznej. Badania Palmiero i zespołu (2016) wskazują, że szacowana wartość redukcji transmisji mowy wynosi 3–4%. Ocenia się, że maseczki działają na zasadzie filtrów dolnoprzepustowych z częstotliwością odcięcia ok. 1000–2000 Hz (Palmiero i in., 2016; Corey i in., 2020). Tezę tę potwierdzają m.in. wyniki uzyskane przez Duy Duong Nguyen i in. (2021) – poziomy spektralne w zakresie wyższych badanych częstotliwości (tj. z przedziału 1–8 kHz) zostają wytłumione o 2 dB w maseczkach chirurgicznych, a w KN95 – 5.2 dB.

Rozdział 4

Czynniki zewnętrzne i wewnętrzne wpływające na percepcję mowy

Francuskiemu filozofowi, Pascalowi Quignardowi, przypisuje się stwierdzenie „ucho nie ma powieki", które pojawiło się w jednej z jego prac przy stawianiu w opozycji zmysłu wzroku i słuchu (Quignard, 2016). Chociaż z punktu widzenia nauki o percepcji takie przedstawienie zagadnienia wydaje się być uproszczeniem („To, co widziane, może zostać zniesione przez powieki, może zostać zatrzymane przez przegrody lub zasłony, może zostać natychmiast zmienione na niedostępne przez ściany. To, co słyszalne, nie zna ani powiek, ani przegród, ani zasłon, ani ścian. Niedefiniowalne, nie sposób się przed nim uchronić (...) Dźwięk wdziera się. Narusza". – tłum. własne autorki), nie sposób odmówić autorowi chociaż częściowej racji. Doskonale przedstawia to w jednym ze swoich tekstów Springer: „Gdy nie chcemy na coś patrzeć, to po prostu nie patrzymy. Kierujemy wzrok gdzie indziej, odwracamy głowę i to, co nas wizualnie rozpraszało – znika. Z uszami tak się nie da. Trzeba by użyć zatyczek albo ochronnych nauszników, ale na dłuższą metę tak funkcjonować nie można. Musimy słyszeć i w dodatku słyszymy wszystko". Żyjemy w świecie przepełnionym dźwiękiem i to on niejako kształtuje naszą rzeczywistość. Może być wytworem natury, metodą przekazywania informacji, ale również sygnałem niepożądanym, który tę komunikację zaburza. To, w jak dużym stopniu się to odbywa zależy od szeregu czynników. Wiele z nich nie podlega ocenie ilościowej. Pojęcia „dokuczliwości" czy „wysiłku słuchowego", choć ubrane w coraz bardziej wiarygodne (zobiektywizowane) i dokładniej walidowane skale i kwestionariusze, nadal dotyczą zjawisk, które, z uwagi na subiektywny z natury rzeczy charakter, nigdy nie będą precyzyjnie mierzalne. Nie zmienia to jednak faktu, iż, chcąc lepiej poznać procesy percepcyjne odpowiadające tym zachodzącym w rzeczywistym środowisku życia, obecność sygnałów zakłócających w badaniach układu słuchowego jako całości,

jak i konkretnych jego części powinna stać się globalnym standardem, zarówno w zastosowaniach klinicznych, jak i podejściu psychoakustycznym. Ta kwestia zostanie szerzej opisana w rozdziale poniżej. W kolejnych natomiast, przedstawione zostaną inne aspekty wpływające na efektywność procesów słyszenia, ze szczególnym uwzględnieniem sygnału mowy.

Słyszenie jest procesem niezwykle złożonym, zwłaszcza jeśli analizowany jest tak skomplikowany sygnał, jak mowa. Aby mogło zachodzić prawidłowo niezbędne jest spełnienie 3 podstawowych wymogów:

- właściwej detekcji (ang. *audibility*),
- właściwego rozkodowania (ang. *encoding*),
- właściwego przetworzenia (ang. *processing*).

Detekcja przebiega w sposób stosunkowo prosty. Aby powstało wrażenie, działający bodziec (w tym przypadku fala akustyczna) musi przekroczyć wartość progową. Zdarza się, że ze względu na ubytek słuchu, postrzeganie sygnału jest ograniczone – rozwiązaniem w tej sytuacji powinno być odpowiednie leczenie (głównie w przewodzeniowych ubytkach słuchu) lub kompensacja niedosłuchu z wykorzystaniem aparatów słuchowych, urządzeń wspomagających słyszenie lub implantów ślimakowych. Czynnikami utrudniającymi właściwą detekcję są również warunki propagacji fali – akustyka pomieszczenia, obecność zakłóceń, które mogą maskować sygnał mowy. Maskowaniem określa się proces, w wyniku którego dochodzi do podwyższenia progu słyszalności jednego dźwięku na skutek obecności innego dźwięku. Najważniejsze informacje dotyczące tego etapu percepcji przedstawiono we wstępie rozprawy. W rozkodowaniu i przetworzeniu informacji słuchowej istotną rolę pełnią wyższe piętra drogi słuchowej, w tym ośrodki centralne. Efektywność tych procesów zależy od wielu czynników, spośród których wymienić można wiek, zasobność indywidualnego leksykonu, obecność (lub brak) zaburzeń związanych z neurodegeneracją, jak i wiele innych. Najważniejsze opisano dokładniej poniżej.

4.1 Obecność sygnałów zakłócających

Do czynników wpływających na percepcję mowy, jak i wyniki uzyskane w pomiarach zrozumiałości należy również sposób prezentacji sygnału użytecznego. Może się ona

odbywać w ciszy lub w obecności maskera. Aby jak najdokładniej zasymulować warunki codziennej komunikacji, a tym samym zbadać rzeczywistą zrozumiałość mowy, zarówno w przypadku diagnostyki, monitorowania słuchu, jak i oceny zysku z aparatów słuchowych, audiometria mowy powinna być wykonywana również z towarzyszeniem szumu. Dzięki temu możliwe jest zebranie informacji związanych nie tylko z komponentem odpowiedzialnym za peryferyjne aspekty słyszenia czy utratę wzmocnienia, ale również podprogowego przetwarzania słuchowego uwzględnieniem indywidualnych czynników.

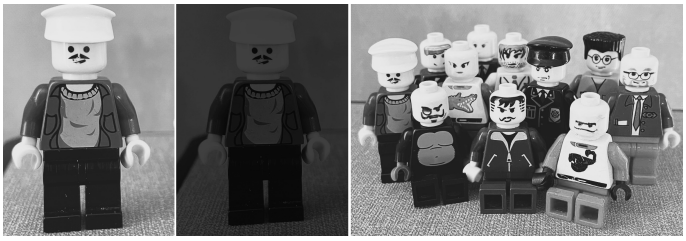
W zależności od tego, jaki aspekt przetwarzania, a finalnie – zrozumienia mowy ma być poddany analizie, w badaniach psychoakustycznych wykorzystać można różnego rodzaju sygnały maskujące. Uzyskane rezultaty uwarunkowane będą ich cechami akustycznymi oraz percepcyjnym wykorzystaniem (lub nie) ich obecności. Osoby z niedosłuchem z reguły potrzebują wyższego stosunku sygnału do szumu, aby osiągnąć ten sam poziom „wydajności" słuchowej (określonej przez stopień zrozumiałości) w porównaniu do prawidłowo słyszących. Różnice pomiędzy uzyskiwanymi przez te dwie grupy wynikami różnią się znacznie w zależności od charakterystyki sygnału zakłócającego. Gdy masker jest stałym w czasie szumem o tym samym długoterminowym średnim widmie, mieści się ona z reguły w zakresie 2–5 dB (Glasberg i Moore, 1989). W przypadku maskowania sygnałem o charakterystyce pojedynczego mówcy (Carhart i Tillman, 1970), mową odwróconą w dziedzinie czasu lub szumem zmodulowanym amplitudowo (Duquesnoy, 1983; Eisenberg i in., 1995), różnica pomiędzy słuchaczami bez oraz z ubytkiem słuchu jest znacznie większa i może sięgać kilkunastu decybeli – mowa prezentowana na tle sygnału zmodulowanego, przypominającego pojedynczego mówcę, jest rozumiana przez osoby z niedosłuchem gorzej, niż w przypadku występowania w tle stacjonarnego szumu, nawet o widmie w kształcie mowy (Moore i in., 2003). Wydaje się to wynikać z ograniczonej możliwości wykorzystania luk w sygnale zakłócającym. Mogą być one dwojakiej natury – czasowe lub spektralne. Obecność czasowych wynika z tego, że istnieją momenty, w których ogólny poziom maskera staje się niższy/niski, na przykład w czasie krótkich przerw, czy wypowiadaniu głosek o, z natury swojej, niższej amplitudzie np. „f". Podczas tych czasowych przerw, stosunek sygnału do szumu jest wysoki, a to pozwala na uzyskanie krótkich „przebłysków"sygnału pożądanego (mowy). Luki widmowe natomiast powstają, gdy widmo mowy, analizowane w krótkim odcinku czasu, jest zwykle inne niż widmo sygnału zakłócającego. Chociaż części widma mogą być całkowicie zamaskowane przez sygnał tła, inne mogą być prawie w ogóle nie maskowane – SNR może często przekraczać 20 dB. Z tego powodu, fragmenty widma sygnału docelowego mogą być „dostrzeżone"i wykorzystywane percepcyjnie do wnioskowania o strukturze całego

dźwięku mowy (Peters, Moore, Baer, 1998).
Nie są do końca znane powody, dla których osoby z niedosłuchem mają trudności z percepcją mowy prezentowanej na tle sygnałów zmodulowanych. W szczególności nie jest jasne, czy problem te wynika głównie z nieumiejętności wykorzystania „luk" czasowych czy z spadków spektralnych. Osoby z niedosłuchem na ogół wykazują ograniczoną rozdzielczość czasową w przypadku bodźców o wolno zmieniających się obwiedniach, czego efektem może być zmniejszenie zdolności do wykorzystywania luk czasowych w sygnale maskującym (Festen, 1987; Glasberg i in., 1987; Moore, 1995). Udowodniono również zmniejszoną selektywność częstotliwościową, co prowadzi do zmniejszonej zdolności do wykorzystania spadków widmowych (Glasberg i Moore, 1986; Tyler, 1986).
Bardziej szczegółowy opis wpływu sygnałów maskujących na zrozumiałość mowy, zwłaszcza z uwzględnieniem efektu *masking release* opisano w części eksperymentalnej (E4).

4.2 Maskowanie energetyczne i informacyjne

Istnieją dwa główne rodzaje zakłóceń mogących wpływać na percepcję mowy. W pierwszym mechanizmie fizycznie oddziałują one na sygnał mowy, co określa się mianem maskowania energetycznego (Pollack, 1975). U podstaw zjawiska leży obciążenie indywidualnych zasobów uwagi i zaburzenie zdolności do rozdzielania strumieni sygnałów w celu ich skutecznego rozpoznawania (Brungart, 2001; Schneider i in., 2007).
W przypadku gdy szum interferuje z mową na poziomie percepcyjnym stosuje się określenie maskowanie informacyjne (Pollack, 1975; Watson i in., 1976; Shinn-Cunningham i in., 2005).
O ile sposób działania maskowania energetycznego jest stosunkowo oczywisty, o tyle maskowanie informacyjne (które z reguły nie występuje w odizolowaniu od działania maskowania energetycznego) stanowi rozbudowane zagadnienie. Można rozpatrywać je na trzech poziomach – właściwościach akustycznych, językowych oraz charakterystyce słuchacza (Best, 2023).
Na Rys. 11 w sposób bardzo uproszczony przedstawiono mechanizm maskowania na zasadzie wizualnej analogii. Obiekt podstawowy – figurka w białym nakryciu głowy – jest w niej odpowiednikiem sygnału, któremu nie towarzyszą zakłócenia. Jej postrzeganie jest stosunkowo łatwe, o ile wydolność układu odbiorczego nie jest ograniczona np. wadą wzroku (czy, w przypadku sygnałów akustycznych – niedosłuchem). Pojawienie się zaburzenia (zaciemnienie rysunku, a tym samym

utrudnienie jego widzenia) można utożsamić z maskowanieniem energetycznym. To jak bardzo będzie ono efektywne (jak bardzo zmniejszyła się widoczność figurki) zależy m.in. od poziomu maskera oraz jego składu częstotliwościowego. Trzecia część rysunku ilustruje maskowanie informacyjne. Zauważenie figurki nie opiera się wyłącznie na prostej detekcji, ale również rozpoznaniu jej między innymi, podobnymi (konkurującymi) obiektami.



Rys. 11: *Uproszczona wizualna analogia mechanizmu maskowania, w którym obiektem podstawowym jest figurka w białym nakryciu głowy: panel lewy – bez maskowania, panel środkowy – maskowanie energetyczne, panel prawy – maskowanie informacyjne*

Na poziomie akustycznym efektywność maskowania informacyjnego zależy od tego, jak bardzo sygnał maskujący przypomina mowę lub do jakiego stopnia konkurująca z sygnałem docelowym mowa jest zniekształcona. Istotne znaczenie ma również czas trwania maskera – im jest on dłuższy, tym maskowanie bardziej skuteczne (Ezzatian i wsp., 2012). Ciekawe badania przeprowadzone przez Browna i Bidelmana (2022) sugerują natomiast, że muzyka, która często towarzyszy sytuacjom, w którym warunki komunikacyjne są najtrudniejsze (np. przyjęcia) może być jeszcze bardziej efektywnym maskerem, jeśli towarzyszy jej wokal.

Drugą płaszczyzną w zakresie której należy analizować maskowanie informacyjne są czynniki lingwistyczne. Udowodniono, że w przypadku zakłócania mowy mową (*speech-on-speech*), jest ono skuteczniejsze, gdy język sygnału docelowego i maskującego jest taki sam lub podobny (pochodzi z tej samej grupy językowej) – Calandruccio i zespół., 2010. Znaczenie ma również płeć osoby mówiącej. Jeżeli jest ona tożsama w sygnale wiodącym i konkurującym, efekt maskowania jest silniejszy. Efektywność maskowania może być zwiększona, gdy treść wypowiedzi zakłócającej jest istotna dla słuchacza. Kiedy w konkurencyjnym strumieniu mowy, nawet dotychczas ignorowanej, słuchacz zauważy własne imię, jego skupienie na sygnale pożądanym maleje (Wood i Cowan 1995). W tym kontekście warto jeszcze raz podkreślić kwestię uwagi i jej (nadal nie do końca zdefiniowanej) roli w percepcji słuchowej. Brungaart i Simpson (2002) koordynowali badanie, w ramach którego analizowano wpływ uwagi na percepcję mowy w utrudnionych warunkach, oceniając, w jaki sposób słuchacz radzi sobie z rozdzielaniem sygnałów konkurujących, z których każdy stanowiła mowa. Choć otrzymane wyniki są niejednoznaczne

i sugerują konieczność prowadzenia dalszej weryfikacji modeli uwagi słuchowej, wyraźnie zasygnalizowany jest silny związek skuteczności separacji sygnałów (a więc ignorowanie zakłóceń) ze stopniem koncentracji uwagi na sygnale docelowym. Cunningham i Best (2008) uzupełniają te założenia o konieczność właściwego grupowania pożądanych obiektów (sygnałów). Według ich założeń, słuchacze powinni znać cechę, która identyfikuje obiekt i pozwala im skupić oraz utrzymać na nim uwagę (podążać za nim). Zachowanie tej zdolności poznawczej jest kluczowe w relacjach społecznych bazujących na komunikacji (najczęściej werbalnej). W niekorzystnych warunkach fragmenty odbieranej wiadomości często pozostają dla słuchaczy niedostępne, w wyniku maskowania przez konkurencyjne źródła, trudności w wyborze sygnału pożądanego oraz ograniczoną efektywność przełączania uwagi. Mimo to udaje się utrzymać wystarczającą zrozumiałość poprzez uzupełnianie brakujących części na podstawie informacji usłyszanych, kontekstu wypowiedzi oraz powiązania z danymi przechowywanymi w pamięci i/lub osobniczym leksykonie. Szybkość każdego z tych etapów przetwarzania jest niezwykle istotna ponieważ interakcje społeczne, jak i reagowanie na zmiany w środowisku wymagają śledzenia przepływających informacji w czasie rzeczywistym. Składa się nań jednak wiele cech osobniczych.

Trzecią grupą czynników wpływających na odseparowanie zakłóceń od sygnału docelowego są właśnie cechy związane z odbiorcą. Niektóre populacje są szczególnie wrażliwe na maskowanie informacyjne – dzieci, osoby starsze, z niedosłuchem lub afazją.

W przypadku dzieci przyczyną szczególnej podatności na zakłócenia jest niewystarczające wykształcenie mechanizmów centralnego przetwarzania odpowiedzialnych za segregację strumieni oraz wspomnianą wcześniej uwagę (Moore i Linthicum, 2007), natomiast osób starszych – deficyty o charakterze neurodegeneracyjnym, najczęściej związane z wiekiem (Buss i wsp., 2019). Istnieje kilka mechanizmów pozwalających, na poziomie percepcyjnym, ograniczyć wpływ maskowania informacyjnego na zrozumiałość mowy. Po pierwsze są to wskazówki przestrzenne. Zmniejszenie efektywności maskowania informacyjnego możliwe jest, gdy masker i sygnał pożądany ulegają przestrzennej separacji (spatial release from masking), o czym pisali m.in. Kociński i Sęk (2005). Właśnie dlatego ludzie przebywając w większym gronie intuicyjnie kierą głowę (ucho) w stronę swojego rozmówcy. Inną strategią jest natomiast kierowanie uwagi na wskazówki pozasłuchowe, w tych najbardziej istotną – ruch warg, co zostało szczegółowo opisane w podrozdziale 3.1.

4.3 Szybkość mowy

Percepcja mowy w dużym stopniu zależy od efektywności przetwarzania czasowego w obrębie drogi słuchowej. Zadanie to jest szczególnie utrudnione gdy mowa jest przyspieszona. W ogólności, zdolność rozumienia mowy skompresowanej w czasie zmniejsza się wraz ze wzrostem szybkości mowy (m.in. Dupoux i Green, 1997), a prawidłowość ta jest zachowana również w przypadku mowy prezentowanej w hałasie (np. Tun, 1998). Najnowsze badania wskazują, że chociaż górna granica 50% zrozumiałości mowy skompresowanej czasowo przez młodych, prawidłowo słyszących badanych wynosi od 11 do 16 sylab na sekundę, zmienność międzyosobnicza jest znaczna (Gransier i in., 2022). W przypadku osób starszych obserwuje się większe trudności z rozpoznawaniem mowy przyspieszonej niż prezentowanej w normalnym (naturalnym) tempie (m.in. Vaughan i Letowski, 1997). U podstaw tego zjawiska leżą problemy z przetwarzaniem czasowym dźwięków, których etiologia może dotykać różnych etapów obwodowego i/lub centralnego przetwarzania słuchowego, a także spowolnienia poznawczego będącego efektem starzenia się organizmu (Schneider i in., 2005), nawet jeśli progi słyszenia pozostają w normie.

Między innymi dlatego testy wykorzystującę mowa przyspieszoną moga być stosowane do określenia lokalizacji odbiorczego niedosłuchu, w szczególności w diagnostyce dysfunkcji ośrodkowych w obrębie pnia mózgu, ośrodków korowych i podkorowych (Pruszewicz, 2010).

Ciekawym przypadkiem, coraz szerzej opisywanym przez literaturę są osoby niedowidzące i niewidome. Dietrich i wsp. wykazali, że potrafią one rozpoznawać mowę przyspieszoną do ponad 20 sylab na sekundę, a więc niemal dwukrotnie więcej niż w przypadku badanych z prawidłowym wzrokiem (2013). Umiejętność ta związana jest z tym, żę u osób niewidomych część kory mózgowej, która normalnie reaguje na wzrok, reaguje na mowę. Kluczowe znaczenie dla przeprogramowania obszarów mózgu kontrolujących słuch w obszarze, który normalnie przetwarza wzrok ma wiek, w którym dana osoba straciła wzrok.

4.4 Warunki propagacji – akustyka pomieszczenia

Warto podkreślić, że badania zrozumiałości mowy nie są wykorzystywane jedynie do analizy funkcjonowania układu słuchowego i ewentualnych jego dysfunkcji. W szeroko rozumianej akustyce wnętrz niezwykle istotne jest, by wyniki subiektywnych ocen jakości transmisji mowy były jak najbardziej zależne od

parametrów fizycznych badanego kanału komunikacyjnego (Houtgast i in., 1985; Steeneken, 1992), bo wtedy warunki odsłuchu mogą być dokładniej kontrolowane. Z tego powodu w weryfikacji stopnia zrozumiałości mowy w danym pomieszczeniu w wielu przypadkach zaleca się stosowanie list logatomowych, złożonych ze słów, które pozbawione są nacechowania semantycznego. Poprzez minimalizowanie wpływu skojarzeń poznawczych (niska redundancja), listy logatomowe pozwalają poddać ewaluacji przede wszystkim „słyszalność" (ostrość słuchu), a nie wydajność przewidywania leksykalnego, silnie wynikającego z kontekstu, jak ma to miejsce w przypadku testów wykorzystujących zdania (Stickney i Assmann, 2001). Istotne jest, aby listy te były zrównoważone fonetycznie, zgodnie z występowaniem statystycznym początkowych spółgłosek, samogłosek i końcowych spółgłosek w danym języku

4.5 Wiek

Liczne badania wskazują na to, że zmiany związane z wiekiem negatywnie rzutują na przetwarzaniu cech akustycznych w domenie czasu, do których zaliczyć można np. zmieniającą się obwiednię sygnału. Z tego właśnie powodu osoby starsze, nawet posiadające słuch prawidłowy, częściej mają problem ze zrozumieniem mowy w warunkach obecności sygnałów zakłócających, niż ma to miejsce w przypadku słuchaczy młodszych (Frisina i in., 1997). Jednym z przykładów wskazujących na prawdziwość tego stwierdzenia jest chociażby efektywność zjawiska masking release. Szczegółowe dane dotyczące związku wieku ze stopniem zrozumiałości mowy wśród badanych ze słuchem prawidłowym, jak i z ubytkami słuchu, przedstawiono w dalszej, eksperymentalnej części pracy.

Ograniczenie zdolności do rozumienia mowy u osób starszych nie może być jednoznacznie związane wyłącznie z efektami starzenia w obwodowym układzie słuchowym (ang. *presbyacusis*), choć te są bardzo istotne. Norma ISO 7029 definiuje przewidywane progi słyszenia wyznaczone w audiometrii tonalnej dla osób w różnym wieku, z których bezsprzecznie wynika, że już powyżej 40 r.ż. audiogramy zaczynają znacząco opadać dla częstotliwości powyżej 1000–1500 Hz. Oczywiście olbrzymią rolę odgrywają także czynniki poznawcze, istotne jest też uwzględnienie centralnego przetwarzania słuchowego na poziomie ponad peryferyjnym (pień mózgu i kora mózgu).

Mimo że podwyższenie progów słyszenia i postępujący wiek zostały zidentyfikowane jako główne czynniki przyczyniające się do pogorszenia rozpoznawania mowy w akustycznie wymagających warunkach, deficyty mogą występować również u osób

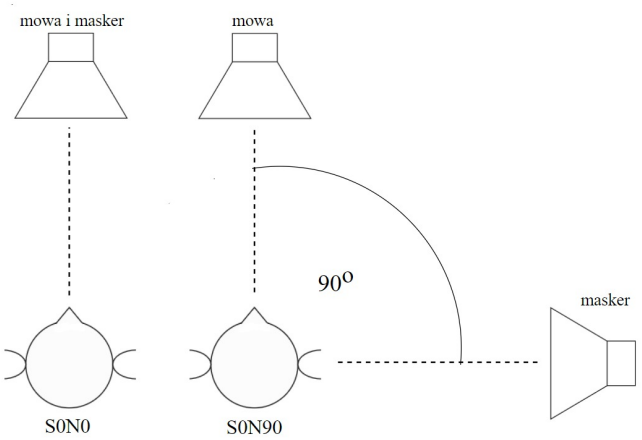
młodych. Poza aspektami lingwistycznymi uwzględnić w tym kontekście można także wykrywanie sygnałów dychotycznych, maskowanie wielokrotne, segregację strumieni i detekcję modulacji, które mogą być zaburzone nawet w przypadku słuchu prawidłowego lub właściwej kompensacji niedosłuchu (przez aparaty słuchowe; Humes i in., 2013). W badaniu przeprowadzonym na młodych i starszych osobach z normalnymi warunkami słuchowymi, Füllgrabe i Moore (2014) udowodnili, że istnieje silna korelacja między rozumieniem mowy w hałaśliwych środowiskach a wrażliwością na struktury czasowe.

4.6 Słyszenie binauralne

Jednym z najważniejszych czynników kształtujących percepcję słuchową jest binauralność (obuuszność; z łac. *bi* – podwójny, *auris* – ucho). To dzięki niej możliwe są, poza bardziej precyzyjną lokalizacją, odseparowanie źródeł dźwięku i zmniejszenie wpływu zakłóceń na odbiór sygnału docelowego. Przestrzenne rozdzielenie źródła mowy i szumu stanowi kluczowy mechanizm w prawidłowym rozpoznawaniu mowy w warunkach akustycznie wymagających, co zostało wykazane w wielu publikacjach (np. Bronkhorst, 2000 oraz, dla języka polskiego, Sęk i in. 2004; Kociński i in., 2005). Parametrami charakteryzującymi efektywność przetwarzania obuusznego w kontekście zrozumiałości mowy są *Intelligibility Level Difference* (ILD) oraz *Binaural Intelligibility Level Difference* (BILD). Pierwszy z nich dotyczy korzyści uzyskanych przez słuchacza z przestrzennego odseparowania sygnału pożądanego (w tym przypadku mowy) i maskującego, a wyraża się poprzez różnicę SRT (poziom sygnału lub stosunek sygnału do szumu (SNR), przy którym badanych prawidłowo powtarza 50% zaprezentowanego materiału testowego; szczegółowy opis w podrozdziale 5.2.3.) pomiędzy dwoma binauralnymi konfugiracjami przestrzennymi:

1. sygnał mowy (S) i szum (N) prezentowane z kierunku „na wprost" słuchacza (0°) – S0N0,
2. sygnał mowy pochodzi z kierunku „na wprost", a maskujący pod kątem 90°– S0N90.

Wspomniane konfiguracje przedstawiono na Rys. 12.



Rys. 12: *Binauralne konfiguracje przestrzenne pomiaru SRT – S0N0 oraz S0N90*

Ze względu na efekt cienia głowy (*head shadow effect*) i obuuszne przetwarzanie, separacja mowy i maskowania prowadzi do poprawy SRT.

W ramach wcześniej realizowanych badań autorki rozprawy przeprowadzono pomiary ILD (Pastusiak i in., 2019). Skorzystano w nim z przestrzennego rozseparowania mowy i sygnałów zakłócających z uwzględnieniem słuchaczy prawidłowo słyszących oraz osób z symetrycznym niedosłuchem odbiorczym. Aby rozróżnić wspomniany wyżej wpływ efektu cienia głowy od przetwarzania obuusznego na ośrodkowym (centralnym) poziomie drogi słuchowej przeprowadzono dodatkowy pomiar BILD, w którym SRT mierzony jest w zdefiniowanej wyżej konfiguracji S0N90, a następnie w tym samym ustawieniu, ale z wyłączeniem ucha, w które skierowany jest sygnał maskujący – S0N90 (MON). BILD to różnica między SRT dla konfiguracji S0N90 mierzonym monauralnie i obuusznie. Z uwagi na korzyści uzyskane dzięki obuusznemu przetwarzaniu i występowaniu akustycznego cienia głowy, wyniki uzyskane w warunkach obuusznej prezentacji S0N90 są niższe niż w prezentacji monauralnej, o ile progi słyszenia osoby badanej go pozostają w normie.

Otrzymane wyniki faktycznie wykazały obniżenie SRT po separacji przestrzennej mowy i szumu maskującego zarówno dla grupy kontrolnej, jak i słuchaczy z ubytkiem słuchu. Różnice uzyskane w zależności od warunków odsłuchu były jednak mniejsze w grupie badanej niż w kontrolnej, co wskazuje na to, że osoby z niedosłuchem nie mogą czerpać tak dużych korzyści z przestrzennego oddzielenia mowy i sygnału zakłócającego, jak słuchacze ze słuchem prawidłowym. Także zmienność między słuchaczami jest wyższa w grupie badanej niż w grupie kontrolnej, na co wskazuje wyższe odchylenie standardowe SRT. Jeśli chodzi o pomiar ILD, poprawa SRT spowodowana efektem cienia głowy i obuusznego przetwarzania wynosiła średnio

8.1 ± 1.8 dB w grupie kontrolnej i 3.5 ± 1.9 dB w grupie słuchaczy z niedosłuchem. Parametr BILD osiągał wartość od 3 do 6 dB (średnio 4.6 dB) w grupie kontrolnej i zmalał ponad dwukrotnie w grupie osób z niedosłuchem do średnio 1.9 dB (z odchyleniem 1.4 dB).

Powyższe dane potwierdzają, że u osób z niedosłuchem odbiorczym wykorzystanie separacji przestrzennej jednocześnie prezentowanych sygnałów mowy i maskera jest mniej efektywne w zakresie poprawy zrozumiałości mowy. Obserwacja ta dotyczy również parametru BILD, który określa przyrost zrozumiałości mowy wynikający z recepcji sygnału dwuusznie, w porównaniu z prezentacją jednouszną. Dane referencyjne uzyskane w grupie kontrolnej są zgodne z wynikami opisanymi w literaturze, również dla innych niż polski języków (Wagener i Brand, 2006; Bronkhorst, Plomp, 1989; Beutelmann i in., 2010).

Aby lepiej zrozumieć mechanizmy, które leżą u podstaw zmniejszonej przewagi obuusznej w grupie badanej, wartości SRT mierzone w różnych konfiguracjach zostały zestawione z indywidualnymi progami słyszenia wyznaczonymi przed pomiarami. Określano również wzajemne powiązanie wartości SRT uzyskanych dla konfiguracji mono- i binauralnej S0N90. Szczegółowe informacje znajdują się w publikacji autorki niniejszej rozprawy (Pastusiak i wsp., 2019; w szczególności Rys. 6 i 7) Na podstawie uzyskanych wyników w ogólności stwierdzić można, że:

- w przypadku osób z niedosłuchem odbiorczym wpływ na zrozumiałość mowy w szumie mają czynniki nieuchwytne w teście wykorzystującym wyłącznie proste sygnały tonalne (audiometria tonalna),

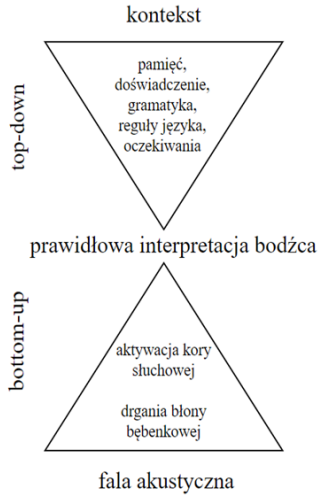
- w grupie słuchaczy z niedosłuchem wykazano mniejsze korzyści binauralne wynikające z przestrzennego odseparowania mowy i szumu, niż w grupie kontrolnej,

- nie stwierdzono związku między stopniem ubytku słuchu (określonym przez PTA) a deficytami binauralnymi wpływającymi na zrozumiałość mowy w hałasie.

Powyższe wnioski sugerują, że obuuszne pomiary zrozumiałości mowy w hałasie mogą, a wręcz powinny być włączone do baterii testów diagnostycznych, aby móc dokładniej scharakteryzować indywidualny deficyt słuchu wpływający na zrozumiałości mowy w warunkach akustycznie wymagających.

4.7 *Cocktail-party effect*

Analizując wpływ sygnałów zakłócających na zrozumiałość mowy, nie sposób nie wspomnieć o zjawisku *cocktail-party*, opisywanym przez literaturę już od lat 50. XIX wieku. Istnieje szereg mechanizmów, które pozwalają na efektywne wykorzystanie użytecznej informacji językowej w trudnych warunkach akustycznych.
Po pierwsze, niezbędna jest właściwa analiza tzw. sceny słuchowej (zbioru dźwięków obecnych w otoczeniu). W swojej pracy z 1990 r. Bregman wprowadza rozróżnienie między sekwencyjną i równoczesną integracją. Pierwsza z nich odnosi się do dźwięków, które pochodzą z jednego źródła, ale są oddzielone w czasie (np. sylaby i wyrazy) i ich łączenia w spójny strumień słuchowy przy równoczesnym oddzieleniu od sygnałów zakłócających. Integracja równoczesna polega na grupowaniu jednocześnie występujących komponentów widma częstotliwościowego (np. harmoniczne, formanty) w reprezentację akustyczną jednego źródła dźwięku oraz odseparowanie jej od innych sygnałów. Wymienione procesy są komplementarne i pozwalają na zintegrowanie ze sobą elementów sygnału pożądanego (głos interesującego nas mówcy) oraz odseparowanie innych, niepożądanych (głosy innych osób, muzyka, szuranie krzeseł, brzęk naczyń i sztućców itp.).
W analizie słuchowej biorą również udział procesy oddolne (ang. *bottom-up*), napędzane przez bodźce docierające do receptora, jak i odgórne, angażujące wyższe procesy poznawcze i zależne od wcześniejszych doświadczeń i oczekiwań słuchacza (ang. *top-down*). Obydwa schematycznie przedstawiono je na Rys. 13 poniżej oraz opisano w rozdziale 7.



Rys. 13: *Schematyczny przebieg procesów oddolnych i odgórnych w percepcji słuchowej*

4.8 Ubytek słuchu

Przy podwyższonych progu detekcji dźwięku spowodowanych ubytkiem słuchu, niektóre części sygnału docelowego (np. mowy) mogą być zbyt ciche (poniżej progów) i stać się niesłyszalne. We wstępie pracy wskazano, że najważniejsze informacje dotyczące sygnału mowy, których właściwe wykorzystanie pozwala na jego zrozumienie należą do domeny czasowej, przestrzennej oraz spektralnej. W przypadku osób ze słuchem prawidłowym, u których nie występują dodatkowe zmiany o charakterze neurodegeneracyjnym, wyżej wymienione informacje są wysoce redundantne. Dzięki temu odseparowanie mowy od sygnałów zakłócających w zasadzie nie stanowi większego problemu – mowa może być w pełni zrozumiała nawet przy ujemnych wartościach SNR. Obecność ubytku słuchu często jednak negatywnie, w stopniu zależnym od głębokośći niedosłuchu, wpływa na mechanizmu pozwalające korzystać w dostarczanych do układu słuchowego informacji. Ubytkom słuchu niezwykle często towarzyszy ograniczona rozdzielczość widmowa. W praktyce oznacza to bardziej efektywne maskowanie mowy o częstotliwościach zbliżonych do częstotliwości występujących w sygnale maskującym – sygnał pożądany i zakłócający nie mogą być rozdzielone tak precyzyjnie, jak ma to miejsce w przypadku prawidłowo funkcjonującej separacji spektralnej (Festen i Plomp, 1983).
Pogorszona rozdzielczość czasowa, również niezwykle często towarzysząca niedosłuchowi, ogranicza natomiast możliwość wykorzystania faktu, iż w wielu sytuacjach hałas w tle nie jest stały w czasie, ale podlega pewnej fluktuacji (Bacon i in., 1998). Osoby prawidłowo słyszące czerpią korzyść z odcinków czasu, w których stosunek sygnału do szumu przyjmuje bardziej korzystną wartość – krótkie fragmenty sygnału mowy nie są (lub są mniej efektywnie) maskowane przez zmodulowany sygnał zakłócający. Ten zysk percepcyjny może osiągać wartości rzędu kilkunastu decybeli. W przypadku występowania niedosłuchu możliwość skorzystania z tego mechanizmu nazywanego *listening in the gaps*, jest, co zmierzono w ramach realizacji pracy doktorskiej oraz szczegółowo opisane w części eksperymentalnej, mocno ograniczone.

Chociaż ubytek słuchu jest jednym z najbardziej istotnych czynników wpływających na pogorszenie stopnia zrozumiałości mowy należy pamiętać, że nie jedynym. Co więcej, może zdarzyć, że zaburzenia percepcji mowy występują u osób, których audiogramy przyjmują prawidłowy przebieg dla całego badanego zakresu częstotliwości. Przykładem takiej sytuacji jest *hidden hearing loss* (HHL), czyli stan w którym pacjenci doświadczają trudności w słyszeniu mowy,

zwłaszcza w warunkach niekorzystnych akustycznie mimo prawidłowych wyników standardowych testów audiometrycznych. Związane jest to z dysfunkcją lub uszkodzeniem komórek słuchowych lub połączeń synaptycznych w obrębie ślimaka lub nerwu słuchowego. Etiologia HHL jest nie do końca poznana. Przyczyn można dopatrywać się m.in. w ekspozycji na hałas o dużym natężeniu. Kujawa i Liberman (2009) wykazali na modelach zwierzęcych, że nawet pojedyncza ekspozycja na duży hałas może uszkodzić zakończenia ślimakowych neuronów aferentnych. Osłabienie to obserwuje się nawet wtedy, kiedy komórki słuchowe wydają się być nieuszkodzone. Inne badania (np. Paul i in., 2017) wskazują na to, że HHL przyczynia się do powstawania deficytów w kodowaniu modulacji amplitudy i przetwarzania czasowego sygnału, przede wszystkim mowy (Bharadwaj i in. 2015), a wielkość tych zmian jest często związana z wiekiem (Clinard i Tremblay 2013; King i in. 2014). Sergeyenko i in. (2013) wykazali, że istnieje silna korelacja między uszkodzeniami synaps a zmniejszonymi odpowiedziami nadprogowymi u myszy podczas starzenia się, co potwierdza pogląd, że synaptopatia[5] (zaburzona komunikacja między synapsami) jest kluczowym czynnikiem przyczyniającym się do postępującej neuropatii słuchowej[6] (zaburzenia synchronizacji neuronalnej objawiającej się nieprawidłowości w badaniu ABR pomimo zarejestrowanej otoemisji OAE) i HHL nawet przy braku wspomnianej wyżej nadmiernej ekspozycji na hałas. Badania prowadzone w ostatnich latach wskazują, że jedną z przyczyn HHL niekoniecznie musi być synaptopatia czy uszkodzenie samych komórek słuchowych,

[5]Synaptopatią określa się dysfunkcję połączeń synaptycznych komórek słuchowych wewnętrznych ze zwojem spiralnym. Choć przyczyny mechanizmów jej powstawania nie są jednoznacznie określone, wiele badań sugeruje silne uwarunkowanie genetyczne w zakresie zaburzeń w wytwarzaniu glutaminianu w pęcherzykach synaptycznych tych synaps (Moser, 2016). L-Glutaminian uważany jest za neuroprzekaźnik w pierwszej synapsie słuchowej pomiędzy komórkami włoskowymi wewnętrznego ucha a dendrytami komórek typu I zwojów spiralnych. Wyniki te sugerują, że nadmierne uwalnianie glutaminianu w ślimaku jako skutek niedotlenienia lub urazu akustycznego powoduje uszkodzenia otaczających neuronów, zwłaszcza dendrytów pierwszego aferentnego nerwu słuchowego (Rothman, 1984).

[6]Po raz pierwszy przypadek niedosłuchu odbiorczego ze znacznym ograniczeniem zrozumienia mowy przy zachowanej otoemisji przy braku potencjałów wywołanych z pnia mózgu lub pojawieniu się ich wyłącznie przy wysokich progach opisali pod koniec lat 90. XX wieku Starr i in. (1991). Stan ten określono auditory neuropathy, a więc właśnie neuropatią słuchową. Na skutek uszkodzenia komórek słuchowych wewnętrznych, zwoju spiralnego, włókien nerwu ślimakowego oraz dysfunkcji neurotransmiterów synaptycznych dochodzi do zaburzeń synaptycznego kodowania informacji akustycznych, a także synchronizacji pobudzeń neuronów zwoju spiralnego. Między innymi dlatego percepcja mowy jest ograniczona nieproporcjonalnie w stosunku do progów słyszenia w audiometrii tonalnej. Etiologia neuropatii słuchowej jest niejasna, natomiast do jej potencjalnych przyczyn mogą należeć reakcja immunologiczna, hiperbilirubinemia lub infekcje. Istotny jest również czynnik genetyczny – szacuje się, że w 42% schorzenie występuje rodzinnie (Sininger, 2002). Protezowanie słuchu, głównie w celu osiągnięcia poprawy w zrozumiałości mowy i obniżenia wysiłku słuchowego jest trudne i nie zawsze przynosi satysfakcjonujące efekty (w badaniach Jijo (2013) było to zaledwie 30% pacjentów). Protetykom słuchu zaleca się zastosowanie niewielkich wzmocnień (zwłaszcza przy prawidłowych audiogramach), ale obejmującej szeroki zakres częstotliwości kompresji dynamiki sygnału wejściowego (Obrębowski, 2010).

ale nieprawidłowa czynność komórek Schwanna odpowiedzialnych za wytwarzanie osłonki mielinowej okrywającej włókna nerwowe, w tym w układzie słuchowym (Wan i Corfas, 2017).

4.9 Kontekst

Różnice między zrozumiałością całego zdania a identyfikacją pojedynczych słów zwykle są związane z kontekstem (Skrodzka, 2014). Codzienne wypowiedzi zawierają dużo informacji kontekstowych, które pomagają słuchaczowi wydedukować niezrozumiałe słowa lub ich fragmenty. Zgodnie z niektórymi teoriami percepcji mowy, część sygnału, która nie została zrozumiana jest przechowana w pamięci krótkoterminowej, a następnie analizowana w kontekście kolejnych, zrozumianych już fonemów, sylab lub wyrazów. Możliwość wykorzystania kontekstu jest zatem niezwykle korzystna i niejednokrotnie przyczynia się do poprawy stopnia zrozumiałości mowy oraz obniżenia wysiłku słuchowego. Dobrą analogią może być przykład przedstawiony poniżej (Rys. 14), który, wprawdzie dotyczy modalności wzrokowej, jednak pozwala lepiej zrozumieć omawiane zjawisko. Mimo że część zdania jest nieczytelna, stosunkowo łatwo jest je odczytać, właśnie poprzez skorzystanie z kontekstu.

Rys. 14: *Schematyczne przedstawienie możliwości wykorzystania informacji kontekstowej (zdanie: „Ala ma kota")*

To, jak efektywnie może być wykorzystany kontekst zależy od wielu czynników. Poza *stricte* akustycznymi (np. warunki prezentacji sygnału), znaczącą rolę odgrywa znajomość tematu rozmowy, mówcy oraz wydolność aparatu percepcyjnego, zarówno na poziomie peryferyjnym, jak i ośrodkowym. Wykazano, że gdy słuch jest uszkodzony centralnie, kontekst wypowiedzi nie może być wykorzystany (Skrodzka, 2014).

Jak wynika z przedstawionych wyżej treści, poza negatywnym skutkiem obecności zewnętrznych czynników zakłócających, oraz ewolucyjnymi mechanizmami pozwalającymi zwiększyć efektywność ekstrakcji informacji, działającymi na niższych piętrach drogi słuchowej oraz samego języka (wspomniane tu redundancja i kontekst) istnieją też wysokopoziomowe aspekty wspomagające

efektywne pozyskiwanie informacji. Są one jednak często mocno subiektywne i zindywidualizowane.

4.10 Kompensacja poznawcza

Gdy słyszalność zawodzi, strategie kompensacji poznawczej (tłumienie nieistotnych informacji, poleganie na kontekście itp.) mogą częściowo wyrównać taką utratę informacji. W takim przypadku słuchacze wkładają więcej wysiłku umysłowego, aby skoncentrować się na interesujących ich źródłach dźwięku i przeznaczają więcej zdolności umysłowych na zrozumienie i zapamiętanie tego, czego słuchają. Tak więc zdolności przetwarzania poznawczego, takie jak uwaga, pamięć i szybkość przetwarzania, odgrywają istotną rolę w słuchaniu w niekorzystnych warunkach. Zdolności poznawcze różnią się u poszczególnych osób, a niektóre z nich zmniejszają się wraz z wiekiem (Murman, 2015). Ten rozrzut przynajmniej częściowo wyjaśnia różnice w wydajności rozumienia mowy lub postrzeganego wysiłku słuchowego u osób z podobnym ubytkiem słuchu, mierzonym za pomocą audiometrii tonalnej. Akustyczne przebiegi dźwięków mowy zmieniają się w sposób zależny od dźwięków sąsiadujących (poprzedzających i następujących) w niezwykle skomplikowany sposób. Percepcja, a następnie zrozumienie mowy ciągłej (płynnej, pozbawionej niepotrzebnych przerw i pauz) nie zależy jedynie od informacji która zawarta jest w sygnale akustycznym. Część wyrazu, która jest wysoce prawdopodobna biorąc pod uwagę całe zdanie może być „usłyszana" nawet wtedy, gdy odpowiadająca mu część fali dźwiękowej jest silnie zniekształcona lub nawet nieobecna. W jednym ze swoich badań Warren (1970) wykazał, że zastąpienie fragmentu mowy zakłóceniem (np. kaszlem) pozwala słuchaczom usłyszenie brakującej części zdania, o ile zakłócenie to jest wystarczająco głośne i zawiera składowe o częstotliwościach zbliżonych do domyślnego, przerywanego sygnału.

Opisane powyżej elementy i mechanizmy fizjologiczne w drodze słuchowej są niezwykle subtelne i wrażliwe na wszelkie zakłócenia. Każdy defekt, każdy artefakt może prowadzić do obniżenia wydajności procesu, w którym mają one istotne znaczenie - w tym przypadku zrozumiałości mowy. Uszkodzenie jednego z poziomów drogi słuchowej skutkuje przekazaniem niepełnej informacji do kolejnych etapów przetwarzania i analizy bodźca. Przykładem może być przestrzenne odmaskowanie (*spatial release from masking*) – w przypadku niesymetrycznego niedosłuchu, już na wejściu do układu przetwarzania sygnał mowy jest bardziej zniekształcony niż

w przypadku symetrycznego słuchu prawidłowego.
W związku z tym, tak istotną rolę pełnią skuteczne i wiarygodne narzędzia diagnostyczne, które pozwalają na ocenę stanu każdego elementu i procesu biorącego udział w rozumieniu mowy.

Rozdział 5

Audiometria mowy jako metoda diagnostyki słuchu

W tym rozdziale po raz kolejny przedstawiono argumenty przemawiające za potrzebą stosowania badań wykorzystujących sygnał mowy w praktyce klinicznej oraz gabinetach protetyki słuchu. Na wstępie zdefiniowano podstawowe pojęcia związane z audiometrią mowy składające się na jej techniczne aspekty. Zaprezentowane treści uzupełniono odniesieniami do wyników badań eksperymentalnych przeprowadzonych przez autorkę niniejszej rozprawy. Wymieniono także rodzaje dostępnych testów oraz krótko scharakteryzowany każdy z nich.

Pierwsze próby ustrukturyzowanej oceny zrozumiałości mowy zostały przeprowadzone na początku XX wieku (Chermak, 1997), jednak dopiero pięć dekad później zaczęto wykorzystywać je w celach klinicznych. Stało się to za sprawą Hudginsa, Hawkinsa, Karlina i Stevensa w Bell Telephone Laboratories (Iwankiewicz, 1961). Od tego czasu obserwuje się wzrost zainteresowania metodami pozwalającymi ocenić zdolności komunikacyjne pacjentów prawidłowo słyszących oraz z niedosłuchem różnego pochodzenia zarówno w ciszy, jak i, co uzasadnione rzeczywistymi warunkami środowiskowymi, w obecności sygnałów maskujących. Przez lata opracowano wiele testów audiologicznych wykorzystujących materiał językowy. Narzędzia te mogą dostarczyć cennych informacji klinicystom, badaczom oraz samym użytkownikom aparatów słuchowych i ich rodzinom w celu optymalizacji leczenia oraz poprawy zdolności komunikacyjnych. Ich szczegółowy opis został zawarty w rozdziale 6.

5.1 Zasadność oceny zrozumiałości mowy

Bez wątpienia audiometria tonalna jest podstawowym narzędziem służącym do diagnozowania i monitorowania ubytków słuchu. Mimo swojej dostępności i powszechności przy jednoczesnym zachowaniu dużej wartości klinicznej, nie jest wystarczająca do pełnej oceny zdolności komunikacyjnych. W początkowej części niniejszej rozprawy omówiono fizyczne cechy sygnału mowy. Zaznaczono, jak bardzo jest on złożony – charakteryzuje się szerokim zakresem częstotliwości, natężenia oraz zmienną charakterystyką (Moore, 1999). Należy pamiętać, że percepcja mowy nie opiera się wyłącznie na ocenie mierzalnych, obiektywnych cech dźwięku. Oczywiście, ich analiza jest istotna, ale nie wyczerpuje w pełni zagadnienia. W tym kontekście warto powołać się na różnice między *słyszeniem* a *słuchaniem*. Słyszenie jest odbiorem dźwięków przez układ słuchowy, natomiast słuchanie procesem ich percepcji i interpretacji przy wykorzystaniu bardziej wyszukanych mechanizmów. Fundamentalne znaczenie mają tutaj chociażby binauralność (a więc również lokalizacja źródła dźwięku, separacja przestrzenna i lepsze rozumienie mowy w utrudnionych warunkach – *cocktail-party effect*) czy wysoka redundantność mowy i możliwość posłużenia się kontekstem, które pozwalają efektywnie skorzystać z informacji językowej, nawet gdy sygnał jest mocno zniekształcony, a nawet częściowo utracony.

Badanie z wykorzystaniem tonów, nawet bardzo skrupulatnie przeprowadzone, nie jest w stanie odzwierciedlić dynamiki rzeczywistego sygnału mowy. Audiometria tonalna mierzy jedynie zdolność słyszenia, nie dając pełnego obrazu percepcji słuchowej, która obejmuje zarówno fizyczne, jak i kognitywne aspekty odbioru dźwięków. Potwierdzenie tej tezy odnaleźć można w wynikach przedstawionych w części eksperymentalnej rozprawy.

5.2 Techniczne aspekty oceny zrozumiałości mowy

5.2.1 Wyrazistość i zrozumiałość

Z badaniami audiometrycznymi wykorzystującymi materiał językowy nierozerwalnie związane są dwa pojęcia: „wyrazistość" i „zrozumiałość". Pierwsze z nich wykorzystuje się w odniesieniu do prawidłowej identyfikacji głosek, sylab i logatomów (przypominających wyrazy struktur nieposiadających nacechowania semantycznego – patrz. rozdział 6.5.). Pojęcie zrozumiałości wiązać należy natomiast ze strukturami wyższego rzędu, mających określony sens (znaczenie)

lub zawierających określoną treść myślową – słów oraz zdań. We współczesnym wykorzystaniu klinicznym spotyka się zarówno testy określające wyrazistość, jak i zrozumiałość, a rodzaj zastosowanego narzędzia zależy przede wszystkim od tego, czy celem badania jest czy dąży się do oceny funkcji receptywnej drogi słuchowej (opartej głównie na fizycznych parametrach sygnału mowy takich jak m.in. natężenie, skład widmowy, częstotliwość), czy funkcji integracyjnej (związanej z umiejętnością identyfikacji usłyszanych jednostek oraz prawidłowym rozpoznaniem treści w nich zawartych na podstawie słownika pamięci osobniczej).
Aby określić zrozumiałość (lub wyrazistość) należy zaprezentować badanemu materiał językowy, a następnie określić, jaka jego część została poprawnie powtórzona. Tak zdefiniowaną zrozumiałość lub wyrazistość można przedstawić prostą zależnością (1):

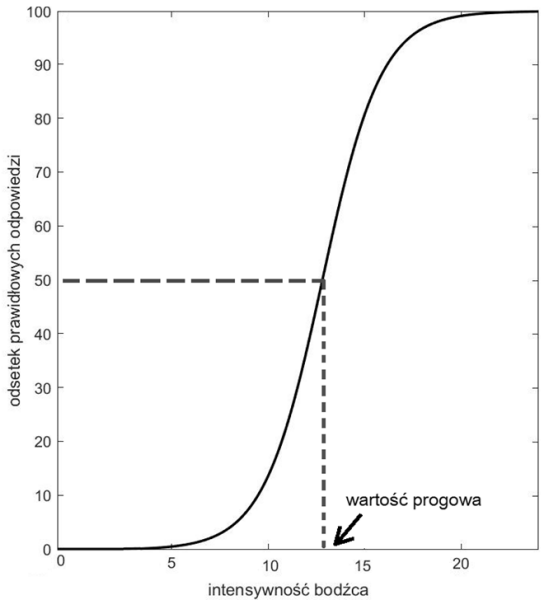
$$I = \frac{p}{n} \cdot 100\% \tag{1}$$

w której I stanowi zrozumiałość (najczęściej wyrażoną w procentach), p liczbę poprawnych odpowiedzi, a n liczbę wszystkich zaprezentowanych elementów testu.

5.2.2 Krzywa psychometryczna

Krzywa psychometryczna opisuje związek pomiędzy wynikami uzyskanymi w wykonywanym przez badanego zadaniu, a określonymi fizycznymi aspektami prezentowanych w trakcie eksperymentu bodźców. Prezentacja ta może odbywać się na wiele sposób. Przez lata standardem psychofizycznym była metoda stałych bodźców. Polega na przedstawianiu badanym serii bodźców o różnych intensywnościach w losowej kolejności (Stevens, 1957). Intensywności bodźca dobiera się tak, żeby obejmowały zarówno poziomy wyraźnie poniżej, jak i powyżej progu percepcji (ustalonego z reguły na 50% prawidłowych powtórzeń). Z uwagi na to, że każdy poziom intensywności bodźca jest prezentowany wiele razy metoda ta, choć dokładna, jest bardzo czasochłonna. Aby skrócić czas badania i uniknąć wpływu dodatkowych zmiennych takich jak np. zmęczenie osoby badanej, coraz częściej korzysta się z metod adaptacyjnych, w których parametry kolejnej prezentacji warunkowane są odpowiedzią słuchacza na wcześniejszy bodziec (Levit, 1971). Metoda adaptacyjna szerzej opisana jest w dalszej części rozprawy. Takie krzywe modelowane są matematycznie za pomocą kumulanty rozkładu normalnego lub funkcji logistycznej (m.in. Treutwein i Strasburger, 1999).
W przypadku opisywanego w pracy doktorskiej obszaru badań, funkcja ta określa

zdolność słuchacza do zrozumienia mowy w funkcji poziomu jej prezentacji lub, jeśli mowa prezentowana była w towarzyszeniu maskera, stosunku sygnału do szumu (SNR). Często wyznaczanymi parametrami krzywej są poziomy bodźca niezbędne do uzyskania konkretnego poziomu wydolności (np. 50% zrozumiałości mowy) i nachylenie krzywej, stanowiące iloraz zmiany stopnia zrozumiałości mowy oraz przyrostu/spadku poziomu bodźca lub SNR. Przykładową funkcję psychometryczną przedstawiono na Rys. 15.



Rys. 15: *Przykładowa krzywa psychometryczna z zaznaczoną wartością progową*

Istnieje szereg czynników wpływających na stromość krzywych. Wykazano, że maskowanie mową pozwala uzyskać mniej strome funkcje psychometryczne niż wykorzystanie szumów zmodulowanych amplitudowo bądź stacjonarnych (Dirks i Bower, 1969; Elliott, 1979; Arbogast, Mason, i Kidd, 2005). Nachylenie funkcji psychometrycznej wykazuje tendencję rosnącą wraz ze wzrostem liczby zmodulowanych sygnałów maskujących – zwiększenie liczby maskerów z jednego do dwóch zwiększa nachylenie średnio o 4% na dB, zbliżając do stanu przy wykorzystaniu maskerów stacjonarnych. Spadki spektralne i czasowe, które mogłyby zostać percepcyjnie wykorzystywane zaczynają się wypełniać (Miller, 1947; Cooke, 2006), wobec czego szansa na to, że sygnał docelowy pokryje się czasowo z co najmniej jednym z szumów maskujących staje się większa, a modulacja amplitudowa w „mieszaninie" sygnału docelowego i maskerów staje się płytsza i krótsza. Zmniejszona szansa na zrozumienie mowy prowadzi do zwiększenia nachylenia. Kiedy jednak liczba różnych sygnałów maskujących osiągnie trzy lub cztery, każdy dodatkowy masker ma znikomy wpływ na nachylenie krzywej (Macpherson, 2014). Na stromość uzyskanych funkcji psychometrycznych mogą wpływać również

stopień przewidywalności przebiegu sygnału docelowego (Kalikow i wsp., 1977), podobieństwo sygnału maskującego i maskowanego. W jednym z przytaczanych w pracy badań dotyczących pomiaru wysiłku słuchowego przybliżono praktyczny aspekt przeprowadzenia i wykorzystania oceny nachylenia krzywej psychometrycznej. Badanie polegające na wyznaczeniu nachylania krzywych psychometrycznych przedstawiono w części eksperymentalnej.

5.2.3 SRT (*Speech Recognition Threshold*)

Najczęściej określanym w badaniach zrozumiałości mowy parametrem jest punkt na krzywej psychometrycznej, w którym wynosi ona 50% (zaznaczony również na Rys. 15). Dowiedziono, że przy 50% poprawnej identyfikacji jednostek językowych, słuchacz powinien, przy zachowaniu odpowiedniego, ponadprogowego poziomu intensywności bodźca, percypować znaczenie poszczególnych fraz języka (Świdziński, 2011), a tym samym zrozumieć komunikat. Wynika to z redundancji języka definiowanej jako pewna nadmiarowość informacji – do zrozumienia mowy potocznej nie jest niezbędna pełna zdolność dyskryminacji poszczególnych dźwięków wchodzących w skład wypowiedzi. Elementy, których nie uda się wysłyszeć mogą być „odtworzone" w procesie poznawczym na podstawie kontekstu, czy to wyrazowego (w przypadku niedosłyszenia pojedynczych głosek lub sylab), czy zdaniowego (przy braku zrozumienia wyrazów w obrębie zdania), na który wpływ ma zasobność indywidualnego leksykonu (Giovannone, 2021).
W związku z powyższym, wyznaczenie poziomu, przy którym badany prawidłowo rozpoznaje połowę zaprezentowanego materiału jest zasadne i może być wykorzystywane w diagnostyce oraz ocenie zysku z protezowania słuchu. Wielkość tę określa się mianem progu rozumienia mowy (z ang. *Speech Recognition Threshold*). W przypadku pojawienia się potrzeby zdefiniowania progów słyszenia dla innych ocen procentowej zrozumiałości, stosuje się dokładnie tę samą metodologię, co w przypadku SRT, ale zamieniając 50% na dowolnie wybraną wartość. Obecnie często stosuje się tak skonstruowane miary, np. SRT20 i SRT80, które odpowiadają kolejno: 20% i 80% zrozumiałości mowy. Z parametrów z tych skorzystano również we własnych badaniach opisanych w niniejszej rozprawie w części eksperymentalnej. Określenie procentowej zrozumiałości mowy dla różnych punktów (poniżej i powyżej 50%) pozwala szybko wyestymować nachylenie krzywej psychometrycznej.

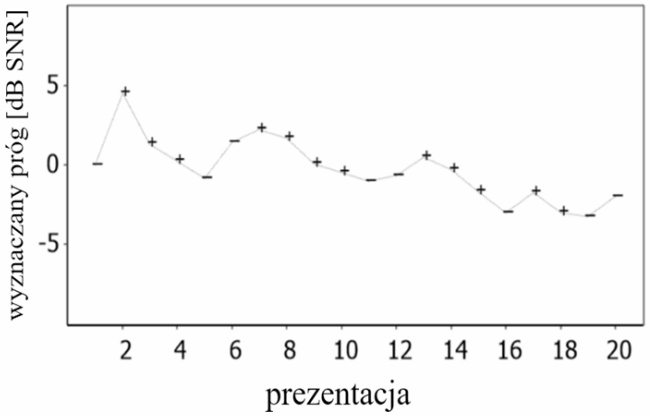
5.2.4 TCT (*Time-Compression Threshold*)

Jednym z rodzajów testów wykonywanych w celu oceny wydajności przetwarzania centralnego są testy mowy przyspieszonej. Zmiany szybkości odtwarzania sygnału mowy powodują zmiany szybkości modulacji amplitudowej, co, przy niewłaściwej rozdzielczości czasowej oraz zaburzonych procesach poza peryferyjnych, powoduje obniżenie stopnia zrozumiałości mowy. W testach mierzących mowę skompresowaną z reguły wyznacza się tzw. próg czasowej kompresji (ang. *Time-Compression Threshold, TCT*), czyli tempo mowy, dla którego 50% prezentowanych słów/zdań zostało poprawnie powtórzonych przez osobę biorącą udział w badaniu. Niemiec i Kociński (2016) w badaniu wykorzystującym polski test zdaniowy matrix wykazali, że średnia wartość TCT wśród młodych, prawidłowo słyszących badanych wynosi 3.7. Rezultat ten jest zbieżny z wynikami otrzymanymi przez Versfelda i Dreshlera przy użyciu innego materiału językowego (2002). W ogólności można przyjąć, że wartość TCT maleje wraz z wydłużaniem się czasu pogłosu. Nie jest to jednak zależność liniowa, jak w przypadku związku TCT i STI. W nim z kolei, im wyższa jest wartość STI, tym bardziej można przyspieszyć mowę zachowując jej zrozumienie na poziomie 50%.

5.2.5 Procedura adaptacyjna

Ze względu na to, że klasyczna metoda wyznaczania poszczególnych punktów krzywej psychometrycznej jest procedurą dość czasochłonną, opracowane inne metody definiowania wartości SRT, które zostały wykorzystane również w badaniach będących przedmiotem niniejszej rozprawy doktorskiej. Najczęściej wykorzystywaną jest procedura adaptacyjna *1-up/1-down*. Dzięki podejściu adaptacyjnym możliwy staje się bezpośredni pomiar SRT, bez konieczności wykreślania krzywej psychometrycznej. Poziom prezentacji kolejnych elementów testu (lub SNR, w przypadku pomiarów w obecności szumu) modyfikowany jest w zależności od odpowiedzi słuchacza. Jeśli powtórzył on prawidłowo określoną, zdefiniowaną w protokole pomiarowym część wygenerowanych jednostek, kolejna prezentacja odbywa się na niższym poziomie (lub przy mniej korzystnym SNR). Poziom prezentacji kolejnego sygnału użytecznego natomiast wzrasta (lub zwiększa się SNR, w przypadku wykorzystania maskera) gdy badany odpowiada błędnie. Po wystąpieniu odpowiedniej liczby tzw. punktów zwrotnych, a więc miejsc, w których następuje zwrot kierunku zmian danego parametru (poziomu sygnały, SNR itp.), badanie zatrzymuje się, a w wyniku uśrednienia ostatnich kilku punktów zwrotnych (ich liczba zdefiniowana jest przez konkretny protokół badania) wyznacza się

próg zrozumiałości mowy. Główną zaletą procedury adaptacyjnej jest możliwość dostosowania poziomu trudności zadania do indywidualnych umiejętności badanych (w tym przypadku zdolności prawidłowego powtórzenia zaprezentowanego materiału językowego), co może prowadzić do uzyskania bardziej dokładnych wyników oraz zwiększenia motywacji uczestników. Niezaprzeczalną korzyścią przemawiającą za wykorzystaniem tego rodzaju metody jest również znaczne ograniczenie czasu potrzebnego na przeprowadzenie pomiaru. Inny psychofizyczny standard – procedura stałych bodźców – pozwala wprawdzie na porównanie wyników między badanymi grupami przy użyciu tych samych bodźców, jednak badanie z jej wykorzystaniem jest znacznie bardziej czasochłonne. Zapis przykładowego przebiegu pomiaru z wykorzystaniem opisanej procedury przedstawia Rys. 16.



Rys. 16: *Przykładowy przebieg wyznaczania progu z wykorzystaniem procedury adaptacyjnej (próg wyznaczony po 20 prezentacjach: -1 dB)*

Rozdział 6

Rodzaje testów wykorzystywanych w audiometrii i badaniach zrozumiałości mowy

Istnieje wiele rodzajów testów audiometrycznych wykorzystujących sygnał mowy. W celach diagnostycznych najczęściej stosowane są testy logatomowe, składające się z elementów nieprzenoszących informacji semantycznej (Iwankiewicz, 1961; Brachmański i in., 1999), testy liczbowe – zapewniające znaczne skrócenie procedury pomiarowej i tym samym niwelujące wpływ „zmęczenia słuchu" (a w przypadku testu trypletowego, dość wiernie odtwarzające progi zrozumiałości mowy (Ozimek, 2009)) oraz testy składające się z list wyrazów jednosylabowych (Pruszewicz i in., 1994). Poniżej przedstawione zostały najważniejsze informacje dotyczące każdego z wymienionych rodzajów testów.

6.1 Testy liczbowe

Mimo braku zrównoważenia strukturalnego, są dość często stosowanym w diagnostyce audiologicznej narzędziem (Piłka, 2015). Ich podstawową zaletą jest prostota oraz to, że pomiar je wykorzystujący nie wymaga dużo czasu. Ze względu na niską czułość (nawet osoby z niedosłuchem stosunkowo łatwo mogą osiągnąć 100% dyskryminacją przy niskich poziomach prezentacji), zaleca się jednak, by testów liczbowych używać w procedurach przesiewowych (Ozimek, 2009), a w celach klinicznych sięgać po metody o większej wartości diagnostycznej (np. testy zdaniowe). Niemniej jednak, testy liczbowe nadal są wykorzystywane, m.in. w badaniach mowy utrudnionej (opisanych również w 6.9.), w ramach których równocześnie prezentuje się do uszu różne sygnały – w tym przypadku

liczby. Ponad 30 lat temu Musiek i in. (1991) wskazywali na użyteczność tak prezentowanych testów liczbowych w przesiewowej ocenie przetwarzania informacji na poziomie ośrodkowym. Również pomiary dychotyczne z użyciem testów liczbowych wykonywane przez Idrizbegovic i in. (2013) wykazały istotne statystycznie różnice w stopniu zrozumiałości mowy między pacjentami z chorobą Alzheimera oraz grupą kontrolną w zbliżonym wieku.

6.2 Testy trypletowe

Składają się trzech cyfr (np. *dziewięć–dwa–pięć*), które badany ma za zadanie powtórzyć w dokładnie tej samej kolejności. Dzięki „odcięciu się" od wpływu kontekstu, narzędzie to może być wykorzystywane wielokrotnie, bez obawy, że słuchacz zapamięta prezentowane próbki. Trudno jest również odgadnąć tryplet w przypadku nieusłyszenia któregoś z jego składowych. Krzywe psychometryczne uzyskane w badaniach trypletami są strome, a wyznaczone wartości SRT cechują się niewielkim odchyleniem standardowym, w przeciwieństwie do klasycznych testów liczbowych, nadal częściej wykorzystywanych w praktyce diagnostycznej (Kociński i in., 2017). Mimo że progi zrozumiałości mowy dla trypletów są nieco niższe niż testów zdaniowych, wyniki uzyskane dla obu procedur są ze sobą silnie skorelowane (Smits i in., 2004). Test trypletowy jest narzędziem dostępnym, a samo badanie jest krótkie, zaleca się jednak, by wykorzystywać je do celów przesiewowych. Ograniczona zawartość fonematyczna ogranicza bowiem jego reprezentatywność danego języka. Jedną z form użycia tego rodzaju testu jest metoda opisana przez Smitsa i zespół (2004), w której tryplety w języku niderlandzkim wykorzystywane były w skriningu słuchowym wykonywanym przez telefon. Badacze ocenili, że ten krótki (ok. 3 minut), adaptacyjny pomiar SRT nie wykazywał wrażliwości na typ telefonu lub warunki odsłuchu, a odnotowane błędy mieściły się w granicach 1 dB. Korelacja zrozumiałości trypletów z progami słyszenia określonymi w audiometrii tonalnej osiągała natomiast imponującą wartość rzędu 0.77.

6.3 Testy sylabowe

Testy sylabowe składają się najczęściej z głosek ułożonych w schemacie CVC (ang. c*onsonant, vowel, consonant* – spółgłoska, samogłoska, spółgłoska). Często zawierają równocześnie struktury pozbawione nacechowania semantycznego, jak i jednosylabowe słowa, tak aby wyizolować zrozumienie fonetycznego,

pozbawione wpływu wskazówek kontekstowych. Jako że zrozumiałość mowy wzrasta wraz ze zwiększaniem się liczby sylab w wyrazie, zastosowanie tego rodzaju narzędzia znacznie utrudnia zrozumienie mowy, również z uwagi na niskoredundantność wykorzystanych jednostek. Z tego powodu wykorzystywano je przy charakteryzowaniu właściwości pomieszczeń oraz ocenie akustycznej kanałów transmisji dźwięku. W ubiegłym stuleciu, w 1910 r., Campbell używał sylab w układzie CV (spółgłoska, samogłoska) w celu weryfikacji jakości sygnału przekazywanego drogą telefoniczną. Obecnie testy sylabowe często zastępowane są narzędziami pozwalającymi wygenerować więcej jednostek słownych (np. logatomami), jednak nadal można je spotkać chociażby w testach składających się z tzw. par minimalnych, a więc struktur różniących się jedną głoską np. *par–por*.

6.4 Testy rymowe

W badaniach wykorzystujących tego rodzaju narzędzie (raczej niestosowane dla języka polskiego) zadaniem słuchacza jest identyfikacja, które słowo zostało wypowiedziane, z listy podobnie brzmiących słów. Z reguły wykorzystuje się metodę dwualternatywnego wymuszonego wyboru, w której osoba badana wybiera jedno słowo z pary, np. do prezentowanego dźwięku /bi/ należy dobrać odpowiedź „bee"lub „pea"(Greenspan i in., 1998). Test rymowy koncentruje się na subtelnych różnicach akustycznych, jest zatem wrażliwy na nawet niewielkie zaburzenia w percepcji mowy lub jakości systemu komunikacyjnego (Voiers, 1983), dlatego chętnie wykorzystywany jest w pomiarach z zakresu akustyki pomieszczeń. Już w 1986 r. Bradley skorzystał z tego narzędzia porównując różne rodzaje parametrów akustycznych w zakresie skuteczności predykcji zrozumiałości mowy w pomieszczeniach o odmiennej wielkości i warunkach propagacji dźwięku. W swoim przeglądzie z 2015 r., Garlińska i wsp. wskazują natomiast, że testy rymowe mogą być wykorzystywane w ocenie zrozumiałości mowy generowanej przez dźwiękowe systemy ostrzegawcze. Testy rymowe dostępne są w wielu językach (m.in. niemieckim, francuskim i niderlandzkim, a ich historia sięga lat 70. i 80. XX w. (Schmidt-Nielsen, 1992).

6.5 Testy logatomowe

Inaczej: pseudowyrazowe lub wykorzystujące słowa nonsensowne, składają się z elementów, które odpowiadają zasadom obowiązującym w danym języku, ale

nie posiadają odpowiednika leksykalnego, a więc nacechowania semantycznego. Badany musi zatem zrozumieć wszystkie występujące w logatomie fonemy, aby udzielić prawidłowej odpowiedzi w teście odsłuchowym. Brak kontekstu czyni tego rodzaju materiał językowy idealnym narzędziem do pomiaru wyrazistości mowy, której oszacowanie może być istotne np. w badaniach dotyczących układów telekomunikacyjnych czy akustyki pomieszczeń. Z testów logatomowych korzystali z powodzeniem m.in. Kociński i Ozimek (2012), oceniając efektywność pętli indukcyjnej w poprawie zrozumiałości mowy użytkowników aparatów słuchowych, czy Błasiński (2023) w próbie określenia związku parametrów obiektywnych pomieszczenia z wyrazistością logatomową.

6.6 Testy wyrazowe

Podobnie jak liczbowe, cechują się dość prostą strukturą, czyniącą je narzędziem dostępnym i gwarantującym szybkie przeprowadzenie pomiaru. Jego ograniczeniem jest to, że krótkie słowa nagrane w izolacji nie zawierają wielu atrybutów naturalnej mowy, m.in. przejść między słowami i intonacji (Nilsson i in., 1994). Niewątpliwą zaletą jest natomiast to, że przy odpowiednim doborze materiału językowego, testy wyrazowe wiernie odzwierciedlają rozkład fonemiczny danego języka. Obecnie funkcjonują w wielu wersjach, a pierwsze, stosowane do pomiarów zrozumiałości mowy z uwzględnieniem sygnału maskującego dla języka angielskiego opracowano w latach 50. i 60. (Fairbanks, 1958; House i in. 1965). Próby stworzenia polskiej wersji testu datuje się na połowę XX wieku, jednak najczęściej stosuje się listy artykulacyjne NLA-93 z początku lat dwutysięcznych (Pruszewicz, 2000). Skrócenie czasu badania i wykorzystanego materiału słownego nierzadko prowadzi do mniejszej dokładności pomiaru. Podobnie jak w przypadku testów liczbowych innych niż trypletowe, uzyskiwane w pomiarach z testami wyrazowymi krzywe psychometryczne cechują się niewielkim nachyleniem (zagadnienie szerzej opisane w części eksperymentalnej.), a wyniki otrzymane dla badanych z prawidłowym słuchem, stosunkowo dużym odchyleniem standardowym. Dążąc do zachowania wiarygodności, jednoznaczności oceny oraz „wysokiej efektywności" w określaniu stopnia zrozumiałości mowy w różnych warunkach, od lat sugeruje się korzystanie z testów zdaniowych (Kollmeier i in., 1997), zwłaszcza, że testy wyrazowe nie nadają się do bardziej zaawansowanych pomiarów, np. w trakcie dopasowywania aparatów słuchowych lub implantów ślimakowych, ponieważ algorytmy kompresji i redukcji szumów nie działają w pełni z pojedynczymi, zazwyczaj krótkimi jednostkami słownymi (Nilsson i in., 1994).

6.7 Testy zdaniowe

Testy można podzielić na dwie główne kategorie – składające się z codziennych wypowiedzi (Plomp i Mimpen, 1979; Smoorenburg, 1992; Kollmeier, Wesselkamp, 1997) oraz poprawne gramatycznie, ale nieprzewidywalne semantyczne zdania (Hagerman, 1982; Wagener i in., 1999) – test typu matrix. Bez względu na rodzaj zastosowanego materiału, testy zdaniowe są dokładniejszym narzędziem służącym do oceny zrozumiałości mowy niż testy wyrazowe (Nilsson i in., 1994; Wagener i in., 2003), co przejawia się bardziej stromymi krzywymi psychometrycznymi. Dodatkową zaletą testów zawierających dłuższe formy wypowiedzi jest możliwość wykorzystania w badaniach szumów zmodulowanych. W przypadku pojedynczych wyrazów pomiędzy którymi występują przerwy, efekt listening in the gaps może nie zostać zaobserwowany.

Należy pamiętać, że aby zapewnić maksymalną wiarygodność badania oraz mieć pewność, że analizowane są przede wszystkim procesy słuchowe, a nie związane z pamięcią osobniczą, czy zasobnością indywidualnego leksykonu, poszczególne elementy testu lingwistycznego powinny być zrozumiałe dla wszystkich, bez względu na wykształcenie, poziom inteligencji, wiek, zawód czy płeć. Spełnienie tego kryterium zapewnia odpowiedni wybór jednostek wyrazowych wchodzących w skład zdań oraz zrównoważenie testu pod względem semantycznym czy fonetycznym.

Testy zdaniowe zostały opracowane dla wielu języków, każdorazowo przy zachowaniu wyżej wymienionych zasad. Jednym z najciekawszych przykładów jest duński test zdaniowy (Danish Sentence Test – DAST) szczegółowo opisany w 2024 r. (Kressner i wsp., 2024). Przy tworzeniu korpusu testu korzystano z bazy uwzględniającej frekwencyjność określonych słów-kluczy w zasobach pisanych (nie istnieje jej odpowiednik dla mowy) obejmującej okres od 1983 do 2024 r. Po przeprowadzeniu szeregu zabiegów uwzględniających m.in. weryfikację rozkładu fonemów, przystąpiono do oceny zaproponowanych zdań udostępniając je ponad 800 osobom, które w siedmiostopniowej skali Likerta miały określić na ile zdanie jest naturalne oraz neutralne (tzn. czy jego znaczenie nie posiada negatywnych konotacji). Dopiero po tak szczegółowej selekcji możliwe było zarejestrowanie wybranych zdań w formie audio oraz, co nadal stanowi rzadkość – wideo. Przygotowane nagrania, przed ostateczną akceptacją, były weryfikowane przez osoby będące natywnymi użytkownikami języka duńskiego, jak i takie, dla których duński jest drugim językiem. Finalnie opracowana baza składa się ze 1200 zdań zarejestrowanych przy udziale dwóch kobiet i dwóch mężczyzn i ma bez wątpienia ogromny potencjał aplikacyjny nie tylko w diagnostyce, ale i badaniach psychoakustycznych, rehabilitacji słuchowej czy ocenie transmisji sygnału.

6.7.1 Polski test zdaniowy (PTZ)

Bazuje na koncepcji Plompa (1979), zgodnie z którą jako materiał językowy wykorzystuje się zdania pochodzące z mowy codziennej, a pierwsze prace nad jego kształtem rozpoczął na początku lat 60. Iwankiewicz (1961). Takie narzędzie ma zdecydowanie większy stopień trudności i większą wartość diagnostyczną w porównaniu z testami liczbowymi. Na rzecz wyróżniającej się trafności fasadowej świadczy również uwzględnienie w wybranych do zbioru testowego zdaniach warunków zrównoważenia fonemicznego. Dodatkowo, za Skrodzką (2014), kryteriami wyboru zdań wchodzących w skład list testowych były neutralność znaczeniowa, poprawność gramatyczna, liczba sylab w zdaniu i wyrazie oraz długość zdania (śr. 4–6 wyrazów na zdanie). Pomiary SRT i nachylenia krzywych psychometrycznych wykazały, że dla słuchaczy ze słuchem prawidłowym, osiągają one wartości zbliżone niezależnie od wykorzystanej listy. Choć pierwotnie stworzono 25 list po 20 zdań, obecnie dostępne są także inne warianty, w których dominują przede wszystkim testy z listami 20-zdaniowymi oraz 13-zdaniowymi. Pomiary z wykorzystaniem PTZ można wykonywać adaptacyjnie zmieniając poziom sygnału (lub stosunek sygnału do szumu) w zależności od odpowiedzi słuchacza, co znacznie skraca czas badania. Test znajduje szerokie zastosowanie nie tylko w diagnostyce słuchu, ale również analizie zrozumiałości mowy w pomieszczeniach, np. z zainstalowaną pętlą indukcyjną (jak u Kocińskiego i in., 2015). PTZ został wykorzystany także w badaniu realizowanym w ramach projektu Alzheimer Prediction Project, które zostało dokładniej opisane w rozdziale E10 części eksperymentalnej.

6.7.2 Test matrix

Jak wykazano, istnieje szereg testów służących do oceny zrozumiałości (lub wyrazistości) mowy. Różnią się one czułością, możliwością wykorzystania, długością pomiaru i stopniem skomplikowania. Choć z założenia realizują zbliżoną ideę, testy dostępne w różnych językach, a nawet w tych samych, ale w różnych ośrodkach klinicznych bądź badawczych, nie są takie same. Znacznie ogranicza to możliwość porównywania uzyskanych wyników i wyciąganie wniosków obejmujących szerszą populację. Tym, co odróżnia rozumienie wyrazów od pełnych zdań jest możliwość wykorzystania kontekstu. Wśród autorów wielu prac przeglądowych nie ma jednoznacznej zgody w kwestii tego, w jaki sposób – pozytywny czy negatywny – wpływa na wiarygodność testów zadaniowych. Chociaż tego rodzaju testy odzwierciedlają naturalny proces komunikacyjny, którego immanentną cechą jest możliwość skorzystania z kontekstu, zdolności poznawcze są cechą jednostkową,

co powoduje dodatkową zmienność wyników. Eliminacja tego aspektu możliwa jest wyłącznie przez zastosowanie testów zdaniowych o nieprzewidywalnej semantyce.
Próbą standaryzacji metody pomiarowej oceniającej zrozumiałość mowy a „odcinającej" się od nadmiernego wpływu kontekstu, stała się formuła po raz pierwszy zaproponowana przez Hagermana, dla języka szwedzkiego (1982). Zakłada zastosowane w ramach badania audiometrii mowy test o strukturze matrycowej, zdefiniowanej zgodnie z określonym protokołem. W późniejszych latach test typu matrix został opracowany dla innych języków m.in. polskiego (Ozimek, 2009, 2010), duńskiego (Wagener, 2003), rosyjskiego (Warzybok i in., 2015) czy mandaryńskiego (Kollmeier, 2015). Obecnie dostępnych jest kilkanaście wersji językowych i ich liczba stale się zwiększa. Test znajduje szerokie zastosowanie w diagnostyce słuchu nie tylko w formie klasycznej prezentacji audio, ale również audio-wizualnej, jak np. w przypadku języka francuskiego (Le Rhun i wsp., 2024) czy niemieckiego (Llorach i in., 2022). Wykorzystuje się go również w ocenie skuteczności metod cyfrowej obróbki sygnału wejściowego w aparatach słuchowych w kontekście poprawy stosunku sygnału do szumu (Deniz i in. (2024) w badaniu dla języka tureckiego), a także ocenie zysku z protezowania słuchu z wykorzystaniem aparatów słuchowych oraz implantów ślimakowych (np. u Willberg i wsp. (2022) dla fińskiej wersji językowej).
We wszystkich wersjach językowych test ma zbliżoną budowę. Na 50-elementową macierz (Rys. 17) składa się pięćdziesięcio wierszowych kolumn, w których znajdują się kolejno: imiona, czasowniki, rzeczowniki, przymiotniki i rzeczowniki. Z każdej kolumny losowo wybierana jest jedna komórka, co pozwala na tworzenie pięcio wyrazowych zdań o stałej syntaktyce (budowie), ale nieprzewidywalnej zawartości semantycznej. Macierz pozwala na wygenerowanie 100 000 kombinacji zdaniowych. W praktycznych zastosowaniach klinicznych z reguły wykorzystuje się listy 20-zdaniowe, natomiast w przypadku konieczności przeprowadzenia bardziej szczegółowych analiz, materiał językowy rozszerza się do 30 zdaniowych list. Dokładność progów szacowanych na podstawie list 20-zdaniowych szacuje się na ok. 1 dB.

Rys. 17: *Test zdaniowy matrix w języku polskim*

Inną istotną cechą test matrix, która stanowi niewątpliwą przewagę tego rodzaju materiału językowego nad wykorzystaniem zdań z mowy codziennej, jest zrównoważenie list. Należy pamiętać o tym, że aby listy artykulacyjne były tak bardzo, jak to tylko możliwe, obiektywne, a otrzymywane wyniki powtarzalne i reprezentatywne, test musi zostać poddany odpowiednim „zabiegom językoznawczym" (Lang, 2003), by spełniać szereg zasad dotyczących m.in. poprawności językowej, częstości występowania określonych struktur w języku czy sposobu nagrania list wchodzących w jego skład. Między innymi z tego powodu obecność w teście słów z określonych grup akustyczno-fonetycznych uwarunkowana jest frekwencją występowania poszczególnych głosek i fonemów w mowie żywej. Sama realizacja akustyczna badania również powinna przeprowadzona z najwyższą starannością, aby zminimalizować wpływ dodatkowych czynników (związanych z głosem mówcy bądź metodami jego rejestracji) na badaną zrozumiałość mowy (Demenko, 2011).

Jak wspomniano, zdania budowane z macierzy testu matrix, w przeciwieństwie do mowy codziennej, cechują się niezmienną, uniwersalną składnią, przy jednoczesnym zachowaniu semantycznej nieprzewidywalności. Z jednej strony słuchacz może więc wykorzystać wiedzę a priori wynikającą ze stałej konstrukcji zdania, z drugiej jednak, z uwagi na niską redundancje nie ma możliwości odgadnięcia treści zdania na podstawie kontekstu (którego nie ma), co zwiększa wiarygodność testu w kontekście oceny zrozumiałości mowy oderwanej od aspektów związanych z biegłością językową czy zasobnością osobniczego słownika.

Swoistą odmianą testu matrix jest Polski Test Matrycowy dla Dzieci, PTMD

(Ozimek i in., 2012). Zasadnicza budowa pozostaje ta sama, jednak, z uwagi na specyfikę grupy docelowej, matryca składa się z trzech kolumn po 16 elementów. Aby uniknąć zupełnej losowości generującej zdania abstrakcyjne semantycznie (np. Kot podlewa ogórki), permutacje elementów testu mogą być wykonywane jedynie w grupach czteroelementowych. Niewątpliwą zaletą testu jest możliwość zastosowania formuły obrazkowej – dzieci są wówcza proszone o wskazywaniu odpowiadających zaprezentowanym zdaniom ilustracji zamiast udzielania odpowiedzi werbalnych. Przykład zaprezentowano na Rys. 18.

Rys. 18: *Przykładowy ekran odpowiedzi w polskim teście matrycowym dla dzieci*

Podobnie jak klasyczny test matrix, PTMD można wykorzystywać do pomiarów zrozumiałości mowy w szumie (Kociński i in., 2017).

6.8 Inne rodzaje testów

Omawiając temat testów językowych wykorzystywanych do oceny stopnia wyrazistości lub zrozumiałości mowy, warto wspomnieć również o testach mowy utrudnionej. Potrzeba ich stosowania wzięła z intensywnego rozwoju metod badania zaburzeń przetwarzania słuchowego (ang. CAPD – *Central Auditory Processing Disorders*), zapoczątkowanego w latach 70. ubiegłego wieku. Ich diagnostyka jest szczególnie utrudniona, ponieważ zwykle nie manifestują się ubytkami w audiometrii tonalnej. Również klasyczna audiometria mowy przeprowadzona w ciszy nie daje jednoznacznych rezultatów. W testach mowy utrudnionej materiał językowy podlega odpowiedniej obróbce w taki sposób, aby pozbawić go redundancji, np. poprzez przefiltrowanie lub przyspieszenie. Wieloletnie obserwacje udowodniły bowiem, że to właśnie takie sygnały mogą być istotnym klinicznie wskaźnikiem sugerującym występowanie deficytów w zakresie przetwarzania centralnego (Wojnowski i in.,

2006). Jako, że, zgodnie z definicją, CAPD dotyczy również takich mechanizmów jak lokalizacja i lateralizacja dźwięku, czasowe aspekty słyszenia czy zdolność rozpoznawania sygnałów akustycznych, zauważalne zaburzenie jednego lub więcej ośrodkowych mechanizmów słuchu, protokół diagnostyczny uwzględnia również testy dychotyczne, w których do każdego z uszu prezentowany jest równolegle inny materiał odsłuchowy lub testy mowy przerzucanej przez Calearo (1963), które polegają na prezentowaniu podzielonych na sylaby pięciu zdań na przemian do prawego i lewego ucha. Zdania różnią się długością (od trzech do dziesięciu sylab) oraz intonacją (część z nich cechuje się wyraźnie rosnącą intonacją). Bez względu na wybraną metodą, nadrzędnym celem testów mowy utrudnionej jest weryfikacja efektywności procesów integracyjnych.

6.9 Wpływ treningu na wyniki

Jednym z najważniejszych procesów związanych z percepcją jest uczenie się. O ile na co dzień stanowi ono podstawę funkcjonowania człowieka w środowisku (pozwalając np. na akwizycję języka, naukę jazdy na rowerze, pływania) i jego relacji względem niego (np. reagowanie na zagrożenia, korzystanie ze sprawdzonych, efektywnych metod działania, unikanie rozwiązań nieprzynoszących pożądanych rezultatów), w badaniach psychofizycznych może być pewnym utrudnieniem. Wprawdzie w pomiarach zdolności poznawczych niezaprzeczalne (i pozytywne) znaczenie ma znajomość przez badanego reguł, jakimi rządzi się eksperyment, jednak trzeba pamiętać o tym, że zbyt intensywny i długotrwały kontakt badanego z materiałem testowym może, przez wielokrotną ekspozycję, spowodować wyuczenie się bodźców na pamięć. Nie inaczej jest w przypadku testu typu matrix, w którym zagadnienie to określa się mianem „efektu treningu". Jak wykazano, obserwuje się go w kilkunastu wersjach językowych tego narzędzia (Kollmeier i in. 2015). Obecność tego zjawiska związana jest zatem bardziej ze strukturą testu i przebiegiem pomiaru, nie zależy natomiast od wykorzystanego języka. Efekt treningu polega on na osiąganiu przez słuchaczy coraz „lepszych" wyników wraz z kontynuowaniem procedury pomiarowej. Jego miarą jest różnica w wartościach progów zrozumiałości mowych uzyskanych przy pierwszej i ostatniej wygenerowanej w tych samych warunkach akustycznych liście testowej. Wielokrotnie szacowano wielkość tego efektu stwierdzając, że największa różnica SRT występuje pomiędzy pierwszą i drugą listą. Wagener i wsp. (1999) wykazali, że wynosi ona ok. 2 dB, a następnie, z każdą kolejną listą, maleje, nie przekraczając 1 dB, co, w przypadku tego rodzaju pomiarów można uznać za wartość zaniedbywalną (lub akceptowalnie małą).

Aby wyniki pomiarów były jak najbardziej wiarygodne, konieczne jest zaprezentowanie list treningowych przed rozpoczęciem procedury wyznaczania SRT (ale również np. wysiłku słuchowego). Dzięki temu niwelowany jest się wpływ opisywanego efektu, (Wagener, 1999) a zatem uniezależnia się otrzymany rezultat od dodatkowych zmiennych. Za wystarczające określono wygenerowanie dwóch list dwudziesto wyrazowych w ramach treningu poprzedzającego badanie właściwe. Takie listy treningowe prezentuje się z reguły w formacie zamkniętym (*closed-set*). Na ekranie komputera lub tabletu wyświetla się macierz testową, która zawiera wszystkie 50 elementów testu, spośród których wybierane są usłyszane przez badanego jednostki słowne składające się na zdanie. Wcześniejsze prace wykazały, że pomimo zaznajomienia się z wyrazami wchodzącymi w skład macierzy, słuchacze nie osiągając wyższej procentowej zrozumiałości mowy, z uwagi na niemożność nauczenia się zdań na pamięć, która wynika z liczby możliwych do wygenerowania kombinacji. Zaprezentowanie nawet kilkudziesięciu list zdaniowych w czasie treningu bez wątpienia nie wyczerpuje zatem limitu dostępnego materiału.
O ile listy treningowe prezentuje się, jak wspomniano, zazwyczaj w formule closed-set, o tyle w praktycznym zastosowaniu klinicznym częściej wykorzystuje się format otwarty (*open-set*). Badany nie widzi wówczas matrycy na ekranie, a jedynie powtarza jej usłyszane elementy. Przebieg procedury pomiarowej jest wówczas szybszy niż w protokole otwartym, a także wygodniejszy, zwłaszcza dla słuchaczy w podeszłym wieku, mogących mieć problem z obsługą manualną klawiatury bądź kursora lub percepcją wzrokową. Z uwagi na to, że stwierdzono występowanie niższych wartości SRT w przypadku prezentacji *closed-set* w porównaniu z formułą otwartą (różnice te wynosiły, w zależności od analizowanego języka, od 0.6 do 1 dB), konieczne jest zaznaczenie w protokole badania sposobu przeprowadzenia pomiaru, zwłaszcza przy zestawiany danych uzyskanych np. dla różnych ośrodków lub wersji językowych.

Testy zrozumiałości mowy, zarówno te wymienione, jak i inne, które zostały poddane odpowiedniej walidacji, choć stanowią istotny element diagnostyki słuchowej, mają również swoje ograniczenia. Warunki laboratoryjne, mimo że dobrze kontrolowane i pozwalające na wykorzystanie bardzo zróżnicowanych scenariuszy testowych, nie są w stanie w pełni odzwierciedlić rzeczywistych sytuacji komunikacyjnych. Przeprowadzana od wielu dekad analiza takich pomiarów była i jest kluczowa dla zrozumienia, dlaczego pacjenci doświadczają trudności w codziennej komunikacji. Trudności te często prowadzą do wzmożonego wysiłku poznawczego, co z kolei wpływa na jakość życia i ogólne funkcjonowanie nie tylko osób z ubytkiem

słuchu. Chęć dokładniejszego zbadania tego zjawiska a tym samym, uczynienie diagnostyki słuchu zagadnieniem bardziej kompleksowym leży u podstaw rozważań na temat szczególnego rodzaju obciążenia poznawczego, jakim jest wysiłek słuchowy. Jego definicję, metody oceny, jak i, finalnie, wyniki dotyczących go badań przeprowadzonych przez autorkę niniejszej rozprawy przedstawiono w eksperymentalnej części pracy.

Rozdział 7

Wysiłek słuchowy

W tym rozdziale zostaną przedstawione definicja wysiłku słuchowego, czynniki mające wpływ na jego wielkość oraz subiektywne i obiektywne metody służące do jego oceny, ze szczególnym uwzględnieniem ACALES – adaptacyjnej kategorialnej skali oceny wysiłku słuchowego (ang. Adaptive CAtegorical Listening Effort Scaling).

Ocena wydolności słuchu (ang. *hearing acuity*) zazwyczaj obejmuje metody takie jak audiometria tonalna lub audiometria mowy. Chociaż są one kluczowe dla diagnozy klinicznej i ewentualnej interwencji medycznej (leczenie, kompensacja niedosłuchu, rehabilitacja słuchowa), nie są w stanie uchwycić i scharakteryzować istotnego aspektu wchodzącego w skład szeroko rozumianego doświadczenia słuchowego – poziomu wysiłku, jaki podejmuje słuchacz, aby osiągnąć swój cel (zrozumieć komunikat słowny).

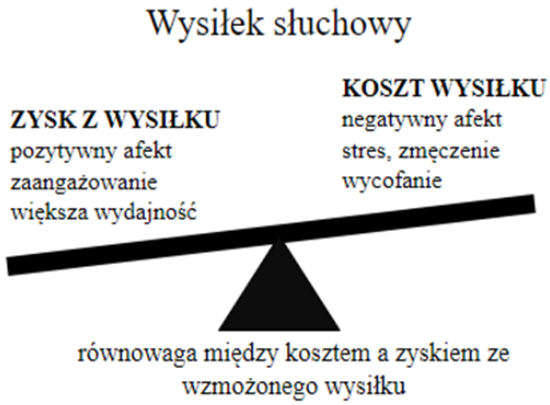
Prawidłowy odbiór, a później zrozumienie mowy są złożonymi zadaniami obejmującymi nie tylko peryferyjny układ słuchowy, ale także wymagającymi zaangażowania procesów poznawczych wyższego rzędu (Johnson i in., 2015), uwagi i pamięci (Davis, 1964; Marslen-Wilson, 1987). W przypadku osób ze słuchem prawidłowym lub z niewielkim niedosłuchem wszystkie niezbędne „operacje zakulisowe", pozwalające na selektywne przetwarzanie określonego dźwięku i jednoczesne odfiltrowanie nieistotnych informacji pełnią działające równolegle mechanizmy centralne (*top-down*) i oddolne (*bottom-up*). W domenie słyszenia, przetwarzanie *bottom-up* nakierowane, jak wskazuje nazwa, na dolny poziom kontroli i opiera na się na przyjęciu wyłącznie informacji akustycznej, bez uwzględnienia innych wskazówek pozasłuchowych. Na rozumienie składa się analiza poszczególnych dźwięków mowy oraz ich sekwencji bez zważania na kontekst, w jakim zostały przedstawione. Podejście *top-down* opiera się natomiast na wykorzystaniu wcześniejszej wiedzy, kontekstu (np. sytuacyjnego – miejsca, w którym toczy się rozmowa), oczekiwań słuchacza oraz informacji o charakterze niewerbalnym

(mimika, ułożenie ciała). Czynniki te, razem z dźwiękami mowy, są analizowane, by finalnie doprowadzić do interpretacji tego, co zostało powiedziane. Prace Moore'a (m.in. 1999) pozwalają stwierdzić, iż czynnikiem istotnie wpływającym na przebieg percepcji mowy jest także znajomość reguł, jakimi rządzi się dany język. Uwzględnia ona czynniki gramatyczne, fonetyczne, semantyczne i inne, np. cechy charakterystyczne mówcy, które pomagają określić prawdopodobieństwo wystąpienia danego fonemu, słowa czy nawet ciągu wyrazów/zdań (Gleason i Ratner, 2005).

Bardziej efektywne funkcjonowanie w sytuacjach wymagających komunikacji językowej, zwłaszcza w trudnych akustycznie warunkach stanowi również mechanizm tzw. „selektywnego wzmocnienia". Koncepcja opisana przez Kerlina (2010) odnosi się do możliwości skupienie się na jednym dźwięku lub rozmówcy pomimo obecności wielu innych sygnałów towarzyszących. Selekcja ta może być realizowana przez różne mechanizmy neuropsychologiczne, takie jak skupienie uwagi czy filtrację sensoryczną, nie jest natomiast fizycznym wzmocnieniem sygnału lub poprawą SNR. Istotność mechanizmu selektywnego wzmocnienia ma silne uzasadnienie ewolucyjne – adaptacja ta pozwalała na szybką identyfikację i interpretację dźwięków w otoczeniu, co z kolei przyczyniało się do zwiększenia szans na przetrwanie. I choć w codziennym funkcjonowaniu okoliczności jej wykorzystania wydają się być mniej dramatyczne, z uwagi na ograniczone możliwości operacyjne mózgu, a także zasobów pamięci osobniczej, wyłonienie istotnych informacji spośród wielu bodźców jest kluczowe dla zachowania wydajności poznawczej na optymalnym poziomie (Lavie, 2005). Nadmiarowość informacji percepcyjnej może bowiem prowadzić do spadku efektywności przetwarzania docierających do receptorów danych, co skutkuje wzrostem wysiłku kognitywnego.

W ostatnim czasie wzrasta zainteresowanie badaniami dotyczącymi obciążenia poznawczego związanego z percepcją mowy, zwłaszcza w warunkach akustycznie wymagających (m.in. Pichora-Fuller i in., 2016; Lemke i Besser, 2016), które określa się jako wysiłek słuchowy (ang. *listening effort*; LE). Przez wiele lat niejednoznacznie opisywane, ostatecznie zdefiniowane w dokumencie Cognition in Hearing Special Interest Group of the British Society of Audiology (McGarrigle i in., 2014) zjawisko rozumiane jest jako „wysiłek umysłowy wymagany do słuchania i zrozumienia komunikatu słuchowego". Rozszerzeniem tej formuły jest bardziej generyczny zapis mówiący o celowej alokacji zasobów umysłowych w celu przezwyciężenia przeszkód w dążeniu do celu podczas wykonywania zadania, w tym przypadku – słuchowego (Pichora–Fuller i in., 2016). Ten zapis koncentruje się na mowie, ale pozwala też włączyć w obszar zainteresowania inne sygnały pojawiające się w warunkach rzeczywistych takie jak np. syrena karetki pogotowia w zakorkowanym centrum

miasta, czy oceniać subiektywnie pojmowaną łatwość lokalizacji źródła dźwięku lub podążania za melodią jednego z instrumentów w trakcie koncertu (Shinn-Cunningham i Best, 2008). Definicja odnotowuje również udział motywacji jako warunku wstępnego przy decydowaniu o realizacji wysiłku. Przykładem może być wykład akademicki. Jeśli jego temat jest interesujący dla słuchacza, będzie on skłonny poświęcić więcej uwagi aby wysłuchać (i zrozumieć) mówcę, nawet jeśli warunki odsłuchu są zaburzone (Tchorz, 2022). Związek ten schematycznie przedstawiono na Rys. 19.



Rys. 19: *Wysiłek słuchowy jako przykład procesu równoważenia zysku i kosztów percepcyjnych*

Wielkość wysiłku słuchowego i możliwość jego kontroli są cechami osobniczą zależnymi od wielu czynników. Kramer i in. (2006) wykazali, że słuchanie w warunkach utrudnionych jest w przypadku osób z niedosłuchem znacznie bardziej obciążające (Kramer i in., 2006; Ohlenforst, 2017) niż w przypadku badanych z prawidłowym progiem słyszenia. Skarżą się one częściej na zmęczenie związane z wyższym poziomem koncentracji wymaganej do zrozumienia mowy w codziennych środowiskach akustycznych (np. w biurze, kawiarni, komunikacji miejskiej). W dłuższej perspektywie może to wywoływać chroniczne uczucie stresu (Hetú i in., 1988), ograniczyć zdolność do wykonywania innych operacji umysłowych w sytuacjach wielozadaniowych (Kociński, 2007; Sarampalis i in, 2009; Anderson Gosselin i Gagné, 2011), co przekłada się na wydolność zawodową, efektywność nauki, ale również obniża komfort w sytuacjach społecznych. Z tego względu tak istotne jest zwrócenie uwagi również na ten aspekt, nie tylko samą zrozumiałość mowy. Dla audiologów i protetyków słuchu wartość określająca stopień wysiłku słuchowego mogłaby bardzo dobrze uzupełniać obecne wykorzystywane narzędzia takie jak audiometria tonalna i audiometria mowy (zwłaszcza prezentowanej w szumie), czyniąc ocenę kliniczną słuchu bardziej kompleksową.

Mimo że rozumienie mowy jest zachowane, zwiększenie stopnia degradacji sygnału i/lub deficyty w jego przetwarzaniu, mogą skutkować wzrostem zmęczenia oraz nakładów uwagi niezbędnych do prawidłowego percepcyjnego wykorzystania sygnału docelowego (Kramer, Kapteyn i Houtgast, 2006; Nachtegaal i in., 2009). Gdy słuchacz będzie wykorzystywał zróżnicowane zasoby niezbędnej uwagi (wysiłku poznawczego), w obserwacji klinicznej możliwe będzie uzyskanie podobnych wartości określających stopień zrozumienia mowy przy bardzo różnych warunkach jej prezentacji (Broadbent, 1958; Rabbitt, 1968; Peelle, 2018). Zasadne jest zatem wnioskowanie, że u podstaw wysiłku poznawczego leży inna zmienna niż w przypadku samej zrozumiałości mowy. Szereg dowodów naukowych potwierdza to założenie. Rudner i in. (2012) wykazali na przykład, że badani oceniali słuchanie prezentowanego materiału językowego jako wymagające coraz mniej wysiłku wraz ze wzrostem wartości opisującej stosunek sygnału do szumu, mimo że stopień zrozumiałości mowy nie ulegał zmianie (poprawie). Oczywiście kwestie percepcji i wysiłku, czy stopnia koncentracji, które jej towarzyszą są ze sobą związane, jednak wydaje się, że wartości wskaźników opisujących jedno i drugie nie są od siebie liniowo zależne.

7.1 Czynniki wpływające na wysiłek słuchowy

Przetwarzanie języka mówionego poprzedzone wyodrębnieniem kluczowych informacji sensorycznych z szybko zmieniającego się sygnału akustycznego wymaga szeregu analiz o charakterze percepcyjnym i kognitywnym (Strand i in., 2018), które wiążą się z określonym wysiłkiem poznawczym. To, jak duży on będzie zależy od wielu czynników. Spośród najbardziej istotnych wymienić należy obecność i rodzaj sygnału zakłócającego (szumu tła), treści i stopnia skomplikowania słyszanej mowy, degradacji sygnału docelowego (np. wskutek niekorzystnych warunków pogłosowych albo nieodpowiedniego przetwarzania sygnału w aparacie słuchowym) charakterystyki mówcy oraz, naturalnie, słuchacza (Mattys, Davis, Bradlow i Scott, 2012). W ogólności stwierdzić można, że w przypadku, gdy sygnał wejściowy jest zniekształcony, zaburzony lub jego odbiór jest w jakiś sposób ograniczony, słuchacze zmuszeni są angażować znacznie większe zasoby poznawcze celem pozyskania użytecznej informacji, niż w przypadku sygnału niezaburzonego, prezentowanego w korzystnych warunkach i rejestrowanego przez nieograniczony aparat odbiorczy (Peelle, 2018).

Poniżej krótko scharakteryzowano najistotniejsze czynniki mogące wpływać na kształtowanie się wysiłku słuchowego. Łatwo zauważyć, że są one podobne (lub niekiedy nawet tożsame) do tych, które wpływają na zrozumiałość mowy, a zostały wymienione w rozdziale 4. Badanym efektem, co należy podkreślić, są jednak inne zjawiska percepcyjne (zrozumiałość mowy i wysiłek słuchowy), które niekoniecznie pozostają ze sobą w liniowej zależności.

a. **Obecność szumów maskujących**: Wyniki wielu badań wskazują, że zgłaszany wysiłek słuchowy jest bardziej wzmożony w przypadku mowy prezentowanej na tle sygnałów zakłócających, niż w ciszy (np. Rudner, Lunner Behrens i in., 2012). Co więcej, uczestnicy eksperymentów psychoakustycznych z reguły zgłaszają większy wysiłek słuchowy wraz ze spadkiem SNR (zmiany w kierunku mniej korzystnym) np. Larsby i in., 2005; Seeman, Sims, 2015; Zekveld, Kramer i Festen, 2010. Potwierdzenie tych obserwacji można znaleźć również w badaniach przeprowadzonych przez autorkę niniejszej rozprawy, opisanych w części eksperymentalnej.

b. **Konkurencyjna mowa:** W przypadku prezentacji sygnału mowy na tle wypowiedzi innych (tzw. *speech-on-speech*[7]) wysiłek słuchowy jest bez wątpienia wyższy niż ciszy (Meister i in., 2023). Dominującym mechanizmem w takich sytuacjach jest maskowanie informacyjne – zarówno sygnał docelowy, jak i maskujący mogą być potencjalnym źródłem użytecznej informacji, stąd ich segregacja i strumieniowanie wymaga dodatkowych nakładów uwagi i wysiłku (Koelewijn i in., 2012). Zasadnicze znaczenie przy segregacji równocześnie występujących sygnałów o nacechowaniu semantycznym ma charakterystyka głosu mówcy, poziom sygnału oraz umiejscowienie źródła dźwięku w przestrzeni. Nie bez znaczenia jest binauralność. Wszystkie te mechanizmy warunkują również istnienie tzw. *cocktail-party effect*.

c. **Zakłócenia na drodze propagacji fali akustycznej:** Obecność różnego rodzaju zakłóceń takich jak echo czy nadmierny pogłos może negatywnie wpływać na stopień zrozumiałości mowy (Festen, i Plomp, 1990), ale także wskaźnik wysiłku słuchowego (Picou, 2016). Szerzej zagadnienie opisano w części eksperymentalnej pracy.

[7]Maskowanie typu *speech-on-speech* (mowa na tle mowy) obejmuje nie tylko zakłócenie na poziomie fonetycznym/fonemicznycm, ale również semantycznym, na co wskazują m.in. badania Tuna i wsp. (2002).W ogólności można powiedzieć, że im sygnał użyteczny (mowa docelowa) jest bardziej podobny do mowy zawartej w sygnale maskującym (np. ten sam język vs różne języki, ten sam poziom treści semantycznej vs różne poziomy treści semantycznej), tym słuchacze mają większe problemy z prawidłowym zrozumieniem zaprezentowanego materiału. Nie bez znaczenia jest również znajomość języka docelowego oraz obecnego w szumie tła (Brouwer, 2012).

d. **Jakość sygnału:** Istotne znaczenie ma jakość samego dźwięku i obecność ew. zniekształceń (np. zniekształcenia w trakcie rozmowy telefonicznej). Jeśli mowa jest niewyraźna, niewłaściwe artykułowana lub akcentowana, prawidłowa percepcja wymaga większego skupienia, co może prowadzić do zwiększenia wysiłku słuchowego (Schiller, 2023).

e. **Wiek:** W miarę starzenia się, zdolność do skupienia się na mowie i filtracji zakłóceń może ulegać pogorszeniu, co może zwiększać wysiłek słuchowy (Rosemann i in., 2020; Kwak, 2021; Kaplan i wsp., 2022), ale są również wyniki badań, które nie wskazują na taką zależność (Rahne, 2023).

f. **Ubytek słuchu:** słuchacze z niedosłuchem wskazują na większy wysiłek słuchowy niż osoby ze słuchem prawidłowym, badane w tych samych warunkach akustycznych (Pichora-Fuller i in., 2016; Hicks i Tharpe, 2002). Takie wnioski wynikają również z badań autorki niniejszej rozprawy przedstawionych w drugiej części pracy.

g. **Korzystanie z aparatów słuchowych (w przypadku niedosłuchu):** zastosowanie algorytmów redukcji hałasu (szumu otoczenia) w aparatach słuchowych może zmniejszyć wysiłek niezbędny do zrozumienia mowy, nawet jeśli nie dochodzi do poprawy jej zrozumiałości (Desjardins i Doherty, 2014; Sarampalis i in., 2009).

h. **Kontekst:** z uwagi na redundantną naturę sygnału mowy, przewidywanie kontekstu zdania zmniejsza wysiłek słuchowy (Hunter, 2022). Wielkość korzyści wynikającej z predykcji semantycznej uzależniona jest m.in. od zasobności indywidualnego leksykonu, wyjściowej pojemności pamięci roboczej, a także znajomości materiału testowego. Z tego ostatniego względu, w badaniach nad wysiłkiem słuchowym prowadzonych przez autorkę, chcąc odciąć się od wpływu kontekstu wykorzystano semantycznie nieprzewidywalne zdania.

i. **Biegłość językowa:** badania prowadzone m.in. przez Ganesha i in. (2011) wskazują na to, że wysiłek słuchowy oraz czas reakcji w badaniach *dual-task paradigm* oceniających zrozumiałość mowy jest większa, gdy słuchacze korzystają z angielskiego jako drugiego języka – wysiłek słuchowy jest większy, gdy badany słabo lub gorzej zna język prezentowanej mowy. Podobne rezultaty w badaniu wykorzytstującym pupilometrię otrzymali Carraturo i in. (2023).

j. **Zmęczenie:** Ogólne zmęczenie lub stres mogą ograniczać efektywne utrzymanie uwagi prowadząc do nadmiernego wysiłku słuchowego (w porównaniu z sytuacją optymalnej alokacji atencji). Zależność ta działa również w drugą stronę. Badania

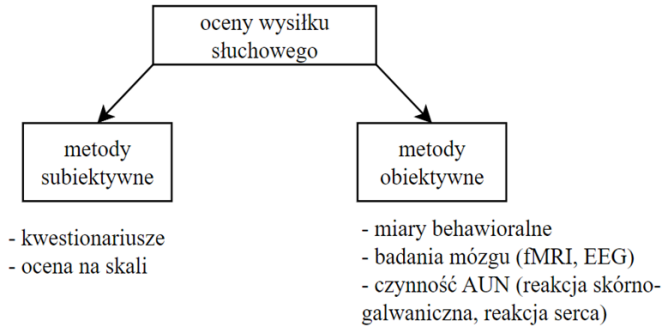
wskazują, że zwiększony wysiłek poznawczy (w tym słuchowy) mogą prowadzić do wzrostu dyskomfortu fizycznego, a także zmian funkcjonalnych układu pokarmowego, krwionośnego lub nerwowego (Mizuno i in., 2011; Solhjoo i in., 2019).

k. **Wielozadaniowość:** W trakcie realizacji zadań poznawczych wymagających jednoczesnego przetwarzania informacji lub zapamiętywania mowy (np. w sytuacjach *dual-task* opisanych niżej) oraz aktywności multimodalnych, wysiłek słuchowy może być wyższy niż w pojedynczych działań unimodalnych, wymagających samego słuchania (Kuchinsky, 2023)

l. **Aspekt wizualny:** obecność wskazówek wzrokowych (np. poruszających się ust osób mówiącej) może zmniejszać wysiłek słuchowy, w porównaniu z sytuacją, w której dostępna jest tylko informacja akustyczna (np. Mishra i in., 2013; Sommers i Phelps, 2016) – więcej informacji, również w odniesieniu do zrozumiałości mowy przedstawiono w podrozdziale 3.1.

7.2 Przegląd istniejących metod pomiaru wysiłku słuchowego

Istnieje szereg metod służących określaniu wartości parametrów opisujących wysiłek słuchowy, zarówno obiektywnych, jak i zależnych od odpowiedzi słuchacza (Klink i in., 2012; Rudner i in., 2012). Schematyczny podział wraz z przykładami przedstawiono na Rys. 20.



Rys. 20: *Schematyczny podział metod oceny wysiłku słuchowego*

Do pierwszych należą m.in. paradygmaty jedno- i dwuzadaniowe oraz pomiary fizjologiczne (np. pupilometria, reakcja skórno-galwaniczna, EEG), natomiast drugich – kwestionariusze oraz ocena na skalach liczbowych lub nominalnych (Larsby i in., 2005). Inne klasyfikacje biorą pod uwagę cechy miar wysiłku słuchowego to

m.in. skuteczność, konieczność (lub brak) wykorzystania specjalistycznego sprzętu i wykształcenia audiologicznego, czułość na niewielkie zmiany oraz możliwość połączenia pomiaru z testem zrozumiałości mowy (Johnson i in., 2015).

7.2.1 Metody obiektywne

a) **EEG**

Hipoteza mówiąca o tym, że wysiłek słuchowy może być powiązany z aktywnością EEG opiera się na założeniu, że mózg wykorzystuje ograniczoną ilość zasobów (neuronalnych) współdzielonych przez procesy sensoryczne, percepcyjne i poznawcze („Teoria ograniczonych zasobów – hipoteza wysiłku"; Rabitt, 1968). Jeżeli przetwarzanie sygnału jest szczególnie wymagające (np. rozmowa w hałaśliwym otoczeniu), więcej zasobów musi zostać przeznaczonych na kodowanie sensoryczne, co prowadzi do mniejszej ilości zasobów dostępnych dla wyższego poziomu przetwarzania – wysiłek słuchowy wzrasta (Bernarding i in. 2012). W swojej pracy Oates (2002), pisząc o wpływie ubytków słuchu na zrozumiałość mowy i wysiłek słuchowy wskazuje na to, że pomiary EEG wykazały, że niektóre obszary mózgu, reprezentujące przetwarzanie poznawcze były bardziej aktywne podczas kompensacji zmniejszonego dopływu aferentnego do kory słuchowej. Aktywność neuronalna mierzona w bezpośrednim powiązaniu z działającym na słuchacza bodźcem (ang. *time-locked EEG activity*) wydaje się też być bardziej czułym i precyzyjnym wskaźnikiem odzwierciedlającym zmiany w pobudzeniu wejściowym (np. hałas w tle lub niedostępna z uwagi na ubytek słuchu informacja) niż miary behawioralne (np. pomiary czasu reakcji) lub subiektywna ocena wysiłku słuchowego (Ohlenfrost, 2017). Badanie EEG jest małoinwazyjne, stosunkowo dostępne, ma dobrą rozdzielczość czasową, ale ograniczoną rozdzielczość przestrzenną. Utrudnia to dokładne lokalizowanie źródeł aktywności mózgowej odpowiedzialnej również za wysiłek słuchowy.

b) **fMRI**

Bazuje na technice MRI (ang. *magnetic resonance imaging*). Badana osoba umieszczana jest w silnym polu magnetycznym (rzędu 1.5–3 Tesli), w którym jądra atomów posiadające słabe właściwości magnetyczne zachowują się jak mikroskopijne magnesy. Dodatkowe cewki wytwarzają krótkie impulsy promieniowania elektromagnetycznego o częstotliwości radiowej (Lauterbur, 1973). Jądra wodoru absorbują energię fal radiowych, zmieniają swój stan, a potem oddają energię emitując fale o tej samej częstości, co określa się mianem rezonansu.

Powrót jonów do pierwotnej pozycji po ustaniu oddziaływania pola magnetycznego (czas relaksacji) jest różny w różnych typach struktur. Urządzenie pomiarowe przetwarza odebrany sygnał na obraz, gdzie różnym czasom relaksacji odpowiadają różne odcienie szarości. Badanie takie pozwala zatem ocenić struktury wewnętrzne i ewentualne nieprawidłowości występujące w ich obrębie. Rozszerzeniem techniki stosowanej w eksperymentach percepcyjnych, również w zakresie oceny wysiłku słuchowego jest badanie funkcjonalne, które umożliwia zbadanie nie tylko budowy, ale także działanie danego narządu czy układu narządów, w tym interesującego w kontekście poniższej pracy, mózgu. fMRI bazuje na rejestracji zmian przepływu krwi, które pojawiają się w odpowiedzi na prezentowane bodźce zewnętrzne. Poszczególne obszary aktywizują się w trakcie wykonywania związanych z pobudzeniem czynności i zużywają wówczas energię czerpaną z procesu spalania glukozy, który zachodzi przy udziale przenoszonego przez krew tlenu. Utlenowana hemoglobina ma inne właściwości magnetyczne niż pozbawiona tlenu, co przekłada się na różne czasy relaksacji jonów, możliwe do zaobserwowania techniką rezonansu. Sygnał ten odbierany jest przez urządzenie pomiarowe i na jego podstawie możliwe jest zlokalizowanie miejsca emisji. Ze względu na wysoką anatomiczną dokładność pomiaru wynoszącą 1–2 mm (Glover, 2011) możliwe jest analizowanie odpowiedzi neuronalnej zarówno na poziomie kory, jak i struktur podkorowych. Badanie wiąże się jednak z dyskomfortem – badany musi przyjąć pozycję leżącą, jest unieruchomiony oraz przebywa w ograniczonej przestrzeni, w której panuje hałas o dużym natężeniu. Czynniki te znacznie ograniczają trafność ekologiczną badania, a więc możliwość odniesienia do warunków rzeczywistych.

c) **fNIRS**

Jest metodą pozwalającą na obrazowanie aktywności mózgu przy wykorzystaniu łatwo przenikającego przez tkanki ciała światła podczerwonego (o długości fal 700–900 nm), która po raz pierwszy została wykorzystana do oceny aktywności ludzkiego mózgu na początku lat dziewięćdziesiątych (Ferrari i in., 2012). W trakcie badania mierzy się odpowiedź hemodynamiczną, a więc regulację tlenu we krwi w której najistotniejszą rolę pełnią silnie pochłaniające światło hemoglobina i odtleniona hemoglobina (Ferrari i in., 2004). Większa aktywność określonego obszaru mózgu (większa aktywność neuronalna) oznacza zużywanie większej ilości energii, a tym samym – zwiększony przepływ natlenowanej krwi, co schematycznie przedstawiono na Rys. 21.

Rys. 21: *Schemat odpowiedzi dynamicznej leżącej u podstaw pomiaru fNIRS*

To, w jakim stopniu światło podczerwone jest odbijane lub pochłaniane przez zawartą w krwi hemoglobinę zależy właśnie od tego, czy przenosi ona cząsteczki tlenu czy nie – w zależności od tego czynnika zmienia się widmo absorpcji (Bonetti i in., 2019).

Pomiaru ilości odbitego światła dokonuje się poprzez rozmieszczenie źródeł i detektorów na czepku, który następnie jest zakładany na głowę badanego, który wykonuje określone zadanie. Metoda fNIRS nie jest tak dokładna jak fMRI w zakresie lokalizacji źródeł aktywności. Nie pozwala bowiem na pomiar w obrębie głęboko położonych struktur - rozdzielczość przestrzenna jest z reguły rzędu 1–2 cm, dlatego precyzyjne analizy możliwe są jedynie na poziomie struktur korowych (Steinbrink i in., 2006). Jest jednak dużo tańsza. Aspekt ten pozwala na bardziej efektywne wykorzystanie w badaniach przesiewowych lub wymagających udziału wielu słuchaczy, również dzieci. Ponadto pomiary wykorzystujące światło podczerwone nie są obarczone ryzykiem dla osób z wszczepionym elementem metalowym (np. zastawką serca lub implantem ślimakowym), jak ma to miejsce w przypadku badania MR (Olds i in., 2016). Nie wymaga także utrzymania nieruchomej pozycji w trakcie całego pomiaru.

d) **Pupilometria**

Powszechnie wiadomym jest, że rozmiar źrenicy zmienia się (dzięki aktywności mięśni w tęczówce) w odpowiedzi na zmiany zachodzące w środowisku, np. poziom oświetlenia otoczenia. Taka reakcja, co interesujące, może pojawić się również w trakcie przetwarzania poznawczego, pobudzenia i zwiększonego zainteresowania (Kahneman, 1973). Zazwyczaj im wyższy poziom pobudzenia lub zainteresowania, tym większy rozmiar źrenicy. Zekveld i in. (2011) wykazali, że u osób starszych z niedosłuchem zmniejszenie wielkości źrenicy wraz z poprawianiem się warunków odsłuchowych jest dużo wolniejsze niż w przypadku młodych, prawidłowo słyszących badanych, co można wiązać z mniej dynamicznym obniżeniem wysiłku słuchowego. Udowodniono, że zmiana rozmiaru źrenic zależy od warunków akustycznych w trakcie badania zrozumiałości mowy, zwłaszcza od rodzaju wykorzystanego sygnału maskującego. W przypadku maskowania informacyjnego maksymalne

rozszerzenie źrenicy jest większe niż w przypadku maskowania energetycznego (Koelewijn i in., 2012). Mimo że pupilometria jest metodą chętnie używaną, zarówno w badaniach klinicznych (Zekveld i in., 2011), jak i nieklinicznych (Kramer i in., 2012), interpretacja uzyskanych wyników jest bardzo problematyczna, a sam pomiar musi być przeprowadzony w idealnie kontrolowanych warunkach. Brak również „złotego standardu" który wskazywałby wartości referencyjne (Alhanbali i in., 2019).

e) **Reakcja skórno-galwaniczna**

Stopień wilgotności skóry, uwarunkowany ilością potu, wpływa na jej zdolności do przewodzenia prądu elektrycznego (Boucsein, 1992), co stanowi podstawę reakcji skórno-galwanicznej. Jej pomiar jest wykonywany w okolicy ekrynowych gruczołów potowych, które znajdują się głównie na dłoniach lub podeszwach stóp. Podobnie jak w przypadku wielkości źrenic, zmiany w ilości wytwarzanego potu odbywają się są pod wpływem aktywności współczulnej części autonomicznego układu nerwowego, związanej ze zmianami w pobudzeniu (Martini i in., 2003). W badaniu wykorzystuje się miernik magnetoelektryczny nazywany galwanometrem, który pozwala na rejestrowanie zmian w oporze elektrycznym skóry w trakcie przepływu słabego prądu elektrycznego (metoda Féré'go). Spodziewaną reakcją jest zmniejszanie oporu wraz ze zwiększaniem się potliwości skóry. Wielkością mierzoną jest odwrotność oporności, czyli przewodnictwo wyrażane w simensach. W otrzymanym zapisie przewodnictwa skóry można wyróżnić składową stałą oraz składową zmienną, która reaguje na zmienne bodźce zewnętrzne. Tę ostatnią określa się mianem GSR (ang. *galvanic skin response*). Amplituda GSR (wyznaczana względem składowej stałej) zależy od ilości potu dostarczonego do przewodów potowych oraz liczby pobudzonych gruczołów potowych. Jednym z głównych problemów związanych z wykorzystaniem technik oceny przewodnictwa skóry jest to, że mierzone jest pobudzenie emocjonalne, bez wskazania jego wartości, a więc tego, czy doświadczenie badanego jest pozytywne (malejący wysiłek słuchowy) czy negatywne – wzrost wysiłku (Mackersie i in., 2011). Brak „złotego standardu" jest także jedną z przyczyn, dla których badania tzw. „wykrywaczem kłamstw", w skład którego wchodzi między innymi ocena reakcji skórno-galwanicznej nie jest uznawany w praktyce sądowniczej za jedyny wiążący dowód (art. 171 par. 5 k.p.k. oraz art. 192a par. 2 k.p.k.).

f) **Miary behawioralne**

Metody dotyczące oceny zachowań (ang. *dual task paradigm*) polegają na jednoczesnym wykonywaniu dwóch zadań – głównego, w tym przypadku

związanego z rozpoznawaniem prezentowanego materiału językowego oraz zadania drugorzędnego (np. obserwacji bodźca wizualnego i naciśnięcie przycisku albo zapamiętywania słyszanej mowy). Poziom trudności zadania głównego jest systematycznie zmieniany, na przykład poprzez wprowadzenie szumów maskujących lub bardziej wymagającego testu (np. logatomowego). Określenie w jaki sposób trudność zadania głównego wpływa na wydajność w zadaniu drugorzędnym dostarcza informacji na temat wykorzystywanych zasobów poznawczych, ich alokacji oraz towarzyszącego wysiłku słuchowego (Gosselin i in., 2010).

Oprócz wyżej wymienionych istnieją także inne fizjologiczne miary poziomu wysiłku np. oznaczanie poziomu kortyzolu z próbek krwi lub śliny, jednak są one, przynajmniej w zakresie audiologii, rzadziej stosowane (Hicks i Tharpe, 2002).

7.2.2 Metody subiektywne

Pierwsze próby wykorzystania metod samooceny w ocenie wysiłku poznawczego, który nie ograniczał się wyłącznie do modalności słuchowej, podjęła NASA. Pod koniec lat 80. Hart i Staveland (1988) zidentyfikowali czynniki, które mogły wpływać na różnice w subiektywnym odczuciu obciążenia pracą przy wykonywaniu różnych typów zadań. Opracowali wielowymiarową skalę oceny. Czynnik „wysiłek umysłowy/zmysłowy"został na przykład opisany jako „ilość wymaganej aktywności umysłowej i/lub percepcyjnej (np. myślenie, podejmowanie decyzji, kalkulowanie, zapamiętywanie, szukanie itp)". Osoby badane miały za zadanie umieścić znak na niepodzielonej graficznie skali pomiędzy granicami „niski" i „wysoki". W kolejnych latach trwały prace nad ustandaryzowaniem narzędzi służących do samodzielnej oceny wysiłku, tym razem z podziałem na określone obszary percepcji, w tym słyszenie. Subiektywne metody pomiary wysiłku słuchowego są realizowane z reguły poprzez proszenie słuchaczy o odpowiedź na pytanie o to, ile wysiłku wysiłku bądź koncentracji jest wymagane do wykonania powierzonego zadania - zrozumienia prezentowanej mowy, czasem jej powtórzenia lub zapisania. Badani zazwyczaj odpowiadają przy wykorzystaniu skali (np. Larsby i in., 2005). W ocenie wysiłku słuchowego przydatne mogą być również kwestionariusze. Słuchacze mają za zadanie na przykład określić, jakie sytuacje z codziennego życia są dla nich szczególnie problematyczne i wymagające większego skupienia. Jednym z zastosowań takiego podejścia może być ocena zysku z dopasowania aparatów słuchowych – porównuje się wyniki ankiety przeprowadzonej przed i po zaprotezowaniu, aby uzyskać inny niż sama zmiana stopnia zrozumiałości mowy, miary.

Inną metodą oceny wysiłku słuchowego jest mierzenie współczynnika *word recall*. Badanym prezentowana jest mowa, często w zmieniających się warunkach akustycznych, a następnie proszeni są oni o przywołanie fragmentów wcześniej odtworzonego testu odsłuchowego. U podstaw tego podejścia leży założenie, że zrozumienie materiału językowego prezentowanego w bardziej wymagającym środowisku akustycznym (np. w obecności szumów maskujących i niekorzystnym SNR) wiąże się z wykorzystaniem większej ilości zasobów poznawczych, pozostawiając ich mniej na realizację zadania związanego w powtarzaniem słów. W myśl tej zasady, przywoływanie mniejszej liczby słów jest interpretowane jako odzwierciedlenie zwiększonego wysiłku słuchowego (np. Pichora-Fuller, Schneider i Daneman, 1995; Sarampalis i in., 2009).

W przypadku badań opisywanych w niniejszej rozprawie, wykorzystywano skalę kategorialną ACALES (ang. *Adaptive CAtegorical Listening Effort Scaling*), która została szczegółowo opisana w części eksperymentalnej. Choć metody obiektywne, zwłaszcza neurofizjologiczne, kojarzone są z większą wiarygodnością i powtarzalnością, Johnson i in.. (2015) oraz Holube i in. (2016) wykazali, że zastosowanie subiektywnej skali kategorialnej do oceny wysiłku słuchowego również pozwala otrzymać wiążące rezultaty. Co więcej, metoda ta jest szybka i łatwa do wykorzystania nawet w warunkach nie-klinicznych, a więc choćby w gabinecie protetyka słuchu lub nawet u pacjenta w domu. Możliwe jest również zróżnicowanie wyników uzyskanych w różnych (bardziej i mniej sprzyjających) warunkach akustycznych, czego nie da się osiągnąć w przypadku pomiaru przewodnictwa skóry (reakcja skórno-galwaniczna). Praktycznym efektem jej użycia mogą być m.in. indywidualna ocena korzyści płynących z zastosowania urządzeń wspomagających słyszenie w wymagających akustycznie warunkach (Krueger, 2019), ocena wpływu cech charakterystycznych sygnałów zakłócających na komfortowe rozumienie mowy lub wpływ cech pomieszczenia (np. czasu pogłosu) na stopień skupienia wymagany do prawidłowej percepcji mowy. Należy jednak mieć na uwadze potencjalne wady tego rodzaju narzędzi – ryzyko *biasu*, niespójność wewnętrzną, indywidualną ocenę, brak uwagi słuchacza lub utrudnioną współpracę z nim (McGarrigle i in., 2014).

CZĘŚĆ EKSPERYMENTALNA

W tej części pracy zostaną przedstawione wyniki badań przeprowadzonych przez autorkę niniejszej rozprawy. Jak sugeruje jej tytuł, centralnym zagadnieniem wiążącym wszystkie eksperymenty była zrozumiałość mowy. Chcąc zachować możliwie jak największą spójność, w pomiarach wykorzystywano najczęściej test zdaniowy typu matrix. Wybór ten podyktowany był kilkoma względami. Po pierwsze, stanowi to naturalną kontynuację badań prowadzonych w ramach pracy magisterskiej a dotyczących walidacji tego narzędzia. Narzędzia, które, warto zaznaczyć, staje się coraz szerzej wykorzystywane, zarówno w praktyce klinicznej, jak i badaniach naukowych, uwzględniających m.in. zaawansowane modelowanie zrozumiałości mowy. Test matrix jest, co podkreślono w rozdziale 6.7.2., narzędziem wysoce wiarygodnym i uniwersalnym. Ze względu na jego strukturę i sposób tworzenia zdań z macierzy wyrazów, możliwe jest wielokrotne wykorzystanie bez obawy o to, że słuchacze nauczą się go na pamięć przy kilku pomiarach. Matrix jest testem zrównoważonym fonematyczne, dlatego dość wiernie odzwierciedla charakterystykę danego języka. Można go także stosować do oceny zrozumiałości mowy w szumie. Dzięki temu uzyskane wyniki da się łatwiej odnieść do rzeczywistych sytuacji komunikacyjnych. Kolejną zaletą testu matrix jest jego zunifikowana forma. Reguły konstrukcji i walidacji kolejnych wersji językowych są znormalizowane, co ułatwia porównanie rezultatów otrzymanych w różnych ośrodkach. Bez wątpienia rozszerza to bazę danych, na podstawie której naukowcy wielu dyscyplin, wywodzący się z jednostek klinicznych i badawczych na całym świecie tworzą coraz bardziej skomplikowane modele ludzkiego słuchu, weryfikując wpływ określonych czynników na jego wydolność, a także opracowują metody kompensacji nie tylko ubytków słuchu, ale i wszystkich dysfunkcji, które mogą ograniczać procesy związane z prawidłowym rozumieniem mowy oraz zwiększać wysiłek towarzyszący percepcji informacji słuchowych.

Te ostatnie aspekty są szczególnie istotne, zwłaszcza w XXI wieku. Niedosłuch (o różnej etiologii i głębokości) oraz inne problemy ze słuchem (np. *tinnitus*) dotykają znacznej części społeczeństwa i należą do problemów zdrowotnych mogących istotnie wpływać na wiele aspektów codziennego funkcjonowania, a szereg naukowców coraz częściej twierdzi, że problem ten jest chorobą cywilizacyjną. WHO przewiduje, że do 2050 r. dysfunkcje słuchu różnego stopnia dotyczyć będą 700 mln ludzi na całym świecie. Nieleczony (gdy to możliwe) lub niewystarczająco skompensowany ubytek słuchu skutkuje trudnościami w komunikacji, izolacją społeczną, obniżeniem nastroju, depresją i ogólnym pogorszeniem jakości życia (Mener i in., 2013; Lawrence i in., 2020; Illg i in., 2023). Aby rozwiązać te problemy, ważne jest dokładne zmierzenie zdolności rozumienia mowy, która jest kluczowym elementem komunikacji.

Biorąc uwagę powyższe argumenty kluczowe wydaje się prowadzenie badań sprzyjających budowaniu wiedzy w zakresie percepcji słuchowej, a także weryfikacji dostępnych lub dopiero wprowadzanych narzędzi służących do jej oceny. Podejście takie jest istotne zarówno w kontekście audiologii i protetyki słuchu (stąd zaangażowanie do badań osób z niedosłuchem), jak i szerzej rozumianej psychoakustyki uwzględniającej akustykę wnętrz (E10) oraz analizę komunikatu językowego wykraczającą poza peryferyjny układ słuchowy (E11). Odzwierciedlając naturalny proces gromadzenia doświadczenia i rozszerzania perspektywy badawczej, ważnym założeniem pomiarów realizowanych w ramach niniejszej rozprawy było to, by uzyskane rezultaty stanowiły informacje wejściowe lub wskazówki do realizacji kolejnych eksperymentów. Bazując na przeglądzie literatury oraz własnych próbach badawczych realizowanych na wcześniejszych etapach edukacji przyjęto następujące hipotezy analizowane w kolejnych rozdziałach części eksperymentalnej:

- **w zakresie testu matrix**
  Polski test zdaniowy matrix jest narzędziem wiarygodnym i czułym (E1), a jego parametry są podobne do tych, obserwowanych dla innych wersji językowych, zwłaszcza w zakresie tzw. efektu treningu (E2) oraz kształtu krzywych psychometrycznych, które można określić w pomiarach z jego wykorzystaniem (E3). Test matrix, który po raz pierwszy został tak szeroko zastosowany i opisany w niniejszej rozprawie, nadaje się do nie tylko do oceny wpływu stopnia niedosłuchu na zrozumiałość mowy (E4), ale również skuteczności metod jego kompensacji (E5).

- **w zakresie osób badanych**
  Percepcja słuchowa sygnału mowy w obecności sygnałów zmodulowanych jest bardziej efektywna niż w przypadku współwystępowania szumów stacjonarnych. W największym stopniu decyduje o tym głębokość niedosłuchu – w pionierskim badaniu wykorzystującym test matrix skupiającym się właśnie na tym zagadnieniu a opisywanym w niniejszej rozprawie, spodziewana jest obserwacja istotnych statystycznie różnic między grupą prawidłowo słyszących oraz badanych z niedosłuchem (E4). Do ważnych zmiennych należy również skuteczność rozdzielczości czasowej i wyższych procesów poznawczych (E8).

- **w zakresie nowych narzędzi**
  Do subiektywnej oceny percepcji słuchowej można wykorzystywać szereg narzędzi, które pozwalają ocenić na skali określony atrybut związany ze słyszeniem (np. jakość sygnału czy wysiłek poznawczy). Definiowane przez nie jednoliczbowe wskaźniki dobrze korelują ze stopniem zrozumiałości mowy

(E6, E7). Nowe metody mogą służyć nie tylko badanym, ale i badaczom – przykładem może klasyfikacja Bisgaarda, grupująca słuchaczy na podstawie ich audiogramów, a powstały w ten sposób model jest dopasowany do faktycznych cech warunkujących właściwą percepcję w zakresie zrozumiałości mowy (E9).

- **w zakresie dodatkowych aspektów związanych z percepcją mowy**
  Na badania percepcji z wykorzystaniem testów językowych jest miejsce nie tylko w praktyce audiologicznej, ale również w ocenie jakości transmisji sygnału mowy w rzeczywistych warunkach jej prezentacji. Jednym z parametrów, które je charakteryzują jest czas pogłosu. Bazując na nagraniach zrealizowanych przez współautora przytoczonego w treści eksperymentalnej badania przyjęto założenie, że im będzie on dłuższy, tym zrozumiałość mowy coraz mniejsza, a współistniejący wysiłek poznawczy – wyższy (E10). Tak, jak pierwsza z hipotez została potwierdzona w wielu badaniach na przełomie ostatnich dekad, tak wysiłek słuchowy jest nowym aspektem badawczym tego zagadnienia, który może w szerszym kontekście rzucić światło nie tylko na problem transmisji sygnału mowy, ale także na jego percepcję w przypadkach, gdy sama zrozumiałość nie zmienia się, ale wysiłek potrzebny do zrozumienia mowy jest inny. Ten dodatkowy parametr może być kolejnym elementem, na który trzeba zwrócić uwagę podczas projektowania pomieszczeń audytoryjnych, a który do tej pory był pomijany. Innym potencjalnym zastosowaniem oceny efektywności percepcji słuchowej w zakresie zrozumiałości mowy jest włączenie jej do charakterystyki czynności poznawczych w prodromalnej fazie choroby Alzheimera. Zespół badawczy, do którego należy autorka niniejszej rozprawy przyjął założenia, że stopień zrozumienia mowy u tej specyficznej grupy badanych będzie niższy niż w grupie referencyjnych (osób bez obciążenia neurologicznego), a możliwość skorzystania z przestrzennego rozseparowania mowy i współtowarzyszącego sygnały zakłócającego, ograniczona (E11).

W pierwszym rozdziale tej części pracy (E1), jako swoiste wprowadzenie do zagadnienia, podsumowana zostanie walidacja polskiego testy zdaniowego matrix. Przyjęta hipoteza zakłada, że jest to narzędzie pozwalające na wiarygodną ocenę progów zrozumiałości mowy zarówno w przypadku osób z niedosłuchem, jak i słuchem prawidłowym w różnych warunkach maskowania. Wiarygodność testu utożsamiona jest w tym przypadku z jego czułością (*sensitivity*).

W kolejnej części (E2) krótko przedstawiono wnioski z oceny wielkości „efektu treningu". Jak wskazuje literatura (więcej szczegółów w podrozdziale 6.9.) jest on immanentną cechą testu matrix, związaną z jego formułą. Do tej pory nie scharakteryzowano go jednak dla języka polskiego, dlatego uczyniono to celem omawianego w rozprawie badania. Przyjęto założenie, że polski test zdaniowy matrix zachowuje się podobnie, jak inne wersje językowe, a proponowany dla nich protokół badań służących ocenie progów zrozumiałości mowy można przełożyć na eksperymenty w języku polskim. Idąc dalej w rozważaniach na temat zasadności wykorzystania testu matrix w badaniach dotyczących percepcji mowy i wchodząc głębiej w jego aspekty techniczne należy zatrzymać się nad zdefiniowaniem i pomiarem nachylenia krzywych psychometrycznych. Wszak to właśnie ono, a nie wartości progowe *per se* określają wzrost korzyści percepcyjnych, które badany może osiągnąć przy niewielkich zmianach SNR. W oparciu o dane pochodzące z pomiarów wykorzystujących inne wersje językowe testu matrix, a także obserwacje z walidacji polskiej wersji założono, że czynnikami, które w największym stopniu wpływają na nachylenie krzywych, a więc to, jak dynamicznie zmienia się procentowa zrozumiałość mowy w funkcji stosunku sygnału do szumu są rodzaj zastosowanego maskera oraz to, jak sprawny jest układ odbiorczy (innymi słowy – czy u badanego występuje ubytek słuchu, a jeśli tak, jak głęboki). Rozdział E4 dotyczy oceny wpływu szumów zmodulowanych na zrozumiałość mowy. Efekt *masking release* (opisany w sekcji 4.1.) jest niezwykle interesującym zagadnieniem, którego oddziaływanie można w sposób pośredni obserwować również w innych badaniach autorki. Celem rozdziału jest synteza podstawowych informacji dot. korzystania przez słuchaczy z luk czasowych w obwiedni sygnału maskującego, a także weryfikacja, jakie czynniki (zewnętrzne lub osobnicze) wpływają na jego wielkość. Przyjęto założenie, że zjawisko *masking release* jest mniej skuteczne w przypadku osób z ubytkiem słuchu – im jest on głębszy, tym zysk ze słyszenia w „lukach" słabszy Warto zaznaczyć, że w przytoczonych badaniach brały udział również osoby z niedosłuchem.

Naturalnym dalszym krokiem na ścieżce badawczej było rozszerzenie analizy stopnia zrozumiałości mowy u tej grupy słuchaczy w różnych warunkach akustycznych. Kolejny zaprezentowany rozdział dotyczy zatem wykorzystania testu zdaniowego matrix do oceny zysku z protezowania słuchu, zarówno przy wykorzystaniu własnych aparatów słuchowych badanych, jak i algorytmu symulującego kompensację niedosłuchu (*Master Hearing Aid*) – E5. To kolejny rodzaj pomiaru, który nie był wcześniej wykonywany dla języka polskiego. Choć badanie miało charakter wstępny, uzyskane wyniki są interesujące i wskazują na konieczność opracowania scenariuszy pomiarowych i metod testowania zrozumiałości mowy, które jeszcze wierniej

odzwierciedlałyby codzienne warunki komunikacyjne, a których immanentną cechą jest obecność sygnałów zakłócających.

Celem protezowania słuchu jest nie tylko poprawa zrozumiałości mowy w sensie ilościowym, ale także ułatwienie komunikacji i lokalizacji dźwięku. To, jak wyglądają one w grupie badanych z niedosłuchem obrazują wyniki badania przedstawionego w części E6. Ono również odnosi się do percepcji mowy w warunkach i sytuacjach akustycznych, w których na co dzień przebywają pacjenci gabinetów protetyki słuchu. Niejednoznaczności wynikające z analizowanych danych mogą wynikać z niewystarczająco precyzyjnie sformułowanego protokołu eksperymentalnego, jednak bez wątpienia są również świadectwem wciąż niewystarczającej wiedzy na temat percepcji słuchowej (zwłaszcza osób z niedosłuchem) i zderzenia liczbowych wskaźników oceniających jej efektywność z subiektywną oceną badanych. Myśląc o subiektywnej ocenie jakościowej oraz komforcie słyszenia nie wolno zapomnieć o zyskującym na znaczeniu wysiłku słuchowym. W przytoczonych w rozdziale E7 badaniach znaleźć można zapis walidacji polskiej wersji skali ACALES służącej do jego oceny. Aby uczynić badanie bardziej kompleksowym, a także utrzymać kontynuację wcześniej prowadzonych eksperymentów, charakterystyce wysiłku słuchowego towarzyszyło określanie stopnia zrozumiałości mowy.

Kolejne przywołane badanie (E8) stanowi rozszerzenie eksperymentu o testy odsłuchowe wykorzystujące mowę przyspieszoną. W jeszcze większym niż wcześniej stopniu weryfikowany jest wówczas udział czynników innych, niż wyłącznie słuchowe, warunkujących to, jak efektywnie przebiega percepcja mowy. W części E9 przedstawiono wyniki weryfikacji skuteczności klasyfikacji wykorzystującej model Bisgaarda (2010). Pierwsze próby takiej oceny podjęto już na etapie pracy magisterskiej. W trakcie realizacji studiów doktoranckich, baza pozyskanych przez autorkę danych znacznie się poszerzyła, a ze względu na obranie w części eksperymentów tej samej metodologii pomiarowej, możliwe stało się połączenie wyników i porównania ich z wnioskami z wcześniejszej pracy dyplomowej.

Uzupełnieniem przedstawionych analiz jest krótkie streszczenie pomiarów prowadzonych przez autorkę w szerszym zespole badawczym. Pierwsze z wymienionych (E10) dotyczy określenia wpływu parametrów akustycznych pomieszczenia na zrozumiałość mowy oraz wysiłek słuchowy, drugie natomiast (E11) jest zapisem pilotażowych badań wieloaspektowej oceny czynności układu słuchu u pacjentów w prodromalnej fazie choroby Alzheimera. Obydwa eksperymenty wykorzystywały inny materiał językowy (kolejno: test logatomowy oraz polski test zdaniowych), osią łączącą pozostaje jednak zrozumiałość mowy i ocena czynników mogących mieć nań wpływ. Fakt opisu wyników tych badań w niniejszej rozprawie stanowi także argument przemawiający za tym, że ocena

percepcji słuchowej, wykorzystująca sygnał mowy, może być stosowana nie tylko w diagnostyce osób z niedosłuchem czy dopasowywania aparatów słuchowych, bądź implantów ślimakowych, ale również szerzej – w weryfikacji rozwiązań technicznych przetwarzających mowę w sposób celowy (np. układy transmisji, poprawy zrozumiałości mowy) lub pośrednio (np. pomieszczenia), a także do oceny innych schorzeń, np. neurodegeneracyjnych, które mimo tego, że bezpośrednio nie są związane ze zrozumiałością mowy w klasycznym rozumieniu (badanie funkcjonowania układu słuchowego), to jednak wyniki tych badań mogą wskazywać na degenerację, a w przyszłości jako badania przesiewowe może nawet na konieczność poszerzenia diagnostyki w kierunku np. chorób otępiennych, czy innych, w których percepcja może ulegać zaburzeniu.

W opisanych w kolejnych rozdziałach badaniach (E1–E9) za każdym razem wykorzystano kilka sygnałów maskujących. Wybór podyktowany był zróżnicowaniem ich charakterystyk, a z drugiej strony, chęcią skorzystania z możliwości porównania uzyskanych wyników z badaniami prowadzonych dla innych wersji językowych tego samego testu, poprzez włączenie tych samych bądź bardzo podobnych (jak w przypadku TSN) szumów do protokołu pomiarowego. Wymienić w tym kontekście należy:

- PolMat, TSN – szum wygenerowany na podstawie losowo wybranych zdań pochodzących z polskiego testu zdaniowego matrix (Ozimek i in., 2010). Posiada to samo długoczasowe średnie widma jak materiał językowy testu i prawie stacjonarną charakterystykę czasową, co czyni go efektywnym maskerem (maskowanie energetyczne).

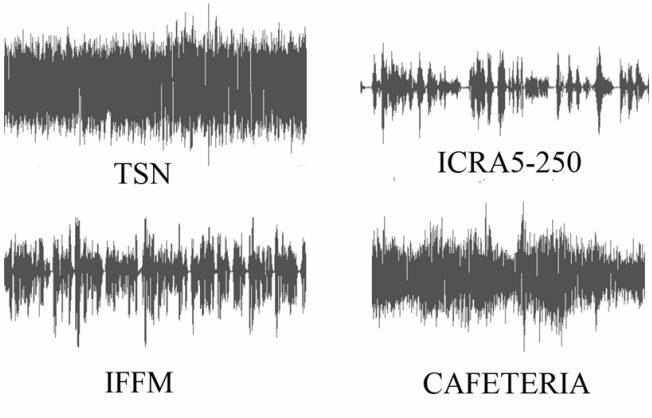
- ICRA5-250 – szum zmodulowany obwiednią sygnału mowy pojedynczego mówcy płci męskiej (Dreschler i in., 2001), zawierający pauzy o długości 250 ms (Wagener i in., 2006);

- IFFM opiera się na ISTS (*International Speech Test Signal*; Holube i in., 2010) stworzonym poprzez nagranie sześciu różnych lektorek czytających bajkę „Północny wiatr i słońce"w swoich językach ojczystych: amerykańskim angielskim, arabskim, mandaryńskim, francuskim, niemieckim i hiszpańskim. Tak przygotowany materiał został poddany segmentacji i losowemu zmiksowaniu, przez co powstały masker, choć zachował charakterystykę czasową i widmową podobną do charakterystyki czasowej i widmowej

jednej mówczyni, jest niezrozumiały. IFFM powstał w wyniku skrócenie maksymalnego czasu trwania pauzy do 250 ms (Holube, 2011).

- Szum cafeteria – najbardziej złożony (pod względem parametrów akustycznych) zastosowany masker, realistyczny szum rejestrowany za pomocą sztucznej głowy KEMAR (Kayser, 2009; Krueger i in., 2017).

Fragmenty przebiegów czasowych sygnałów maskujących przedstawiono schematycznie na Rysunkach poniżej.



Rys. 22: *Fragmenty przebiegów czasowych wykorzystanych sygnałów maskujących – stacjonarnych TSN i cafeteria oraz zmodulowanych ICRA5-250 i IFFM*

Za podstawowy jednoliczbowy wskaźnik stopnia zrozumiałości mowy przyjęto próg 50%, z reguły określany mianem SRT50. Chcąc uprościć zapis i pozostając w zgodzie z podobnymi badaniami opisywanymi przez literaturę zastosowano oznaczenie SRT. W przypadku wartości progowych innych niż 50%, każdorazowo dodano odpowiedni przyrostek, otrzymując np. SRT80 dla progu 80% zrozumiałości mowy.

E1 Walidacja polskiego testu zdaniowego matrix

Mimo iż zalecenia ekspertów podkreślają wartość wykorzystania audiometrii mowy w ocenie wydolności układu słyszenia, badania w zakresie jego diagnostyki opiera się przede wszystkim na audiometrii tonalnej. Jedynie w niektórych wskazaniach medycznych stosuje się dodatkowe testy uwzględniające mowę, ograniczając się jednak często do prezentacji wyrazów w ciszy (Demenko, Pruszewicz, 2011).
Do istotnych aspektów przyczyniających się do ograniczonego wykorzystania audiometrii mowy w audiologii i protetyce słuchu należą bez wątpienia niewystarczająca weryfikacja i walidacja potencjalnie użytecznych narzędzi, które gwarantowałyby możliwość uzyskania wiarygodnych (i powtarzalnych dla danego badanego) wyników. Z tego względu tak ważne wydaje się być podejmowanie prób oceny opracowanych już testów w zakresie z uwzględnieniem potrzeb pacjentów oraz możliwości, jakimi dysponują specjaliści z zakresu diagnostyki i protezowania słuchu. Nawet najbardziej doskonałe narzędzie nie spełni swojej roli, jeśli nie ma warunków do jego praktycznej implementacji.

Wychodząc naprzeciw tym potrzebom przeprowadzono po raz pierwszy z udziałem badanych z niedosłuchem i słuchem prawidłowym, walidację polskiej wersji testu zdaniowego matrix. Za parametry obiektywnie określające jego potencjalną wartość kliniczną uznano czułość (stosunek poprawnie zidentyfikowanych osób z niedosłuchem do całkowitej liczby słuchaczy z niedosłuchem) i swoistość (stosunek poprawnie zidentyfikowanych słuchaczy z prawidłowym słuchem do całkowitej liczby słuchaczy z NH) w wykrywaniu kluczowych aspektów związanych z ograniczeniem percepcji słuchowej. Są to miary uznane w społeczności naukowej, zwłaszcza w badaniach z zakresu medycyny. Warto wspomnieć, że podobne walidacje przeprowadzono w ciągu ostatnich lat dla kilku języków, w tym niemieckiego (Wardenga i in., 2015), francuskiego (Jansen i wsp., 2012) i rosyjskiego (Warzybok i in., 2020). Konsekwentnie wykazywały one wysoką dokładność pomiarów SRT, również u osób z niedosłuchem, wskazując tym samym na przydatność testów matrix w celach diagnostycznych. Niejednokrotnie eksperymenty te koncentrowały się jednak na pomiarach SRT w skrajnie kontrolowanych warunkach maskowania. Zapewnia je zastosowanie szumów stacjonarnych, w szczególności o charakterze *test-specific*, a więc takich, których widmo zazwyczaj pokrywa się ze średnim długoczasowym spektrum sygnału mowy, gwarantując efektywne maskowanie energetyczne (Wardenga i in., 2015; Christiansen, Tau, 2012; Dubno i in., 2002). W omawianym badaniu wykorzystano kilka szumów o różniących się charakterystykach akustycznych (opisanych we wstępie do części eksperymentalnej),

tak aby bardziej kompleksowo zbadać wpływ niedosłuchu i warunków prezentacji sygnału na stopień zrozumiałości mowy.

E1.1 Grupa badana

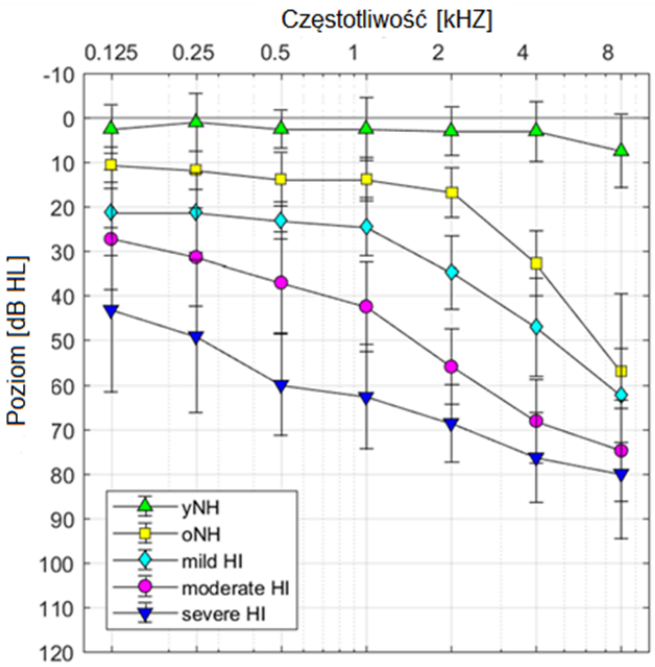
W badaniach walidacyjnych wzięło udział 35 badanych ze słuchem prawidłowym (17 poniżej i 18 powyżej 40 r.ż. oznaczonych jako, kolejno: yNH oraz oNH), a także 57 osób powyżej 40 roku życia z odbiorczym ubytkiem słuchu różnego stopnia, zgodnie z klasyfikacją WHO:

- lekki (średni próg słuchu: 21–40 dB HL) oznaczony jako M (ang. *mild*),
- umiarkowany (średni próg słuchu 41–70 dB HL) oznaczony jako MO (ang. *moderate*),
- znaczny (średni próg słuchu: 71–90 dB HL) – oznaczony jako S (ang. *severe*).

Z uwagi na to, że ubytki słuchu były symetryczne (mediana międzyusznej różnicy PTA wynosiła 4.7 dB) w dalszych analizach brano pod uwagę średnie progi słyszenia ucha „lepszego". W Tab. 1 oraz na Rys. 23 przedstawiono najważniejsze informacje dotyczące badanych.

Tab. 1: *Osoby biorące udział w walidacji polskiego testu zdaniowego matrix (grupy słuchaczy: yNH – młodzi prawidłowo słyszący, oNH – starsi prawidłowo słyszący, M – z lekkim niedosłuchem, MO – z umiarkowanym niedosłuchem, S – ze znacznym niedosłuchem)*

grupa	**N**	**średni wiek (z odchyleniem standardowym)**	**średni próg PTA (z odchyleniem standardowym)**
yNH	18	23.7 (2.6)	3.1 (3.9)
oNH	17	59.3 (12.2)	19.3 (3.9)
M	19	66.3 (8.5)	32.4 (4.5)
MO	27	73.3 (8.9)	50.9 (6.0)
S	11	74.0 (9.5)	66.9 (7.8)



Rys. 23: *Przebieg audiogramów osób biorących udział w badaniu walidacyjnym (grupy słuchaczy: yNH – młodzi prawidłowo słyszący, oNH – starsi prawidłowo słyszący, M – z lekkim niedosłuchem, MO – z umiarkowanym niedosłuchem, S – ze znacznym niedosłuchem)*

E1.2 Materiał i metody

Zachowując metodologię walidacyjną wykorzystaną dla innych wersji językowych testu matrix zastosowano listy 20-zdaniowe prezentowane w formacie *open-set* w protokole adaptacyjnym *1-up/1-down* ze zmieniającą się wielkością kroku (Brand & Kollmeier, 2002). Zadaniem badanych było powtórzenie zaprezentowanych elementów testu.

Aby odzwierciedlić rzeczywiste warunki komunikacyjne, a więc obecność sygnałów zakłócających, pomiary zrozumiałości mowy przeprowadzono w warunkach maskowania, wykorzystując szumy o zróżnicowanej charakterystyce akustycznej – TSN, ICRA5-250, IFFM oraz cafeteria utrzymywane na stałym poziomie 65 dB SPL. Zostały one szczegółowo opisane we wstępie do części eksperymentalnej rozprawy. Za wartość początkową stosunku sygnału do szumu przyjęto 0 dB. Celem ocenienia wpływu wyłącznie ubytku słuchu (bez uwzględnienia wyższych procesów poznawczych), oprócz pomiarów 50% zrozumiałości mowy w szumie zaprezentowano jedną listę testową w ciszy. Pozwoliło to na charakterystykę wydolności komunikacyjnej w najbardziej optymalnych warunkach akustycznych, bez wpływu zakłóceń.

W pomiarach zrozumiałości mowy wykorzystano oprogramowanie Oldenburg Measurement Application (Hörzentrum Oldenburg GmbH, Germany) wraz z kartą dźwiękową EarBox (Auritec) oraz słuchawkami Sennheiser HDA200. Układ został

skalibrowany przy użyciu sztucznego ucha Brüel & Kjær 4153, a badania odbywały się na Wydziale Fizyki i Astronomii UAM w pomieszczeniu spełniającym standardy ANSI S3.1-1999, w ramach grantu Deutsche Forschungsgemeinschaft („Multilingual model-based rehabilitative audiology"; nr 25439187).

E1.3 Wyniki i ich dyskusja

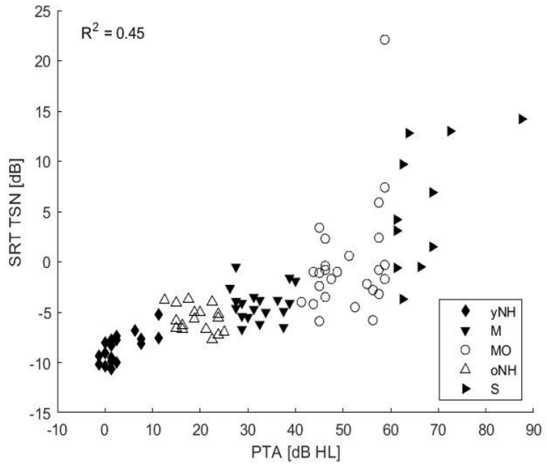
Odnotowano statystycznie istotne różnice między trzema grupami (yNH, oNH i HI) pod względem wieku (test Kruskala-Wallisa, $H(2) = 49.4$, $p < 0.001$) i PTA ($H(2) = 68.2$, $p < 0.001$), gdzie jako HI (ang. *hearing impaired*) oznacza się badanych z niedosłuchem różnego stopnia traktowanych łącznie. Testy Manna-Whitneya, z zastosowaną korektą Bonferroniego, wykazały istotne różnice zarówno pod względem wieku, jak i PTA dla wszystkich porównań ($p = 0.002$ dla porównania wieku między oNH i oHI oraz $p < 0.001$ dla innych porównań).
W Tab. 2 przedstawiono średnie wartości (oraz odchylenie standardowe – SD) 50% zrozumiałości mowy (SRT) we wszystkich badanych warunkach akustycznych.

Tab. 2: *Wartości SRT w szumach maskujących (TSN, ICRA5-250, IFFM, cafeteria) i ciszy wraz z SD (grupy słuchaczy: yNH – młodzi prawidłowo słyszący, oNH – starsi prawidłowo słyszący, M – z lekkim niedosłuchem, MO – z umiarkowanym niedosłuchem, S – ze znacznym niedosłuchem)*

		TSN	**ICRA5-250**	**IFFM**	**cafeteria**	**cisza**
M	średnia	-4.2	-8.1	-8.7	-3.9	32.8
	SD	1.7	3.6	4.1	1.9	8.5
MO	średnia	-0.2	-1.6	-2.0	-1.2	48.0
	SD	5.4	5.3	5.2	3.5	11.5
S	średnia	5.5	5.3	5.4	3.5	68.9
	SD	6.2	5.1	4.7	3.7	9.3
yNH	średnia	-8.5	-21.4	-22.3	-9.9	15.0
	SD	1.4	2.3	1.9	1.1	5.8
oNH	średnia	-5.7	-13.5	-14.3	-5.6	23.6
	SD	1.3	2.4	2.4	1.1	2.2

Jak można zauważyć, różnice pomiędzy średnimi wartościami SRT w szumie stacjonarnym a zmodulowanymi (IFFM i ICRA) są większe w przypadku grup prawidłowo słyszących (zarówno starszych, jak i młodszych) i sięgają nawet kilkunastu decybeli (np. 13.8 dB w TSN vs IFFM dla yNH), natomiast wśród słuchaczy z umiarkowanym (MO) i znacznym (S) ubytkiem słuchu są one niemal niezauważalne (jednoczynnikowa analiza ANOVA z uwzględnieniem HSD Tukeya wykazała, że różnice SRT w szumie stacjonarnym i IFFM dla obu grup nie są znaczące). Wyniki te są zbieżne z danymi przedstawionymi przez Bronkhorsta

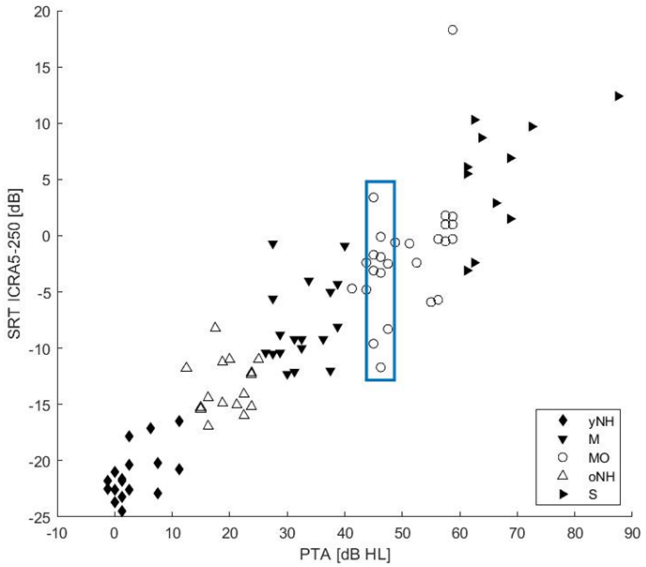
(2000), lecz nieco niższe niż w pracach Wagener i Branda (2005). W przypadku badanych z grupy NH korzyść (a w zasadzie jej brak lub jedynie nieznaczna wartość) wynikająca z pojawienia się fluktuacji w obwiedni maskera (IFFM, ICRA5-250), które sprawiają, że fragmenty prezentowanej na jego tle mowy prezentowane są niejako w ciszy, nie są skorelowane z wiekiem (na podstawie współczynnika korelacji Spearmana; $r_s = 0.44$, $p < 0.0001$). Może być to związane z różną etiologią niedosłuchu lub czasem, który minął od momentu pojawienia się pierwszych jego oznak. Biorąc pod uwagę najniższe wartości SRT uzyskane w szumie IFFM można stwierdzić, że jest on najmniej skutecznym maskerem, jednak różnica między SRT w IFFM oraz ICRA5-250 nie jest statystycznie istotna zarówno dla grupy NH, jak i HI (zgodnie z jednoczynnikową analizą ANOVA; $F = 0.11$, $p = 0.73$). Sugeruje to, iż w tego rodzaju badaniach nie ma konieczności stosowania obydwu maskerów – wyniki uzyskane dla jednego i drugiego są do siebie zbliżone. Pokrywa się to z wnioskami wyciągniętymi przez autorkę niniejszej rozprawy w pracy magisterskiej. Wartości średnich progów zrozumiałości mowy w ciszy dość dobrze korelują ze średnimi progami słyszenia ($R^2 = 0.84$; współczynnik korelacji Pearsona $p < 0.001$). Wyniki audiometrii mowy mogą zatem dość dobrze prognozować to, w jakim stopniu osoba badana będzie rozumiała zaprezentowany materiał testowy (w tym przypadku test zdaniowy matrix). Inaczej rzecz ma się, kiedy mowie towarzyszy sygnał zakłócający. Wprawdzie można zaobserwować tendencję, zgodnie z którą wartości SRT są tym wyższe (tym bardziej korzystny musi być SNR), im większy jest ubytek słuchu (określony jako PTA), jednak korelacja ta nie jest silna (R^2=0.45 dla HI) – Rys. 24.



Rys. 24: *Progi zrozumiałości mowy SRT w szumie stacjonarnym TSN a średnie progi słyszenia PTA (grupy słuchaczy: yNH – młodzi prawidłowo słyszący, oNH – starsi prawidłowo słyszący, M – z lekkim niedosłuchem, MO – z umiarkowanym niedosłuchem, S – ze znacznym niedosłuchem)*

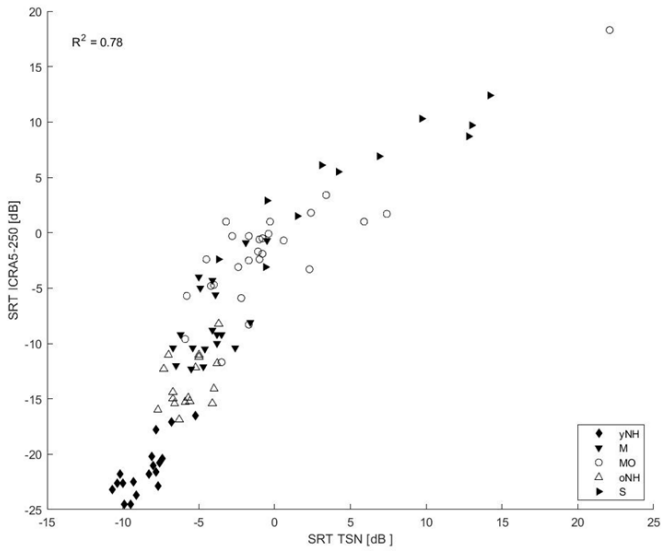
To jednoznaczny argument przemawiający za koniecznością przeprowadzana w trakcie diagnostyki i oceny zysku z protezowania słuchu badań z zakresu audiometrii mowy w warunkach maskowania – wykonując wyłącznie pomiary z wykorzystaniem tonów trudno, nawet ogólnie, przewidywać, jak pacjent radzi sobie w rzeczywistych warunkach komunikacyjnych, które dalekie są od cichego, bezpogłosowego gabinetu protetyka słuchu.

W badaniu z wykorzystaniem polskiego testu zdaniowego matrix, wśród osób z umiarkowanym lub znacznym niedosłuchem zaobserwowano wyraźną zmienność międzyosobniczą – mimo zbliżonych wartości PTA, zrozumiałość mowy osiąga różne wartości. Jest to widoczne np. w zestawieniu średnich progów słyszenia z progami zrozumiałości mowy zmierzonymi w obecności szumu ICRA5-250. W skrajnych przypadkach różnice te są rzędu kilkunastu decybeli, jak zaznaczono na Rys. 25 (osoby o progach słyszenia ok. 45 dB HL uzyskują wyniki SRT w zakresie od -12 do +3 dB).



Rys. 25: *Zależność SRT w szumie zmodulowanym ICRA5-250 od progu słyszenia PTA (grupy słuchaczy: yNH – młodzi prawidłowo słyszący, oNH – starsi prawidłowo słyszący, M – z lekkim niedosłuchem, MO – z umiarkowanym niedosłuchem, S – ze znacznym niedosłuchem); w ramce oznaczono wyniki uzyskane przez badanych z progiem słyszenia ok. 45 dB HL*

Wartości SRT uzyskane w szumie stacjonarnym i zmodulowanym są ze sobą ściśle skorelowane (np. $R^2 = 0.78$ dla TSN i ICRA5-250; współczynnik korelacji Pearsona $p < 0.001$).



Rys. 26: *Zależność SRT w szumach ICRA5-250 i TSN (grupy słuchaczy: yNH – młodzi prawidłowo słyszący, oNH – starsi prawidłowo słyszący, M – z lekkim niedosłuchem, MO – z umiarkowanym niedosłuchem, S – ze znacznym niedosłuchem)*

Kolejny raz manifestuje się zatem obecność dodatkowych deficytów wykraczających poza podniesienie progów słyszenia, które negatywnie wpływają na stopień zrozumiałości mowy prezentowanej w szumie, a które wynikać mogą z poszerzenia filtrów słuchowych, występowania w ślimaku tzw. martwych obszarów, pogorszenia rozdzielczości czasowej i/lub częstotliwościowej oraz innych zaburzeń w przetwarzaniu informacji akustycznej na poziomie wyższych pięter drogi słuchowej lub w jej części ośrodkowej.

Analizując szczegółowo wyniki uzyskane przez osoby z niewielkim lub średnim niedosłuchem (M, MO), które osiągają niskie (<0 dB) wartości SRT w szumie stacjonarnym TSN, można zauważyć, że różnice międzyosobnicze w progach zrozumiałości mowy w szumach zmodulowanych są wyraźniejsze niż w TSN. Fakt ten może sugerować, że wykorzystanie w audiometrii mowy sygnałów maskujących z fluktuującą obwiednią może ułatwiać wychwytywanie nieprawidłowości w zakresie percepcji mowy z większą precyzją niż w przypadku typowego szumu stacjonarnego. To hipoteza, która z całą pewnością wymaga dalszej weryfikacji.

W opisywanym badaniu po raz pierwszy dla języka polskiego określono wiarygodność testu zdaniowego matrix. Za jej mary przyjęto: czułość (*sensitivity*) i swoistość (*specificity*). Na podstawie wcześniej zebranych danych referencyjnych (pomiar z udziałem 18 młodych, prawidłowo słyszących) zdefiniowano punkt odcięcia (ang. *cut-off*), a więc kryterium graniczne pozwalające klasyfikować badanych do grupy NH lub HI. Jego wartość obliczono jako średni próg 50% zrozumiałości mowy (SRT) powiększony o podwójne odchylenie standardowe. Tym sposobem ustalono wartość odcięcia na -6.2 dB (-7.6 + 2·0.7 dB) dla

SRT w TSN i 21.7 dB (16.1 + 2·2.8 dB) dla SRT w ciszy (dla porównania, w rosyjskiej wersji testu zdaniowego matrix są to kolejno -7.2 dB oraz 26.6 dB; Warzybok 2020). Po zastosowaniu kryterium do wszystkich zebranych w badaniu wyników (grupy z ubytkami słuchu różnej głębokości traktowane były łącznie) określono czułość (stosunek poprawnie zidentyfikowanych osób z niedosłuchem do całkowitej liczby słuchaczy z niedosłuchem) i swoistość (stosunek poprawnie zidentyfikowanych słuchaczy z prawidłowym słuchem do całkowitej liczby słuchaczy z NH) o następujących wartościach (Tab. 3).

Tab. 3: *Czułość i swoistość polskiego testu zdaniowego matrix w ciszy i szumie stacjonarnym TSN*

	czułość	**swoistość**
cisza	0.97	1.00
TSN	0.95	0.94

Dla porównania, w rosyjskiej wersji testu zdaniowego w szumie stacjonarnym czułość wynosiła 0.96, a swoistość 0.99 (Warzybok, 2020). Wynik ten świadczy o wyróżniającej się wartości testu zdaniowego matrix – mimo swoich wad i ograniczeń, właściwie zastosowany, gwarantuje uzyskanie wiarygodnych i istotnych klinicznie rezultatów. Mając na względzie to, że w standardowej procedurze diagnostycznej lub czynnościach medycznych związanych z kompensacją niedosłuchu (dopasowanie aparatów słuchowych) czas specjalisty, jak i warunki pomiarowe są znacznie ograniczone, tym istotniejsze wydaje się być sięganie wyłącznie po zwalidowane instrumenty pomiarowe o najwyższej możliwej jakości.

E1.4 Wnioski z eksperymentu E1

Przeprowadzone przez autorkę niniejszej rozprawy i opisane powyżej badanie walidacyjne podkreśla kluczowe znaczenie audiometrii mowy, która powinna być wykonywana nie tylko w ciszy, ale i warunkach utrudnionych, a więc w obecności sygnałów zakłócających. Otrzymane i przedstawione wyniki podkreślają wysoką skuteczność polskiej wersji testu matrix (wykazaną, po raz pierwszy, w badaniu z udziałem osób z niedosłuchem różnego stopnia), na co składają się jego wysokie czułość i swoistość.

Przedstawione spostrzeżenia stanowią kolejny dowód przemawiający za złożonością interakcji między czynnikami słuchowymi i poznawczymi w prawidłowym rozumieniu sygnału mowy. To nie tylko istotne dane naukowe, które stanowią punkt wyjścia do badań przedstawionych w dalszej części pracy, ale, co nawet ważniejsze, informacje, które mają praktyczne implikacje dla praktyki klinicznej.

E2 Pomiary efektu treningu

W podrozdziale 6.9. wspomniano o efekcie uczenia, immanentnej części funkcjonowania człowieka w środowisku, który, choć w ogólności jest cechą niezwykle potrzebną i przydatną, może wnosić pewne ograniczenia w badaniach psychofizycznych. Powtarzana ekspozycja na bodźce testowe prawdopodobnie spowoduje wyuczenie się ich na pamięć, podobnie jest z samą procedurą pomiarową. Niejednokrotnie zaobserwowano, że uczestnicy eksperymentów psychofizycznych (słuchacze, obserwatorzy), do pewnego momentu, uzyskują coraz lepsze rezultaty w miarę kontynuowania badania. Podobnie jest w przypadku testu matrix. Wykazano, że efekt treningu, a więc specyficzna forma uczenia się, występuje dla dotychczas opracowanych wersji językowych testu zdaniowego (Kollmeier, 2015). Do tej pory nie została ona jednak określona dla języka polskiego. Stało się to zatem celem jednego z badań prowadzonych w ramach studiów doktoranckich autorki rozprawy. Wiedza na temat wielkości tego zjawiska jest o tyle istotna, że pozwala usunąć obecność dodatkowego czynnika oddziałującego na stopień mierzonej w badaniu odsłuchowym zrozumiałości mowy.

E2.1 Grupa badana

Do odsłuchów zaangażowano 14 młodych (22–26 lat), prawidłowo słyszących (zgodnie z wytycznymi WHO; średnie progi słyszenia PTA <25 dB HL) osób. Wybór do weryfikacji wielkości efektu treningu badanych należących do grupy referencyjnej (brak stwierdzonych nieprawidłowości w zakresie percepcji sygnałów tonalnych o częstotliwościach 125–8000 Hz) jest zbieżny z metodologią przedstawioną przez Kollmeiera i wsp. (2015) w przeglądzie międzynarodowych wersji testów zdaniowych matrix. Każda z badanych osób była natywnym użytkownikiem języka polskiego i nie miała wcześniejszego doświadczenia udziału w badaniach wykorzystujących test zdaniowy matrix.

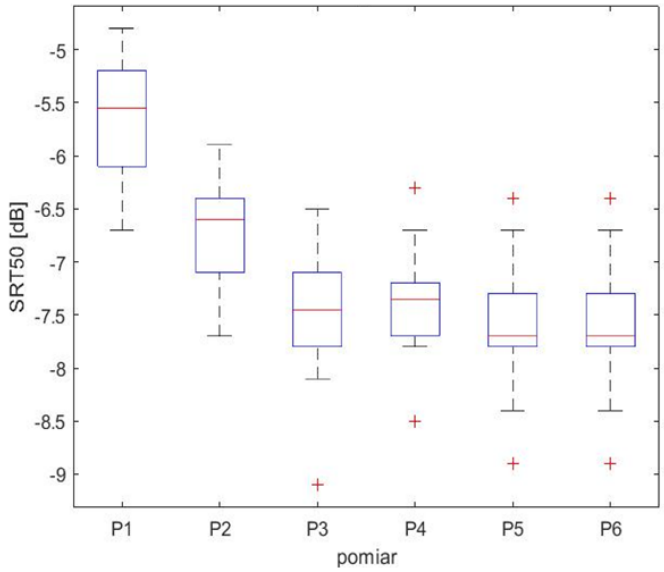
E2.2 Materiały i metody

Aby uzyskać wiedzę na temat efektu treningu i móc, w kolejnych eksperymentach, zniwelować jego wpływ na ostateczne rezultaty, przeprowadzono badania składające się z 6 pomiarów 50% zrozumiałości mowy z użyciem polskiego testu zdaniowego matrix. Podobny protokół badania zastosowano m.in. dla włoskiej wersji językowej (Puglisi i wsp., 2016).

Wykorzystano prostą procedurę adaptacyjną *1-up/1-down* ze zmienną wielkością kroku (Brand i Kollmeier, 2002), a początkowy SNR został ustawiony na 0 dB, przy stałym poziomie szumu stacjonarnego TSN wynoszącym 65 dB SPL. Badani widzieli wszystkie elementy tworzące macierz testową, z których wybierali te, które udało im się usłyszeć (formuła *closed-set*).

E2.3 Wyniki i ich dyskusja

Na Rys. 27 przedstawiono wykresy pudełkowe przedstawiające uzyskane dla poszczególnych pomiarów średnie wartości SRT (wraz ze słupkami błędów). Zaobserwowano, że pomiędzy pierwszym (P1) a ostatnim (P6) pomiarem składającym się z 20-zdaniowych list różnica SRT wynosi 2.4 dB.



Rys. 27: *Zmiany w wartości SRT w kolejnych pomiarach (P1–P5) w szumie stacjonarnym TSN [N=14]*

Największą zmianę (największy udział w kształtowaniu efektu treningu) uzyskano pomiędzy dwoma pierwszymi pomiarami (średnio 1.9 dB). Od trzeciego pomiaru wartości SRT pozostają porównywalne – różnica między trzecią a szóstą serią wynosi zaledwie 0.5 dB. Jednoczynnikowa analiza ANOVA z powtarzanymi pomiarami wykazała główny efekt treningu ($F(5, 65) = 36.5$, $p < 0.001$). Wielokrotne porównania (z poprawkami Bonferroniego) dowodzą temu, że pierwszy i drugi pomiar różniły się statystycznie od pomiarów 3–6 (wszystkie porównania z $p < 0.001$). Nie stwierdzono statystycznie istotnych różnic między pomiarami od P3 do P6. Dane te, z jednej strony, zostały wykorzystane do ustalenia referencyjnego SRT dla prawidłowo słyszących słuchaczy, z drugiej, stanowią dowód

na przedstawione wcześniej stwierdzenie, mówiące o konieczności wykonywania pomiarów treningowych przed zasadniczą częścią badania.Tej zasady przestrzegano przy wszystkich kolejnych pomiarach realizowanych przez autorkę niniejszej pracy.

Rezultaty badania wskazują, że aby wyniki pomiarów zrozumiałości mowy z wykorzystaniem testu zdaniowego matrix były wiarygodne, konieczne jest zaprezentowanie przed częścią główną dwóch list treningowych (tutaj: 20-zdaniowych). Dzięki temu eliminowany jest wpływ efektu treningu, polegającego na uzyskaniu przez badanych lepszych wyników w miarę kontynuowania procedury pomiarowej.

E2.4 Wnioski z eksperymentu E2

Otrzymane przez autorkę wyniki są zgodne z obserwacjami poczynionymi dla innych wersji językowych testu, m.in. Wagener i wsp. (1999). Można więc uznać, że przyjęta hipoteza jest prawidłowa – polski test zdaniowy matrix zachowuje się podobnie, jak jego odpowiedniki w innych językach. Stosowany przy ich wykorzystaniu protokół badania sugerujący konieczność przeprowadzenie przed badaniem zasadniczym dwóch pomiarów treningowych można z całą pewnością przełożyć na eksperymenty w języku polskim, zwiększając wiarygodność ich wyników. Z tego względu w każdym kolejnym badaniu opisywanym w niniejszej rozprawie, zasadniczy pomiar zrozumiałości był poprzedzony prezentacją dwóch list treningowych.

E3 Nachylenie krzywych psychometrycznych

Nachylenie krzywych psychometrycznych (szczegółowo opisane w podrozdziale 5.2.2.) jest kluczowe, ponieważ to ono, a nie wartości progowe *per se*, określają wzrost korzyści percepcyjnych, jakie słuchacz może uzyskać z niewielkich zmian SNR (Ozimek i in., 2009; MacPherson i in., 2014). Stroma funkcja psychometryczna wskazuje, że niewielki wzrost SNR prowadzi do dużego wzrostu zrozumiałości i przeciwnie – jeśli nachylenie jest stosunkowo małe, ta sama zmiana SNR skutkuje mniejszą poprawy wydajności percepcyjnej.

E3.1 Grupa badana

Pomiary zrozumiałości mowy, przy okazji których określono nachylenie krzywych psychometrycznych przeprowadzono na grupie 58 słuchaczy (w tym 18 z symetrycznym niedosłuchem odbiorczym i 40 ze słuchem prawidłowym). Szczegółowe dane dotyczące osób biorących udział w pomiarach przedstawiono w Tab. 4, gdzie yNH oznacza młodszych, prawidłowo słyszących, oNH starszych (>40 r.ż.) prawidłowo słyszących, natomiast HI osoby z niedosłuchem.

Tab. 4: *Charakterystyka słuchaczy biorących udział w badaniu nachylenia krzywych psychometrycznych (w nawiasach wartości odchylenia standardowego); yNH – młodzi prawidłowo słyszący, oNH – starsi prawidłowo słyszący, HI – osoby z symetrycznym odbiorczym niedosłuchem*

grupa	liczba badanych	średni wiek [lata]	PTA [dB HL]
yNH	20	31.8 (6.8)	4.4 (2.6)
oNH	20	58.5 (5.9)	8.9 (3.9)
HI	18	59.9 (15.9)	39.3 (10.3)

E3.2 Materiał i metody

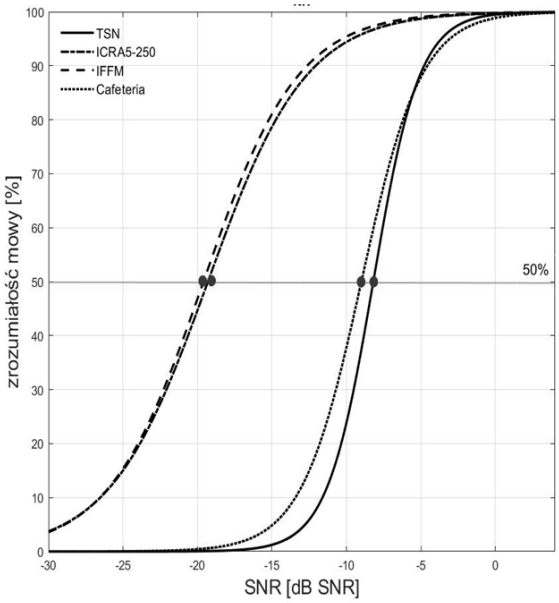
W badaniu dokonano pomiaru zrozumiałości mowy wykorzystując polski test zdaniowy matrix zachowując metodologię opisaną w eksperymencie E1 – listy 20-zdaniowe prezentowane w formacie open-set w protokole adaptacyjnym *1-up/1-down* ze zmieniającą się wielkością kroku (Brand & Kollmeier, 2002). Biorąc pod uwagę rezultaty badania opisanego w rozdziale E2, zasadniczą część pomiaru poprzedzono prezentacją dwóch list treningowych.

Na podstawie wyników oceny 20% i 80% zrozumiałości mowy (kolejno: SRT20 i SRT80) w czterech warunkach maskowania wyznaczono krzywe psychometryczne dla każdego z ustawień pomiarowych., postępując podobnie jak w badaniach

Puglisi (2021). Aby zachować konsekwencję, wybrano maskery, które zastosowano wcześniej w klinicznej walidacji polskiej wersji testu zdaniowego matrix (TSN, ICRA5-250, IFFM, szum cafeteria). Ich właściwości opisano we wstępie do części eksperymentalnej rozprawy.
Nachylenie krzywych psychometrycznych określono przez procentową zmianę zrozumiałości mowy przy 1 dB zmianie SNR lub poziomu prezentacji sygnału mowy w ciszy.

E3.3 Wyniki i ich dyskusja

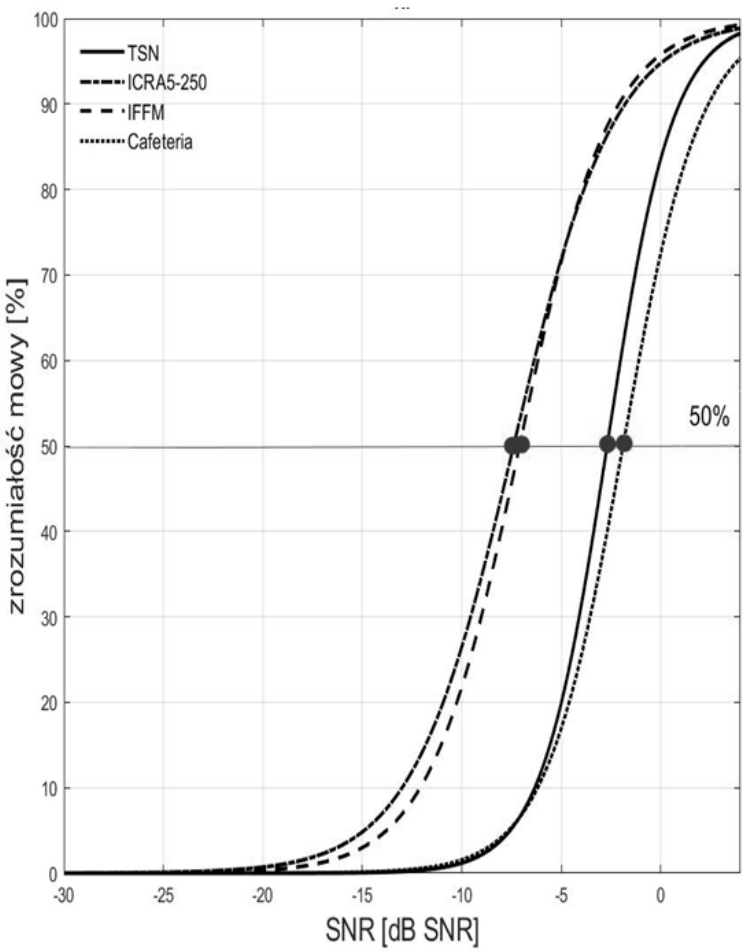
Na Rys. 28 przedstawiono krzywe psychometryczne uzyskane dla grupy yNH we wszystkich szumach maskujących. Wypełnionymi kółkami oznaczono wartości SRT – w tym przypadku jest to stosunek sygnału do szumu (SNR) przy którym badani prawidłowo powtórzyli połowę zaprezentowanego testu.



Rys. 28: *Krzywe psychometryczne z wartościami SRT (czerwone punkty) dla grupy młodych, prawidłowo słyszących badanych (yNH)*

Jak łatwo zauważyć, najniższe wartości uzyskano dla sygnałów zmodulowanych ICRA5-250 oraz IFFM (kolejno -19.3 oraz -17.8 dB SNR), co wynika ze zjawiska *masking release* (opisanego w rozdziale 4.1.). Innymi słowy można stwierdzić, że modulacje amplitudy zwiększają zakres SNR, w którym mowa docelowa pozostaje zrozumiała (co opisywali również Rhebergen i Versfeld, 2005). W przypadku szumów stacjonarnych cafeteria oraz TSN, wartości SRT są nawet kilkanaście decybeli wyższe (-8.2 oraz -9.0 dB). Istotne jest także nachylenie krzywych psychometrycznych – jest ono największe właśnie dla szumów stacjonarnych.

Bardzo podobne relacje można zaobserwować dla grupy badanych z niedosłuchem, z tą różnicą, że wartości SRT dla wszystkich warunków pomiarowych są wyższe (krzywe przesunięte w prawo, w kierunku dodatnich SNR), a zysk percepcyjny wynikający z fluktuacji w obwiedni maskera mniejszy (ok. 10 dB dla yNH *vs* 4 dB dla HI) – Rys. 29.



Rys. 29: *Krzywe psychometryczne z wartościami SRT (czerwone punkty) dla grupy badanych z niedosłuchem (HI)*

Jak wspomniano, nachylenie krzywych estymowane było na podstawie 20% i 80% zrozumiałości mowy. Dokładne wartości je charakteryzujące (wraz z odchyleniem standardowym) przedstawiono w Tab. 5.

Tab. 5: *Nachylenie krzywych psychometrycznych [%/dB] wraz z odchyleniem stand. (SD); yNH – młodzi prawidłowo słyszący, oNH – starsi prawidłowo słyszący, HI – osoby z symetrycznym odbiorczym niedosłuchem*

		TSN	**ICRA5-250**	**IFFM**	**cafeteria**	**cisza**
yNH	nachylenie [%/dB]	16.1	7.6	7.9	12.3	11.1
	SD	3.7	1.4	2.8	3.3	2.5
oNH	nachylenie [%/dB]	16.4	7.8	9.0	12.9	11.5
	SD	3.6	2.0	3.8	3.6	3.5
HI	nachylenie [%/dB]	15.1	9.8	11.0	12.8	11.0
	SD	3.4	3.0	3.0	2.9	2.9

We wszystkich badanych grupach największe wartości nachylenia (a więc najbardziej strome krzywe psychometryczne) uzyskano dla szumu stacjonarnego TSN i wynoszą one, kolejno, 16.1, 16.4 i 15.1 %/dB w przypadku młodszych (<40 lat) i starszych prawidłowo słyszących oraz badanych z niedosłuchem. Wartość nachylenia mierzona była dla punktu 50% zrozumiałości mowy (SRT) z wykorzystaniem funkcji logistycznej, zgodnie z metodologią opisaną przez MacPherson i Akeroyda (2014). Ozimek i in. (2010) przy okazji opracowywania polskiej wersji testu matrix dokonali oceny nachylenia krzywych psychometrycznych, jednak tylko w grupie prawidłowo słyszących w warunkach maskowania szumem stacjonarnym. Otrzymana wartość wynosiła 17.1 %/dB (dla formuły *word-scoring*, która była wykorzystana również w tym eksperymencie). Jest to wynik zbliżony do uzyskanego w badaniu będącym przedmiotem tego opracowania. Nieznacznie niższą wartość uzyskano w grupie słuchaczy z niedosłuchem. W przypadku pomiarów w ciszy oraz szumie cafeteria nie stwierdzono istotnych różnic statystycznych w wartościach nachylenia krzywych w porównaniu wszystkich badanych grup. Różnice te pojawiają się jednak w przypadku szumów zmodulowanych, w których nachylenie krzywych osiąga najniższe wartości Jest to zgodne z doniesieniami literaturowymi. W uproszczeniu można przyjąć, że zmniejszona szansa na zrozumienie mowy prowadzi do zwiększenia nachylenia. Stąd stromość krzywych jest największa w przypadku maskerów najbardziej skutecznych – w tym przypadku jest to szum stacjonarny zapewniający maskowanie głównie o charakterze energetycznym (Dirks i Bower, 1969; Elliott, 1979; Arbogast, Mason i Kidd, 2005), w przeciwieństwie do zmodulowanych sygnałów takich jak wykorzystane w tym badaniu ICRA5-250 oraz IFFM. Jest to

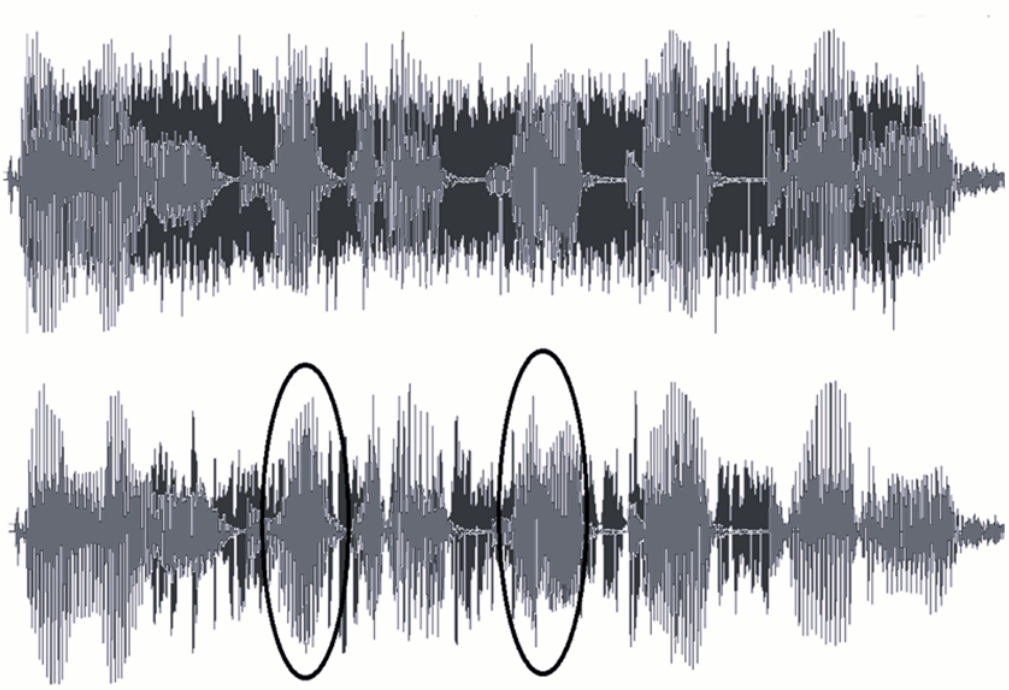
zbieżne z przyjętą hipotezą i pozwala lepiej zrozumieć techniczną stronę oceny słuchu z wykorzystaniem testu matrix w różnorodnych warunkach maskowania.
Wpływ modulacji amplitudowej na nachylenie krzywych psychometrycznych można zrozumieć biorąc pod uwagę zjawisko *listening in the gaps*. Kiedy sygnał docelowy jest prezentowany na tle szumu fluktuującego dochodzi do jego zbiegania się z minimami amplitudy maskera. W tych spadkach (lukach) w obwiedni maskera, lokalny SNR jest zwiększony (bardziej korzystny), umożliwiając słuchaczowi dostrzeżenie (zrozumienie) docelowego sygnału mowy (Miller i Licklider, 1950; Cooke, 2006) – dochodzi do poprawy percepcji mowy, a tym samym obniżenia progi jej zrozumiałości (Miller, 1947; Wilson i Carhart, 1969; Takahashi i Bacon, 1992). Modulacje amplitudy zwiększają zatem zakres SNR, w którym mowa docelowa pozostaje zrozumiała (Rhebergen i Versfeld, 2005), ponieważ „przebłyski" sygnału docelowego występują nawet wtedy, gdy SNR zmniejsza się (staje się mniej korzystny). Skutkiem tej prawidłowości jest właśnie wspomniane już wyżej uzyskiwanie znacznie płytszych krzywych psychometrycznych, niż w przypadku markerów stacjonarnych (Speaks i in., 1967).

E3.4 Wnioski z eksperymentu E3

Analiza otrzymanych w badaniu E3 wyników niewątpliwie świadczy o tym, że istnieje potrzeba zgłębienia zjawiska przy wykorzystaniu polskiego testu matrix prezentowanego na tle szumów maskujących. Stanowi ona także punkt wyjścia badania, które zostało opisane w kolejnym rozdziale (E4). Postępowanie takie jest zgodne z założeniami badań realizowanych w ramach doktoratu – uzyskane rezultaty winny stanowić informacje wejściowe lub wskazówki do realizacji kolejnych eksperymentów.

E4 Ocena wpływu szumów zmodulowanych na zrozumiałość mowy – *masking release*

W eksperymencie dotyczącym nachylenia krzywych psychometrycznych, który został opisany powyżej (E3) potwierdzono, że sygnały zmodulowane są mniej efektywnymi maskerami niż typowy szum stacjonarny. To istotny aspekt w procesie walidacji test matrix, która, po raz pierwszy, została opisana w niniejszej rozprawie. Jest to kluczowa informacja również z punktu widzenia audiometrii mowy – wykorzystanie innych niż standardowy szum stacjonarny maskerów pozwala pełniej scharakteryzować wydolność układu słyszenia, zwłaszcza w kontekście zrozumiałości mowy.
Schematy ilustrujące mechanizm zjawiska przedstawiono na Rys. 30.



Rys. 30: *Zdanie „Tomasz wygra kilka czarnych okien" (szary) zamaskowane szumem stacjonarnym TSN (górny wykres) oraz szumem zmodulowanym IFFM (dolny wykres)*

W górnej części ilustracji przedstawiono graficzną reprezentację nagrania zdania „Tomasz wygra kilka czarnych okien" (kolor szary), na które nałożono szum stacjonarny TSN (kolor czarny). Jak widać, sygnał użyteczny jest niemal w całości zamaskowany przez szum stacjonarny (głównie maskowanie energetyczne), a czynnikiem decydującym w największej mierze o stopniu zrozumiałości mowy jest poziom mowy (lub, dokładniej, stosunek sygnału do szumu). Na drugim wykresie temu samemu zdaniu towarzyszy równoczesna prezentacja szumu zmodulowanego

IFFM. Widoczne są fragmenty sygnału mowy, które pozostają wolne od maskowania. Ich część przypada na spadek amplitudy sygnału zakłócającego związany z fluktuacją jego obwiedni (przykładowe fragmenty oznaczono czarnymi elipsami). Bez wątpienia można stwierdzić, że efektywność maskowania, na poziomie czysto akustycznym, różni się w zależności od wzajemnych relacji sygnału maskującego oraz maskera. W tym przypadku jest ona mniejsza, gdy mowa maskowana jest sygnałem silnie zmodulowanym, co związane jest z występowaniem tzw. luk czasowych (ang. *temporal dips*). Otwartym pozostaje jednak pytanie, jak tę prawidłowość fizyczną można wykorzystać percepcyjnie, zwłaszcza, gdy aparat odbiorczy nie jest w pełni sprawny. Aby na nie odpowiedzieć, a tym samym stwierdzić, w jakim stopniu osoby z niedosłuchem w różnym wieku są w stanie (lub nie) wykorzystać te luki, przeprowadzono pomiary SRT w szumie stacjonarnym (TSN) oraz zmodulowanym ICRA5-250, którego właściwości w kontekście *masking release* są zbliżone do szumu IFFM.

E4.1 Osoby badane

Osoby biorące udział w badaniu zostały przyporządkowane do określonych grup, w zależności od średniego progu słyszenia (średnia arytmetyczna dla częstotliwości 500, 1000, 2000 i 4000 Hz) wyznaczonego w badaniu audiometrycznym oraz w oparciu o klasyfikację zaproponowaną przez WHO (2021), dzieląca ubytki słuchu na stopnie:

- lekki (średni próg słuchu: 21–40 dB HL) oznaczony jako M,
- umiarkowany (średni próg słuchu 41–70 dB HL) oznaczony jako MO,
- znaczny (średni próg słuchu: 71–90 dB HL) – oznaczony jako S

oraz głęboki, w którym próg słuchu przekracza 90 dB HL.Warto zaznaczyć, że wszystkie osoby z grup M, MO oraz S miały niedosłuch odbiorczy bądź mieszany, symetryczny i potwierdzony przebiegiem krzywych audiometrycznych. Szczegółowe informacje dotyczące liczebności poszczególnych grup (N) średniego wieku oraz PTA wraz z odchyleniem standardowym (w nawiasie) przedstawiono w Tab. 6, w której M, MO i S to kolejno grupy słuchaczy z lekkim, umiarkowanym i znacznym ubytkiem słuchu, yNH to młodsi (18–40 r.ż.), a oNH starsi (<40 r.ż.) prawidłowo słyszący.

Tab. 6: *Grupy badanych (liczebność, wiek, próg słyszenia) z odch. standardowym w nawiasie; yNH – młodzi prawidłowo słyszący, oNH – starsi prawidłowo słyszący, M – badani z lekkim niedosłuchem, MO – z umiarkowanym niedosłuchem, S – ze znacznym niedosłuchem*

grupa	**liczba osób**	**wiek**	**PTA [dB HL]**
M	19	66.3 (8.5)	32.4 (4.5)
MO	27	73.3 (8.9)	50.9 (6.0)
S	11	74.0 (9.5)	66.9 (7.8)
yNH	18	23.7 (2.6)	3.0 (3.9)
oNH	17	59.3 (12.3)	19.3 (3.9)

E4.2 Materiał i metody

Pomiary zrozumiałości mowy określonej przez jednoliczbową wartość progową SRT przeprowadzono w pomieszczeniu spełniającym wytyczne ANSI S3.1-1999 przy wykorzystaniu skalibrowanych słuchawek o klasie audiometrycznej Sennheiser HDA200 z kartą dźwiękową EarBox (Auritec) i oprogramowania Oldenburg Measurement Application (Hörzentrum Oldenburg GmbH, Germany). W badaniu wykorzystano listy 20-zdaniowe, tak, jak przy walidacji testu. Zastosowano format open-set w protokole adaptacyjnym *1-up/1-down* ze zmieniającą się wielkością kroku przy stałym poziomie szumu maskującego 65 dB SPL. Część zasadniczą poprzedzone prezentacją 2 list treningowych (w towarzyszeniu szumu TSN). Obecność pomiarów wstępnych wynika z rezultatów badania omówionego w części E2. Badanie realizowane było w ramach grantu Deutsche Forschungsgemeinschaft („Multilingual model-based rehabilitative audiology"; nr 25439187, kier. dr Anna Warzybok-Oetjen).

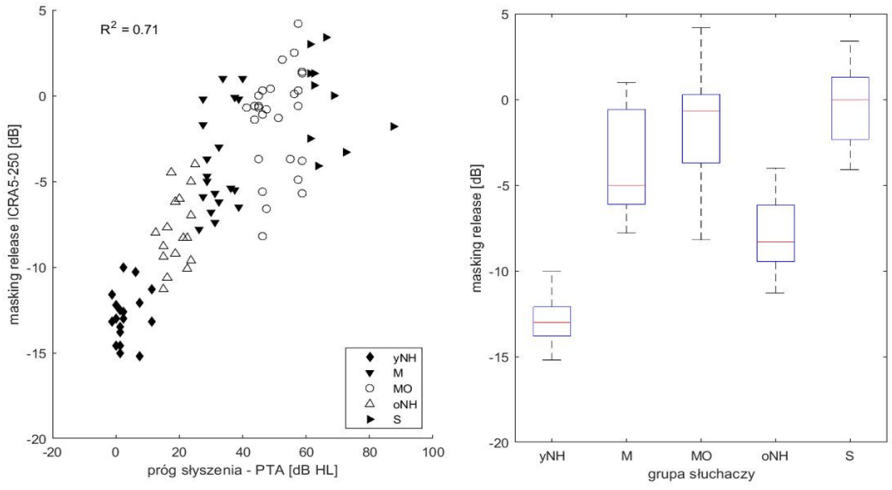
E4.3 Wyniki i ich dyskusja

Wartości średnie SRT w szumie stacjonarnym, fluktuującym oraz ciszy, wraz z odchyleniem standardowym (w nawiasie) przedstawiono w Tab. 7.

Tab. 7: *Średnie wartości SRT w szumie TSN, ICRA5-250 i ciszy w pięciu grupach badanych: yNH – młodzi prawidłowo słyszący, oNH – starsi prawidłowo słyszący, M – z lekkim niedosłuchem, MO – z umiarkowanym niedosłuchem, S – ze znacznym niedosłuchem*

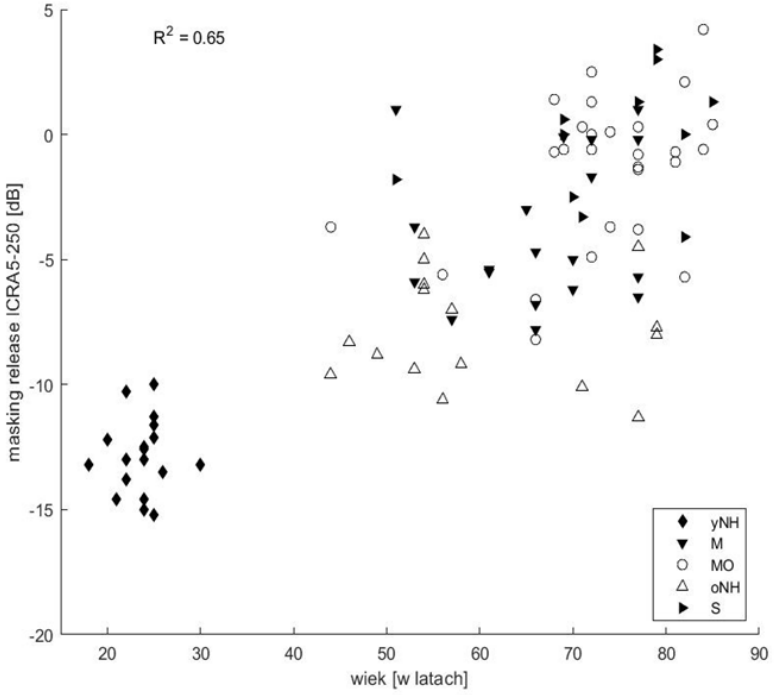
grupa	**TSN**	**ICRA5-250**	**cisza**
M	-4.2 (1.7)	-8.1 (3.6)	32.8 (8.5)
MO	-0.2 (5.4)	-1.6 (5.3)	48.0 (11.5)
S	5.5 (6.2)	5.3 (5.1)	68.9 (9.3)
yNH	-8.5 (1.4)	-21.4 (2.3)	15.8 (5.8)
oNH	-5.7 (1.3)	-13.5 (2.4)	23.6 (2.2)

W szumach stacjonarnym i zmodulowanym, dla wszystkich badanych ze słuchem prawidłowym (yNH, oNH) oraz niewielkim niedosłuchem (M, MO) uzyskano ujemne wartości SRT – poziom sygnału maskującego mógł przewyższać poziom prezentacji testu, a słuchacz mimo to był w stanie prawidłowo powtórzyć 50% zaprezentowanego materiału językowego. Jak wynika z przedstawionych w Tab. 7 danych, szum stacjonarny jest maskerem bardziej efektywnym niż zmodulowany, na co wskazują niższe wartości SRT. Jest to obserwacja zbieżna z wynikami badania walidacyjnego (E1). Różnice pomiędzy średnimi progami zrozumiałości mowy osiągniętymi w tych dwóch scenariuszach pomiarowych świadczą o sile efektu związanego ze zjawiskiem *masking release*. Jego wielkość osiąga średnio 7.8 dB w przypadku starszych prawidłowo słyszących, do nawet 12.9 dB w przypadku młodszych (yNH). Wśród słuchaczy z lekkim i umiarkowanym niedosłuchem (M i MO) są to wartości znacznie niższe – kolejno 3.9 i 1.4 dB. Badani z grupy S, ze średnimi progami słyszenia przekraczającymi 70 dB HL w ogólności nie są wstanie wykorzystać luk w obwiedni sygnału zmodulowanego, co wyraża się poprzez oscylujące wokół zera wartości różnicy między SRT w szumie stacjonarnym i zmodulowanym (średnio 0.2 dB). Uzyskane wyniki dotyczące wielkości *masking release* zestawiono ze średnimi progami słyszenia (PTA). Okazuje się, że wartości tych parametrów dobrze ze sobą korelują ($R^2 = 0.71$) – im większy niedosłuch (wyrażony w tym przypadku poprzez wzrastający średni próg słyszenia, tym mniejsza (bliższa 0) jest percepcyjna korzyść związana z pojawieniem się w maskerze o fluktuującej obwiedni (Rys. 31). W skrajnych przypadkach (wśród słuchaczy z umiarkowanym (MO) i znacznym (S) ubytkiem słuchu) zrozumiałość mowy w szumie stacjonarnym jest większa niż w szumie ICRA5-250.



Rys. 31: *Wielkość uwolnienia od maskowania (masking release) w funkcji średnich progów słyszenia (PTA) w pięciu grupach badanych: yNH – młodzi prawidłowo słyszący, oNH – starsi prawidłowo słyszący, M – z lekkim niedosłuchem, MO – z umiarkowanym niedosłuchem, S – ze znacznym niedosłuchem*

Analizując bardziej szczegółowo rezultaty uzyskane dla osób prawidłowo słyszących zaobserwować można, że o ile w grupie młodszych badanych (yNH) różnice progów zrozumiałości mowy w szumie stacjonarnym i zmodulowanym ICRA5-250 w większości osiągają kilkanaście decybeli (między 10.0 a 15.2 dB), o tyle wśród słuchaczy starszych (oNH) są one średnio 5 dB mniejsze (od 4.0 do 11.3 dB). W pełni uzasadnionym postępowaniem była zatem próba określenia związku pomiędzy wiekiem osoby badanej a korzyścią, jaką czerpie z masking release (Rys. 32).



Rys. 32: *Zależność wielkości masking release od wieku w pięciu grupach badanych: yNH – młodzi prawidłowo słyszący, oNH – starsi prawidłowo słyszący, M – z lekkim niedosłuchem, MO – z umiarkowanym niedosłuchem, S – ze znacznym niedosłuchem*

Współczynnik korelacji wyznaczony dla wszystkich badanych, niezależnie od grupy do której należą, przyjmuje wartość 0.65. Brak wyraźnego związku między wiekiem i różnicą zrozumiałości mowy w różnych warunkach maskowania charakteryzuje natomiast wyniki uzyskane przez słuchaczy z ubytkiem słuchu ($R^2 = 0.17$). Sugeruje to, że bardziej zasadna w tym przypadku jest nie tyle wartość współczynnika korelacji, ile obserwacja rozproszenia wyników poszczególnych słuchaczy. Doskonałym przykładem może być grupa badanych z umiarkowanym ubytkiem słuchu - tylko w relatywnie wąskim przedziale wiekowym (65–70 r.ż.) wartości *masking release* różnią się między sobą nawet o ok. 10 dB. Ponadto, analizując wykres zależności w płaszczyźnie horyzontalnej (oś x), również można zauważyć, że zbliżony zysk w zrozumiałości mowy osiągają słuchacze z podobnym ubytkiem słuchu, ale różniący się wiekiem o niemal dwie dekady. Podobne tendencje

zaobserwowano w grupie starszych prawidłowo słyszących. Była to grupa bardzo jednorodna jeśli chodzi o średni próg słuchu, jednak, w przeciwieństwie do grupy młodszej, zróżnicowana pod względem wieku. (pomiędzy 44 a 77 r.ż.; mediana 54 lata). Zaskakujące jest, jak bardzo skrajne wielkości może przybierać efekt *masking release*. Łatwo zaobserwować to nawet na wykresie. W badanej grupie są osoby, które pomimo braku deficytów słuchowych, mniej efektywnie wykorzystały zjawisko *listening in the gaps* niż ich rówieśnicy z lekkim, a nawet średnim stopniem niedosłuchu.

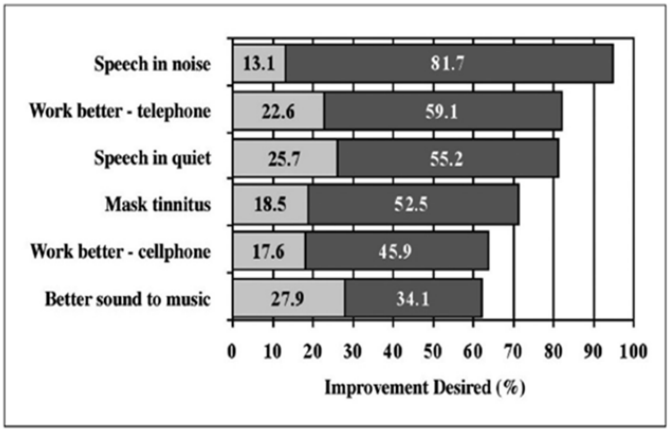
E4.4 Wnioski z eksperymentu E4

W związku z przedstawionymi wynikami pomiarów można stwierdzić, że w badanej grupie zjawisko *masking release* jest mniej skuteczne w przypadku osób z ubytkiem słuchu, co pokrywa się z danymi literaturowymi (Festen i Plomp, 1990, Wagener i Brand, 2005, Francart i in., 2011), jak i wynikami badania walidacyjnego (E1). Na szczególną uwagę zasługuje fakt, iż dla niektórych słuchaczy z niedosłuchem zrozumiałość mowy w szumie zmodulowanym ICRA5-250 jest gorsza niż w szumie stacjonarnym – oprócz procesów związanych z przetwarzaniem informacji na poziomie poznawczym i/lub rozdzielczości czasowej i częstotliwościowej, także sam próg słuchu odgrywa ważną rolę w percepcji mowy w różnych warunkach maskowania. Istotna jest także obserwacja, która nie była brana pod uwagę przy postawieniu hipotezy badania, a dotycząca osób ze słuchem prawidłowym, wskazująca na (istotnie statystyczny) wpływ wieku – im jest on wyższy, tym, w ogólności, korzyść z wykorzystania luk w obwiedni maskera jest mniejsza. To mocny argument wskazujący na olbrzymią wartość dodaną, jaką wnosi audiometria mowy w ocenę wydolności percepcyjnej, względem określania wyłącznie progów słyszenia. Składają się nań, co wielokrotnie sygnalizowano, szereg procesów, zarówno w zakresie prostego odbioru sygnału (wynik audiometrii tonalnej), jak i jego dalszego przetwarzania i interpretacji.

E5 Wykorzystanie polskiego testu zdaniowego matrix do oceny zysku z protezowania słuchu

Wyniki wcześniej omówionych badań (zwłaszcza E1 i E4) jednoznacznie potwierdzają, że ubytek słuchu, w zależności od jego stopnia (który jest czynnikiem istotnym statystycznie) wyraźnie wpływa na stopień zrozumiałości mowy, zwłaszcza w obecności sygnałów maskujących. Obserwuje się także dodatkowy wpływ indywidualnych zmiennych, które przyczyniać się mogą do jeszcze większych trudności w prawidłowej percepcji sygnału mowy (świadectwem ich obecności jest na przykład uzyskanie przez słuchaczy z tymi samimi progami słyszenia wartości SRT różnicących się o nawet kilkanaście decybeli – E1).

Jedną z metod pozwalających na choćby częściową korekcję zaburzeń układu słyszenia i, w efekcie, poprawę zrozumienia sygnału mowy jest dopasowanie aparatu lub aparatów słuchowych (Chisolm i in. 2007). Jego zasadniczym celem jest skompensowanie skutków niedosłuchu w taki sposób aby najwierniej i najefektywniej przywrócić prawidłowe słyszenie, zarówno w domenie częstotliwości (poprzez wykorzystanie np. transpozycji częstotliwościowej), poziomu sygnału (wzmocnienie), jak i zachowania najkorzystniejszego możliwego stosunku sygnału do szumu (zastosowanie filtrów i aktywnej redukcji hałasu). Wszystko po to, aby pozwolić pacjentom na efektywne rozpoznawanie i rozumienie mowy, zwłaszcza w warunkach akustycznie wymagających. Jak wskazują liczne badania, poprawa zrozumiałości mowy w szumie jest najważniejszą potrzebą zgłoszoną przez przyszłych i obecnych użytkowników aparatów słuchowych (Rys. 33).



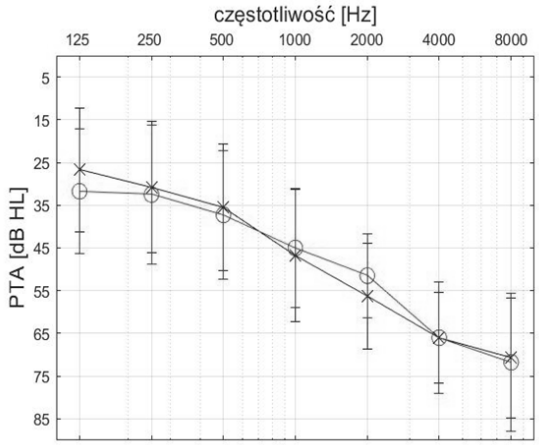
Rys. 33: *Procentowy udział potrzeb (usprawnień) zgłaszanych przez użytkowników aparatów słuchowych, kolejno: poprawa zrozumiałości mowy w szumie, lepsze powiązanie z telefonem, poprawa zrozumiałości mowy w ciszy, maskowanie szumów usznych, lepsze powiązanie z telefonem komórkowym, wyższa jakość dźwięków muzycznych* (źr.: Kochkin, 2002)

Do podobnych wniosków doszli w swoich badaniach Manchaiah i inni (2021) wskazując na to, że najważniejszą potrzebą użytkowników aparatów słuchowych (88% biorących udział w eksperymencie oceniła, że jest ona „bardzo ważna" lub „ekstremalnie ważna") jest „[lepsze] słyszenie przyjaciół i rodziny w hałasie".
Mimo że techniki analizy i przetwarzania sygnałów, a także algorytmy odpowiedzialne za wzmocnienie sygnału pożądanego (mowy) lub usuwanie (ograniczanie natężenia) zakłóceń są coraz bardziej dokładne i zindywidualizowane, wielu pacjentów nadal nie korzysta z aparatów słuchowych, choć je posiada. Szeroko zakrojone badania prowadzone na populacji amerykańskiej wskazują, że problem ten dotyczył w 2022 roku ponad 60% pacjentów, co i tak jest wynikiem bardziej korzystnym niż statystyki z końca lat 90., kiedy aparaty słuchowe były wykorzystywane przez zaledwie 23% (Carr i in., 2022). Jedną z przyczyn takiego zjawiska jest to, że uzyskane korzyści nie są wystarczające do zapewnienia komfortowych warunków komunikacyjnych (Bennett i wsp. 2018; McCormack i Fortnum 2013) – subiektywne oczekiwania pacjentów nie zostają zatem zaspokojone. Sytuacji nie poprawia to, że samo dopasowanie aparatów bądź urządzeń wspomagających słyszenie również jest procesem złożonym, na który wpływ mają czynniki subiektywne, takie jak na przykład doświadczenie audiologa/protetyka słuchu. W wielu krajach sposób oceny korzyści z zastosowania aparatów słuchowych jest niejednoznacznie zdefiniowany lub w ogóle nie jest ustandaryzowany.
Biorąc pod uwagę powyższe kwestie, przeprowadzono badania, których celem było zwalidowanie, po raz pierwszy dla języka polskiego, międzynarodowego podejścia pomiarowego, polegającego na ocenie zrozumiałości mowy nie tylko z wykorzystaniem własnych aparatów słuchowych badanych, ale również dobrze kontrolowanego wirtualnego aparatu słuchowego zaimplementowanego w, opisanym bardziej szczegółowo w dalszej części pracy, programie MHA (ang. *Master Hearing Aid*), który, w założeniu, może wskazać bazę, minimum, jakie powinno być osiągnięte przy indywidualnym dopasowaniu aparatu (lub aparatów) słuchowych.

E5.1 Grupa badana

W zrealizowanych pomiarach wzięło udział 20 dorosłych słuchaczy (śr. wiek: 52.4 lata) z odbiorczym ubytkiem słuchu o wartości co najmniej 40 dB HL, jednak nie przekraczającym 80 dB HL w uchu „gorzej" słyszącym. Ponieważ we wszystkich przypadkach ubytek słuchu był symetryczny (asymetria nie większa niż 15 dB w maksymalnie dwóch częstotliwościach; Cueva, 2004), obliczono średnie wartości progów dla każdego z uszu, które następnie uśredniono. Wszyscy badani mieli

co najmniej 3–letnie doświadczenie w używaniu aparatów słuchowych. Badanie zostało przeprowadzone zgodnie z wytycznymi Deklaracji Helsińskiej i zatwierdzone przez Komisję Etyki Badań Uniwersytetu w Oldenburgu (Nr 71/2015). Przebieg uśrednionych dla obu uszu krzywych progowych (w zakresie 125–8000 Hz) wszystkich badanych, wraz z odchyleniem standardowym, przedstawiono na Rys. 34 (ucho lewe – krzyżyki, ucho prawe – kółka).



Rys. 34: *Średnie progi słyszenia (z odchyleniem standardowym)*

E5.2 Materiał i metody

Do pomiaru zrozumiałości mowy w różnych warunkach maskowania wykorzystano oprogramowanie Oldenburg Measurement Application (HörTech GmbH, Oldenburg) wraz z kartą dźwiękową o dużej mocy (EarBox, Auritec), monitorem studyjnym Genelec 8010 AP oraz słuchawkami Sennheiser HDA200. Zestaw pomiarowy został skalibrowany do dB SPL przy użyciu narzędzi Brüel i Kjær, tj. sztucznego ucha typu 4153, mikrofonu 4134, przedwzmacniacza 2669 oraz wzmacniacza 2610. Wszystkie badania przeprowadzono w akustycznie przygotowanym pomieszczeniu zgodnym z normą ANSI S3.1–1999.

Większość, jeśli nie wszystkie aparaty słuchowe są opracowane na zastrzeżonych systemach, co oznacza, że ich algorytmy nie są w pełni dostępne do celów badawczych. Aby zapewnić bardziej kompleksową analizę tego, jak konkretne ustawienia, metody przetwarzania sygnału czy wybrane procedury dopasowania aparatu wpływają na poprawę zrozumiałości mowy, zmiany w wysiłku słuchowym czy ogólną satysfakcję pacjenta, coraz powszechniej stosuje się ogólnodostępne narzędzia badawcze, które nie działają zgodnie z zasadą „czarnej skrzynki", umożliwiając przeniesienie wyników badań laboratoryjnych (naukowych) do praktycznego projektowania zarówno sprzętu, jak i oprogramowania do

aparatów słuchowych. Jednym z takich rozwiązań jest wykorzystana w tym badaniu *open Master Hearing Aid* (openMHA) – otwarta platforma do badań algorytmów aparatów słuchowych. Chociaż w opisywanych w tej pracy badaniach zaimplementowano jedynie proste symulacje, możliwości MHA są znacznie szersze. W części eksperymentalnej przedstawionego w tym rozdziale pomiaru wykorzystano MHA w konfiguracji NAL-NL2 (najpopularniejszy algorytm dopasowania opracowany przez National Acoustic Laboratories (NAL) do dopasowania HA o szerokim zakresie dynamicznym kompresji; Keidser i in., 2011). Określenie progów SRT oparto o prostą procedurę adaptacyjną *1-up/1-down* ze zmienną wielkością kroku w ramach protokołu *word-scoring* zgodnie z wytycznymi przedstawionymi przez Branda i Kollmeiera (2002) oraz zbieżne z metodologią wcześniej omówionych eksperymentów (E1–E4) przeprowadzonych przez autorkę rozprawy. Początkowy SNR ustalono na poziomie 0 dB, poziom szumu 65 dB SPL, a poziom mowy zmieniał się w miarę dochodzenia do osiągnięcia przez badanego 50% zrozumiałości materiału testowego (PSMT z listami 20-zdaniowymi). Zadaniem słuchacza było powtarzanie słów, które zrozumiał. W celu uniknięcia wpływu efektu treningu typowego dla testów typu matrix, a opisanego w podrozdziale 6.9. oraz doświadczalnie ocenionego w eksperymencie E2, przed częścią główną eksperymentu prowadzono dwa adaptacyjne pomiary SRT w szumie TSN (obydwa z założonymi aparatami słuchowymi, aby upewnić się, że osoby badane właściwie rozumieją protokół badania). Pierwsza lista była prezentowana przy stałym SNR wynoszącym 0 dB (dla słuchaczy z PTA poniżej lub równym 40 dB HL) lub 10 dB (dla słuchaczy z PTA >40 dB HL). Druga lista treningowa była prezentowana w trybie adaptacyjnym z zastosowaniem procedury opisanej powyżej.

Do wyznaczenia SRT w różnych warunkach akustycznych i, tym samym, dostarczenia cennych informacji o indywidualnej zrozumiałości mowy, na którą wpływa wiele, także pozaustrojowych aspektów (Wardenga i in., 2015), wykorzystano trzy różne sygnały maskujące (TSN, ICRA5-250 oraz cafeteria), utrzymywane na domyślnym poziomie 65 dB SPL. Pominięto masker IFFM, aby, z jednej strony ograniczyć czas trwania pomiaru, a z drugiej, z uwagi na to, że, jak wykazano w eksperymencie walidacyjnym E1, jego wykorzystanie nie wnosi dodatkowych informacji w ocenie stopnia zrozumiałości mowy.

Wyżej wymienione maskery zostały już szczegółowo opisane we wstępie do części eksperymentalnej pracy.

Protokół badania obejmował 3 bloki pomiarowe (w kolejności losowej):

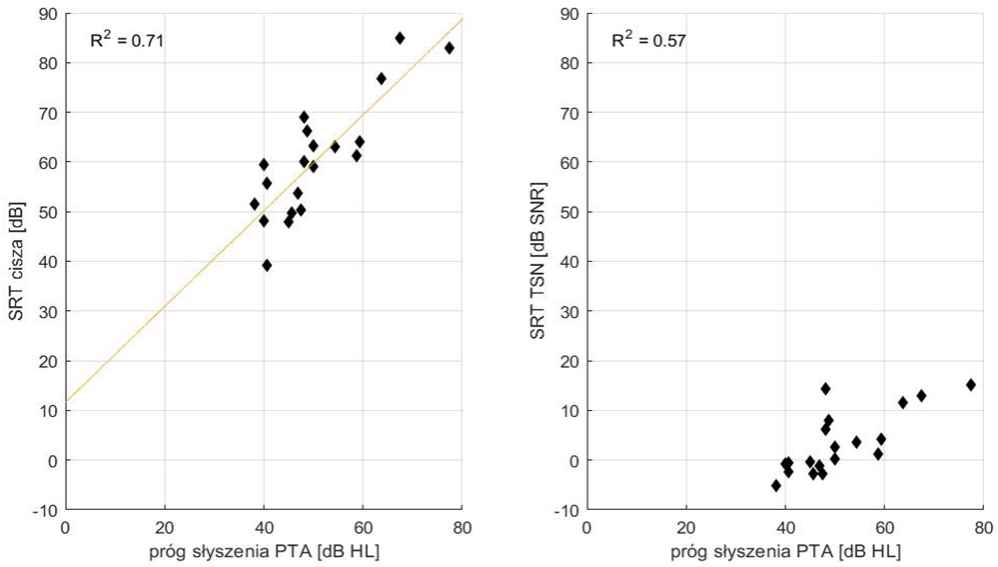
- bez aparatów słuchowych,
- z własnymi aparatami słuchowymi badanego o codziennych, najbardziej komfortowych ustawieniach, które nie ulegały zmianie w trakcie badania,
- Master Hearing Aid z procedurą dopasowania NAL–NL2 (pomiar słuchawkowy),

z krótką przerwą po każdym bloku pomiarowym (lub częściej, jeśli to było konieczne). Wykorzystanie polskiego testu matrix prezentowanego w różnych warunkach akustycznych miało na celu zestawienie potencjalnych rozbieżności pomiędzy zrozumieniem mowy bez aparatów oraz z aparatami słuchowymi. W przypadku kompensacji ubytku słuchu uwzględniono dwie strategie – po pierwsze z wykorzystaniem własnego aparatu słuchowego pacjenta, po drugie – MHA (pomiar słuchawkowy).

E5.3 Wyniki i ich dyskusja

W przypadku badań bez aparatów (bez kompensacji niedosłuchu) zaobserwowano silny związek ($R^2 = 0.71$, na podst. testu korelacji Pearsona) między progami słyszenia i wartościami SRT uzyskanymi w pomiarach bez aparatu słuchowego w ciszy, zarówno dla starszych, jak i młodych badanych. Korelacja między SRT w szumie stacjonarnym TSN i progami słyszenia pozostaje jednak słaba ($R^2 = 0.57$). Progi słyszenia określone przez wartości PTA nie są zatem wystarczające do dokładnego przewidywania stopnia zrozumiałości mowy w wymagającym akustycznie środowisku. Szczegółowe zależności przedstawiono na Rys. 35.

Są to oczywiste obserwacje zbieżne z poczynionymi w innych opisanych eksperymentach, przeprowadzonych przez autorkę niniejszej rozprawy. Podobne jest w przypadku korelacji PTA z SRT w szumach ICRA5-250 i cafeteria – niskie wartości współczynnika, który ją opisuje sugerują, że również czynniki pozasłuchowe (lub przynajmniej nieobwodowe) wpływają na zrozumiałość mowy w utrudnionych warunkach akustycznych. W pomiarach z aparatem (z kompensacją niedosłuchu), zgodnie z oczekiwaniami, nie stwierdzono liniowej korelacji między PTA a wspomaganym (własnym aparatem i MHA) pomiarem w ciszy (R^2 odpowiednio 0.17 i 0.25 dla wszystkich słuchaczy).



Rys. 35: *SRT w ciszy (panel lewy) i szumie stacjonarnym (prawy) w funkcji średnich progów słyszenia PTA*

Najniższe wartości mediany SRT uzyskano dla konfiguracji MHA we wszystkich testowanych scenariuszach (patrz: Tab. 8).

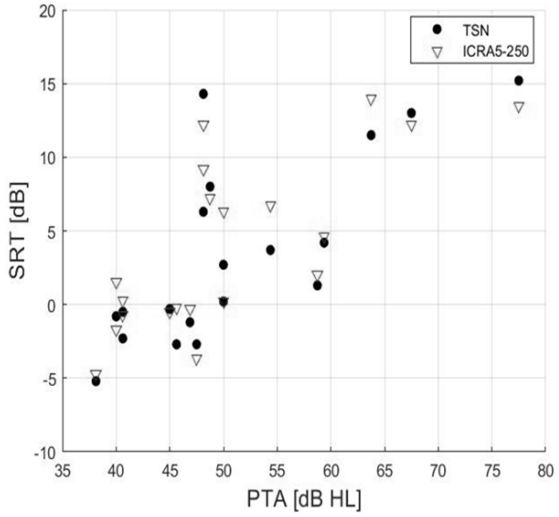
Tab. 8: *Średnie wartości SRT bez kompensacji, z własnym aparatem słuchowym (HA) oraz MHA*

		bez kompensacji	**własny HA**	**MHA**
cisza	zakres [dB]	39.3 do 85.0	35.0 do 68.0	25.3 do 65.7
	średnia [dB]	60.3	46.6	41.4
TSN	zakres [dB]	-5.2 do 15.2	-4.3 do 13.2	-8.2 do 5.3
	średnia [dB]	3.2	2.4	-3.0
ICRA5-250	zakres [dB]	-4.7 do 14.0	-5.5 do 12.5	-15.9 do 12.7
	średnia [dB]	3.9	0.8	-5.5
cafeteria	zakres [dB]	-5.8 do 14.5	brak pomiaru	-7.9 do 34.3
	średnia	0.7	brak pomiaru	-1.0

Uzyskane w badaniu wartości są wyższe niż oczekiwano, a obserwowana korzyść jest mniej znacząca niż w przypadku pomiarów w ciszy. W obecności szumów maskujących można założyć, że „słyszalność" jest częściowo przywrócona, ale indywidualne deficyty nadprogowe, których nie można skompensować zarówno przez aparaty słuchowe, jak i MHA, wpływają niekorzystnie na rozumienie mowy, co objawia się podwyższonymi wartościami SRT. Warto wspomnieć, że najlepszą zrozumiałość mowy zaobserwowano dla zmodulowanego szumu ICRA5-250. Można to wytłumaczyć, biorąc pod uwagę zjawisko „słyszeniania w lukach"(Wagener i in., 2006; Jensen i in., 2019) – poprawę zrozumiałości mowy dzięki modulacji czasowej i/lub widmowej maskera. Zjawisko zostało opisane w podrozdziale 4.1. oraz zmierzone w badaniu opisanym w części E4. Wraz ze wzrostem wartości PTA

próg słyszenia ogranicza wykorzystanie informacji mowy w „lukach czasowych" maskera, ponieważ informacje przypadające poniżej progu słyszenia nie mogą być wykorzystane (są niesłyszalne), ale w tym przypadku deficyty słuchu są częściowo przywracane.

Jak wskazano powyżej, silna korelacja między progami PTA i SRT mierzonych bez aparatu słuchowego w ciszy może być obserwowana zarówno dla starszych, jak i młodych słuchaczy. Sugeruje to, że zrozumiałość mowy w ciszy zależy głównie od progu słyszenia i może być dość dokładnie przewidywana na podstawie jego wartości. To rezultat zbieżny z wynikami badań walidacyjnych E1. Nie jest to jednak tak oczywiste w obecności szumów maskujących. Próg słyszenia nie wystarcza do dokładnego przewidywania rozpoznawania mowy w szumie. W przypadku stacjonarnego TSN, korelacja jest mniejsza w przypadku starszych słuchaczy ($R^2 = 0.39$), co sugeruje silniejszy wpływ innych, ponadprogowych deficytów, zwłaszcza że dla młodszych słuchaczy R^2 wynosi 0.59 (gdy nie jest brany pod uwagę słuchacz z najwyższym PTA – 77 dB HL).



Rys. 36: *Związek PTA z SRT w szumach TSN i ICRA5-250*

Nie stwierdzono liniowej korelacji między PTA a wspomaganym (własny aparat słuchowy i MHA) stopniem zrozumiałości mowy w ciszy. W takich przypadkach ubytek słuchu jest w dużej mierze kompensowany, a zatem słuchacze powinni uzyskać SRT zbliżone do wartości uzyskiwanych przez prawidłowo słyszących. Brak korelacji jest tutaj jak najbardziej oczekiwany. To samo dotyczy pomiarów w szumie stacjonarnym. Nie przeprowadzono pomiarów w szumie cafeteria we własnych aparatach słuchowych (ponieważ nie było to technicznie realizowalne w zastosowanym protokole badawczym), więc nie ma możliwości porównania wartości progowych z własnym aparatem pacjenta w rzeczywistym środowisku.

Jednak, jak udowodniono powyżej, TSN i masker cafeteria dostarczają podobnych informacji dotyczących zrozumiałości mowy.
W części badania z kompensacją ubytku słuchu przeprowadzonej w ciszy, „słyszalność" jest przywracana (dla wszystkich słuchaczy) zarówno w ich własnym aparacie/aparatach, jak i MHA, dlatego nie oczekiwano zależności SRT od ubytku słuchu (określonego wielkością średniego progu PTA4). W MHA zastosowano precyzyjne strojenie, korelacja z PTA jest zatem nieco wyższa, ponieważ dopasowanie odbywa się wyłącznie na podstawie wartości progów słyszenia. W przypadku własnego aparatu dopasowanie jest w pewien sposób zindywidualizowane. Skutkuje to wyższymi korzyściami z MHA w ciszy, co wskazuje na większe tłumienie w konfiguracji MHA niż w przypadku własnego aparatu słuchowego. Mierząc zrozumiałość mowy w szumach przyjęto założenie, że „słyszalność" jest skompensowana, jednak na znaczeniu zyskują jednostkowe deficyty nadprogowe.
Aby odpowiedzieć na pytanie, jak bardzo poprawia się zrozumiałość mowy (w porównaniu do badania bez kompensacji niedosłuchu), w Tab. 9 przedstawiono zakres zysku dla różnych warunków maskowania (TSN, ICRA5-250, cafeteria) i warunków kompensacji (HA, MHA).

Tab. 9: *Zysk w zrozumiałości mowy przy kompensacji (M)HA*

		własny HA	**MHA**
TSN	zysk	-6.7 do 7.5	1.3 do 14.3
	średnia	0.8	6.2
ICRA5-250	zysk	-3.3 do 11.5	1.3 do 14.3
	średnia	3.1	9.4
cafeteria	zysk	nie zmierzono	-1.3 do 11.6
	średnia	nie zmierzono	1.7

W warunkach maskowania największa przewagę wzmocnienia można zaobserwować w zmodulowanym sumie ICRA (średnie 3.1 i 9.4 dB odpowiednio dla HA i MHA). Wartości *masking release* (różnica w SRT między TSN i ICRA) osiągają 2.1 i 2.5 dB (HA i MHA), podczas gdy bez kompensacji wielkość efektu jest nieznaczna lub nieuzyskania w ogóle. Dotyczy to zarówno starszych, jak i młodszych badanych.
Wprawdzie przystępując do badań oczekiwano, że część informacji w lukach czasowych maskera stanie się dostępna tylko ze względu na „odzyskany"niższy próg słyszenia gwarantowany przez własne aparaty słuchowe lub MHA, ale należy mieć na uwadze również inne, pozasłuchowe aspekty słyszenia. Fakt zmniejszonego efektu *release from masking* (lub nawet jego brak) został już wcześniej opisany w literaturze (Festena i Plompa, 1990, a także Krueger, Schulte, Brand i Holube

2017; Zokoll i in. 2017), a powiązano go z kombinacją zmniejszonej rozdzielczości spektralnej i czasowej, a także zwiększoną interferencją modulacji (Jin i in., 2013). Zarówno w przypadku własnych aparatów słuchowych, jak i MHA, korzyści z kompensacji w szumie stacjonarnym korelują z tymi osiągniętymi w szumie ICRA5-250 ($R^2 = 0.61$). Zysk w warunkach cafeteria jest znacznie mniejszy (średnia 1.7 dB) niż w scenariuszach testowych uwzględniających mniej złożone maskery i nie korelują z nimi ($R^2 <0.02$).

Korzyść z zastosowania kompensacji niedosłuchu przez MHA jest obserwowana dla wszystkich badanych i we wszystkich warunkach pomiarowych poza szumem cafeteria. Może to sugerować, że w szumach, w których dominuje mechanizm maskowania energetycznego zniwelowanie tłumienia wynikającego wyłącznie z podwyższenia progów słyszenia pozwala, a które zapewnia algorytm wbudowany w MHA, osiągnąć poprawę rozumienia mowy. Oczywiście, nie jest ona tak doskonała jak w przypadku indywidualnego dopasowania, które bierze (a przynajmniej powinno brać) pod uwagę czynniki inne niż wyłącznie wartości progowe uzyskane w audiometrii tonalnej. Szum cafeteria jest sygnałem najbardziej złożonym. Poza typowym zakłóceniem ambientowym wprowadza także znaczne maskowanie informacyjne, jako że obecny jest w nim sygnał mowy potencjalnie konkurujący z materiałem testowym. W takim wypadku przywrócenie „słyszalności" nie jest zabiegiem wystarczającym.

E5.4 Wnioski

Istotnym pytaniem, które należy postawić analizując wyniki przeprowadzonego badania jest to, czy poprawa zrozumiałości mowy może być przewidywana na podstawie wyników w innych warunkach. Czy szum cafeteria pozwala uzyskać wyniki porównywalne z dobrze kontrolowanymi warunkami laboratoryjnymi (TSN lub ICRA5-250)? Odpowiedzi na te pytania można uzyskać bliżej przyglądając się otrzymanym w eksperymencie rezultatom. Średni zysk wynikający z zastosowania kompensacji MHA w szumie cafeteria wynosi 1.7 dB, a szumie stacjonarnym i zmodulowanym ICRA5-250 kolejno 6.2 i 9.4 dB. Pomiędzy nimi nie ma też korelacji. Stąd wniosek, że warunki laboratoryjne nie pozwalają na ocenę korzyści z zastosowaniem MHA w bardziej rzeczywistym środowisku. Mimo to uzyskane wyniki to interesujące dane, wskazujące na to, że konieczne jest opracowanie scenariuszy pomiarowych i metod testowania zrozumiałości mowy w zróżnicowanych warunkach akustycznych, odpowiadającym codziennym sytuacjom komunikacyjnym.

E6 Wykorzystanie kwestionariusza *Speech, Spatial and Qualities of Hearing* (SSQ) do subiektywnej oceny łatwości komunikacji i lokalizacji dźwięku

Jak wielokrotnie wspomniano w niniejszej rozprawie, zarówno w części teoretycznej, jak i eksperymentalnej, w procesie percepcji mowy oraz towarzyszącemu jej wysiłkowi poznawczemu, poza wydolnością aparatu odbiorczego, istotny jest szereg czynników o charakterze pozasłuchowym. Należą do nich również subiektywne oceny, które można wyrazić korzystając w wielu narzędzi (opisanych w m.in. rozdziale 6.2.). Jednym z nich jest kwestionariusz *Speech, Spatial and Qualities of Hearing* (SSQ) pozwalający ocenić, jak trudne było zrozumienie mowy (lub lokalizacja jej źródła) w różnych sytuacjach z życia codziennego (Gatehouse i Noble, 2004). Jak sama nazwa wskazuje, ewaluacja odbywa się w 3 głównych domenach:

- zrozumieniu mowy (w warunkach maskowania, w tym *speech-on-speech*),
- lokalizacji źródła dźwięku (również ruchomego),
- jakości słyszenia (segregacja źródeł).

Właśnie te aspekty zostały poddane ocenie jako uzupełnienie badania zrozumiałości mowy osób z symetrycznym odbiorczym ubytkiem słuchu.

E6.1 Grupa badana

Jak wspomniano, celem badania było wykorzystanie części formularza SSQ do subiektywnej oceny jakości percepcji słuchowej, ze szczególnym uwzględnieniem zrozumiałości mowy wśród osób z odbiorczym ubytkiem słuchu o wartości co najmniej 40 dB HL, jednak nie przekraczającym 80 dB HL w uchu „gorzej" słyszącym. Byli to Ci sami badani, którzy wzięli udział w pomiarach zrozumiałości mowy z wykorzystaniem Master Hearing Aid (MHA). Uśrednione przebiegi audiogramów przedstawiono na Rys. 34 w części E5.

E6.2 Materiał i metody

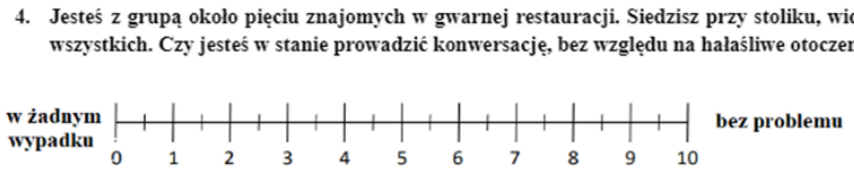
Przygotowany materiał zawierał 13 przetłumaczonych na język polski pytań z różnych kategorii – słyszenie mowy (*speech*), słyszenie w przestrzeni (*spatial*) oraz jakość (*quality*). Pytania zostały wybrane w taki sposób, aby każda z kategorii objęta

była zbliżoną ich liczbą oraz przetłumaczone na język polski przez autorkę niniejszej rozprawy. Miały się również one odnosić do potencjalnych codziennych sytuacji komunikacyjnych badanych. Dokładną treść pytań wraz z podziałem na kategorie przedstawiono poniżej, w Tab. 10. Ich numeracja odnosi się do tej, zastosowanej w pełnej wersji SSQ.

Tab. 10: *Wybrane i przetłumaczone na język polski pytania z kwestionariusza SSQ (numeracja odnosi się do tej, zastosowanej w pełnej wersji SSQ)*

słyszenie mowy (*speech*)	1. Rozmawiasz z inną osobą, a w tym samym pomieszczeniu jest włączony telewizor. Czy możesz śledzić, co mówi osoba, z którą rozmawiasz, bez wyłączania telewizora? 2. Rozmawiasz z inną osobą w cichym, wyłożonym dywanem salonie. Czy możesz śledzić to, co mówi druga osoba? 4. Jesteś w grupie około pięciu osób w gwarnej restauracji. Siedzisz przy stoliku, widząc twarze ich wszystkich. Czy jesteś w stanie prowadzić konwersację bez względu na hałaśliwe otoczenie? 10. Słuchasz kogoś, kto mówi do Ciebie, a jednocześnie próbujesz śledzić wiadomości w telewizji. Czy możesz śledzić to, co mówią obie osoby? 11. Rozmawiasz z jedną osobą w pokoju, w którym jest wiele innych osób. Czy możesz śledzić, co mówi osoba, z którą rozmawiasz? 12. Jesteś w grupie i rozmowa przechodzi od jednej osoby do drugiej. Czy możesz z łatwością śledzić rozmowę, nie tracąc z oczu początku wypowiedzi każdego nowego rozmówcy?
słyszenie w przestrzeni (*spatial*)	6. Jesteś na zewnątrz. Pies głośno szczeka. Czy możesz od razu powiedzieć, gdzie on jest, bez konieczności patrzenia w jego kierunku? 9. Czy potrafisz określić na podstawie dźwięku, jak daleko znajduje się autobus lub ciężarówka? 13. Czy potrafisz stwierdzić na podstawie dźwięku, czy autobus lub ciężarówka zbliża się do Ciebie, czy odjeżdża?
jakość (*quality*)	7. Kiedy słuchasz muzyki, czy potrafisz rozpoznać, grające instrumenty? 9. Czy codzienne dźwięki, które możesz łatwo usłyszeć, wydają Ci się wyraźne (nie rozmyte)? 11. Czy codzienne dźwięki, które można łatwo usłyszeć, wydają się wyraźne (nie rozmyte)? 14. Czy musisz się bardzo koncentrować, słuchając kogoś lub czegoś?

Uczestnicy byli zobowiązani ocenić na skali (0–10), jak dobrze byliby w stanie zrobić lub doświadczyć tego, co opisane jest w pytaniu. Wybór „10" oznaczał, że badany nie ma żadnego problemu ze słyszeniem/wykonaniem określonej czynności w opisanych warunkach, natomiast „0" to zupełny brak możliwości poradzenia sobie w przedstawionej sytuacji. Fragment formularza odpowiedzi przedstawiono poniżej:



Rys. 37: *Fragment formularza odpowiedzi zmodyfikowanego formularza SSQ przetłumaczonego na język polski*

E6.3 Wyniki i ich dyskusja

Nie stwierdzono znaczących związków średnich wartości progów słyszenia z wynikami kwestionariusza (R^2 <0.1). Dotyczy to zarówno pomiarów w ciszy, jak i obecności szumu maskującego (TSN). W przeciwieństwie do badań Moulin i in. (2016) nie zaobserwowano również powiązania między średnimi wynikami ankiety w kategorii *spatial* i różnic międzyusznych wyznaczonych w audiometrii tonalnej. Na podstawie uzyskanych rezultatów można wnioskować, że w przypadku badanej grupy słuchaczy, subiektywna ocena zdolności komunikacyjnych oraz możliwości utrzymania optymalnej efektywności percepcyjnej w różnych sytuacjach nie jest skorelowana z faktyczną miarą zrozumiałości mowy.

Niewykluczone jest, że na taki rozkład wyników miały wpływ dodatkowe czynniki. Po pierwsze, mimo że grupa była stosunkowo jednorodna pod kątem ubytku słuchu oraz doświadczenia z użytkowaniem aparatów słuchowych, badani byli w różnym wieku (SD jest bardzo wysokie, rzędu dwudziestu kilku lat). Wiek, jak wspomniano w rozdziale 4.5. może być zmienną znacznie wpływającą na zrozumiałość mowy oraz percepcję sygnałów w wymagających środowiskach w ogóle. Po podzieleniu badanej grupy na część młodszych i starszych (>60 r.ż.) słuchaczy wyraźnie widać, że średnie oceny w poszczególnych kategoriach są wyższe w grupie młodszych badanych – Tab. 11.

Tab. 11: *Średnie oceny w kategoriach formularza SSQ*

kategoria	**słuchacze młodsi**	**słuchacze starsi (>60 r.ż.)**
speech	6.5	3.7
spatial	7.2	6.0
quality	6.1	5.6

Nawet jeśli przebiegi audiogramów były podobne, niewiele wiadomo o etiologii ubytków słuchu. Biorąc pod uwagę wymienione wcześniej różnice wieku słuchaczy, najprawdopodobniej jest ona odmienna. To także może wprowadzać dodatkowe aspekty utrudniające jednoznaczne powiązanie miar obiektywnych z subiektywnymi. Drugą kwestią, którą należy wziąć pod uwagę jest wybór pytań. Celem takiej właśnie selekcji było zapewnienie najwyższego możliwego związku z doświadczeniami uczestników eksperymentu, niezależnie od ich wieku, statusu zawodowego itp. Nie można jednak wykluczyć, że wybrano niewłaściwe pytania lub było ich zbyt mało.
Poza szukaniem powiązania między progami słyszenia, a średnimi ocenami w SSQ, podjęto próbę odpowiedzi na pytanie, czy subiektywna ocena jakości/łatwości słyszenia w konkretnych sytuacjach pozostaje w związku ze stopniem zrozumiałości mowy w szumie stacjonarnym oraz ciszy. W tabeli poniżej przedstawiono wartości współczynnika korelacji Pearsona dla związków między SRT (określającym próg 50% zrozumiałości mowy) w szumie stacjonarnym TSN oraz ciszy a subiektywnymi ocenami z kwestionariusza SSQ z podziałem na kategorie.

Tab. 12: *Współczynnik korelacji między średnią oceną SSQ a SRT*

kategoria	**SRT w TSN**	**SRT w ciszy**
speech	0.36	0.19
spatial	0.12	0.05
quality	0.08	0.04
średnia	0.14	0.14

Okazuje się, że dla żadnego z 3 aspektów ocenianych w SSQ nie istnieje silne powiązanie ze zmierzonymi progami zrozumiałości mowy.Wyniki ankiety mogą się wpisywać w opisywaną przez wielu autorów tendencję do niedoszacowania negatywnego wpływu ubytku słuchu na komunikowanie się. Wieloletnie badania MarkeTrack prowadzone w Stanach Zjednoczonych (Carr i Kim, 2022) na dużej populacji osób z niedosłuchem wskazują je jako jedną z przyczyn nadal niewystarczająco wysokiej wartości współczynnika adopcji aparatów słuchowych. Pacjencie zauważają, że czasem mają problem w sytuacjach komunikacyjnych, zwłaszcza angażujących wielu interlokutorów (i ma to odzwierciedlenie w badaniach zrozumiałości mowy), natomiast, zwłaszcza jeśli nie noszą aparatów słuchowych długo, uznają, że „nie jest tak źle" (stąd wysokie oceny w ankiecie SSQ).
Te niejednoznaczności są kolejnym dowodem na to, jak bardzo złożone są procesy poznawcze stojące za zrozumieniem mowy i jak precyzyjne i zindywidualizowane powinny być narzędzia służące do oceny ich wydajności, zarówno w ujęciu obiektywnym, jak subiektywnym. Chociaż pożądane byłoby opracowanie jednego obiektywnego parametru, który mówiłby wszystko o ubytku słuchu i sposobach jego

korekcji, wyniki tego, jak i wielu innych badań (również przedstawionych w niniejszej rozprawie) wskazują, że współczesna psychoakustyka, a dokładniej ta jej część, która bezpośrednio dotyczy audiologii, nie ucieknie przed wielowymiarowym badaniem. Wielowymiarowość ta wiąże się z uwzględnieniem klasycznych wartości pomiarowych (obiektywnych), ale i elementów subiektywnych, niezwykle istotnych dla całościowo objętego procesu percepcji. Wyzwaniem pozostaje nadal określenie, z jakich metod subiektywnej oceny korzystać, aby możliwa stała się ich maksymalna obiektywizacja, rozumiana jako zdolność do zapewnienia jednoznacznych i konsekwentnych w obrębie badanej jednostki rezultatów.

E6.4 Wnioski z eksperymentu E6

W badaniu nie udało się podtrzymać przyjętej hipotezy, zgodnie z którą progi zrozumiałości mowy mogłyby być silnie powiązane ze średnimi ocenami jakości percepcji w zakresie mowy, jakości dźwięku i słyszenia przestrzennego sprecyzowanym przez kwestionariusz SSQ. Choć uzyskane w tym kontekście wyniki są niejednoznaczne, niewątpliwie stanowią dobry punkt wyjścia do dalszych, bardziej złożonych badań dotyczących potencjalnego wykorzystania subiektywnych kwestionariuszy odnoszących się do percepcji mowy w warunkach i sytuacjach akustycznych, w których na co dzień przebywają pacjenci gabinetów protetyki słuchu, w praktyce specjalistów z zakresu dopasowania aparatów słuchowych (i innych urządzeń wspomagających słyszenie) oraz oceny zysku z nich.

E7 Wykorzystanie testu matrix do oceny wysiłku słuchowego (*listening effort*)

Jak podkreślono we wcześniejszych częściach rozprawy, zarówno teoretycznej, jak i eksperymentalnej, w diagnostyce słuchu, jego monitorowaniu, a w razie konieczności – protezowaniu (kompensacji niedosłuchu), powinna być wykorzystywana mowa, która, z punktu widzenia codziennego funkcjonowania i komunikacji, jest sygnałem najważniejszym. Naturalną odpowiedzią na tę potrzebę jest uzupełnienie audiometrii tonalnej czy badań obiektywnych (np. pomiaru otoemisji akustycznych) metodami określającymi stopień zrozumiałości mowy. Nie od dziś wiadomo jednak, że w badaniach dotyczących szeroko pojmowanej percepcji nie chodzi wyłącznie o wyznaczenie jednoliczbowych wskaźników jej „wydolności" (najprostszym przykładem mogą być eksperymenty około-progowe), ale również ocenę jakościową, a więc uwzględnienie czynników, które mogą tę percepcję charakteryzować. Do takich należą np. czas reakcji, ale również jakakolwiek ocena bodźca wykraczająca poza tylko jego detekcję. W akustyce doskonałym przykładem może być określenie stopnia dokuczliwości hałasu przy próbie opisu konkretnego środowiska lub źródła dźwięku. Jednym zagadnieniem jest w tym wypadku to czy (i kiedy) sygnał jest wykrywany (wyróżniany spośród tła) lub prawidłowo identyfikowany (właściwie rozpoznany), a innym, niemniej ważnym, jaką reakcję, również emocjonalną, wywołuje. Wprawdzie w regulacjach prawnych, normach i wytycznych wskazuje się w zasadzie wyłącznie cechy fizyczne bodźca (poziom, częstotliwość) lub obiektywne wskaźniki wydajności poznawczej (próg słyszenia, rozdzielczość wzroku), to jednak chcąc dokonać wieloaspektowej analizy psychofizycznej, konieczne jest uwzględnienie także innych czynników. Poza ogromnym znaczeniem naukowym, pozwalającym na jeszcze bardziej szczegółowe poznanie możliwości ludzkiego organizmu i procesów nim rządzących, tak dogłębna analiza ma często znaczenie praktyczne.

W przypadku problematyki poruszanej w niniejszej pracy uzupełnieniem informacji o wydolności słuchu może być wyznaczenie wysiłku słuchowego, zdefiniowanego w rozdziale 7. Przykładem dobrze obrazującym zasadność jego określania jest chociażby sytuacja rozmowy prowadzonej w pomieszczeniu o znacznym pogłosie, z wieloma mówcami, których głosy dochodzą z różnych kierunków. Z dużym prawdopodobieństwem można założyć, że prawidłowo słysząca osoba będzie w stanie zrozumieć większość, jeśli nie wszystkie wypowiedziane słowa (zdania) – jednoliczbowy wskaźnik stopnia zrozumiałości mowy mógłby zatem osiągnąć wartość maksymalną, 100% Stopień koncentracji i wysiłek poznawczy, które słuchacz musi

włożyć w ten proces są jednak niewspółmiernie większe w porównaniu z rozmową prowadzoną z jednym interlokutorem w cichym pomieszczeniu. Innym praktycznym zastosowaniem może być również ocena zysku z dopasowania aparatów słuchowych. Zdarza się, że pomimo zastosowania teoretycznie najbardziej korzystnych ustawień urządzeń wspomagających słyszenie, nie udaje się uzyskać wyższej procentowej zrozumiałości mowy. Może się jednak okazać, że, dzięki wdrożeniu odpowiedniej kompensacji, zmniejszy się tzw. *cognitive load*, co znacznie zwiększy komfort pacjenta, zwłaszcza w warunkach akustycznie wymagających.
Pomiar wysiłku słuchowego odbywać się może we wcześniej zdefiniowanych warunkach akustycznych np. przy określonych z góry poziomach sygnału pożądanego lub stałych stosunkach sygnału do szumu (m.in. van Schoonhoven i in., 2016). Warto pamiętać o tym, że bodźce, z uwagi na możliwy wpływ kolejności ich prezentacji, czy rozpiętość wartości, które przyjmują, powinny być prezentowane losowo, z zachowaniem maksymalnie równomiernego rozłożenia i, przede wszystkim, w indywidualnym zakresie percepcji (Parducci i Perrett, 1971). Jest on szczególnie istotny do uwzględnienia w badaniach o charakterze subiektywnym.

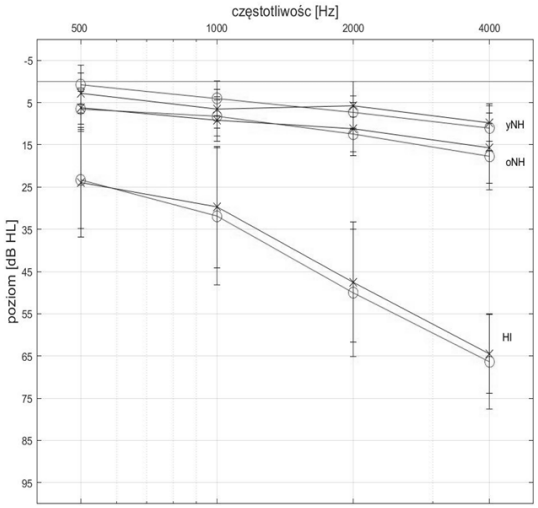
E7.1 Osoby badane

W badaniu wzięło udział 58 pełnoletnich słuchaczy – 20 młodszych prawidłowo słyszących (zgodnie z WHO - średni próg słyszenia nieprzekraczający 20 dB HL), 20 starszych (powyżej 50 r.ż.) prawidłowo słyszących oraz 18 osób z symetrycznym (maksymalna różnica progów słyszenia poniżej 15 dB) niedosłuchem odbiorczym. Wszyscy słuchacze przed badaniem zostali poinformowani o jego celu oraz dobrowolnie podpisali zgodę na udział w nim. Większość badanych nie miała doświadczenia w eksperymentach psychoakustycznych. Z oczywistych względów, żaden z uczestników nie miał wcześniej kontaktu z ACALES. Podstawowe dane dot słuchaczy (wraz z odchyleniem standardowym w nawiasie) zawarto w Tab. 13.

Tab. 13: *Informacje na temat uczestników badania z wykorzystaniem ACALES - yNH i oNH kolejno: młodsi i starsi (≥5 0 r.ż.) prawidłowo słyszący, HI – osoby z symetrycznym niedosłuchem odbiorczym; w nawiasach wartości odchylenia standardowego*

grupa	**N**	**średni wiek**	**średni próg słuchu PTA**
yNH	20	31.8 (6.8)	4.4 (2.6)
oNH	20	58.4 (6.0)	8.9 (3.9)
HI	18	59.9 (15.9)	39.3 (10.3)

Na wykresie przedstawiono średnie wartości progowe z odchyleniem standardowym, dla czterech częstotliwości, wchodzących w skład parametru określającego średni próg słyszenia – 500, 1000, 2000 i 4000 Hz.



Rys. 38: *Uśrednione przebiegi audiogramów uczestników badania z ACALES; kółka – ucho prawe, krzyżyki ucho lewe*

Jak można zauważyć, średnie progi w grupach młodszych i starszych prawidłowo słyszących są do siebie bardzo zbliżone, również dla wyższych badanych częstotliwości (6 i 8 kHz), które nie zostały ujęte na wykresie. W grupie słuchaczy z niedosłuchem dominował ubytek wysokoczęstotliwościowy. W zakresie częstotliwości niskich audiogram przyjmował, w ogólności, kształt prawidłowy lub zbliżony do prawidłowego, by rozpocząć opadanie w zakresie średnich częstotliwości (powyżej 1000 Hz).

E7.2 Materiał i metody

W przypadku omawianego eksperymentu wykorzystano, po raz pierwszy dla języka polskiego, protokół ACALES (Krueger i in., 2017). Do dolnej i górnej granicy zakresu przypisano, kolejno „brak wysiłku" oraz „ekstremalny wysiłek", pomiędzy którymi znajduje się kilka podkategorii oceny. Podejście to opiera się na adaptacyjnej procedurze kategorialnego skalowania głośności ACALOS (Brand i Hohmann, 2002), w której kwantyfikuje się percepcję głośności dla różnych poziomów prezentowanego sygnału od „niesłyszalnego"do „zbyt głośnego". Poziom dźwięku regulowany jest adaptacyjnie, w oparciu o poprzednie oceny słuchacza. W protokole ACALES, odpowiednio, wcześniejsze oceny postrzeganego wysiłku słuchowego, warunkują dalszy przebieg badania.

Na początku procedury określane są granice zakresu parametrów (SNR w badaniach z towarzyszeniem maskera i poziomu prezentacji sygnału mowy w badaniu w ciszy) dla ocen „brak wysiłku" i „ekstremalny wysiłek". W kolejnym korku bodźce są prezentowane przy losowych (ale równomiernie rozłożonych) poziomach SNR leżących we wcześniej zdefiniowanym zakresie, tak aby określić wartości dla ocen pośrednich.
Polską wersję językową skali ACALES, zaproponowaną przez autorkę pracy, a która nie była do tej pory wykorzystywana w żadnym eksperymencie dotyczącym percepcji słuchowej charakteryzowanej w oparciu o materiał językowy w języku polskim, przedstawiono na Rys. 39.

Rys. 39: *Polska wersja skali ACALES (opracowanie własne na podst. Krueger i in., 2017)*

Do każdej z kategorii zawartych w skali jako jednostki numeryczne zostały przypisane kategorialne jednostki wysiłku (*effort scale categorical units*; ESCU). Kategoria „brak wysiłku"odpowiadała 1 ESCU, „bardzo mały wysiłek 3 ESCU, „mały wysiłek"5 ESCU, „umiarkowany wysiłek"7 ESCU, „znaczny wysiłek"9 ESCU, „bardzo duży wysiłek"11 ESCU, „ekstremalny wysiłek"13 ESCU. Liczby w wyrażone w ESCU nie były widoczne dla badanych. Chociaż z punktu widzenia słuchacza, badanie cały czas przebiegało tak samo (oczywiście ze zmiennym stopniem trudności) kolejne kroki można określić jako:

- **I FAZA:** *Określenie indywidualnych granic zakresu SNR dla ocen „brak wysiłku" oraz „ekstremalny wysiłek" (lub „tylko szum").*
  Wartości SNR były adaptacyjne zmieniane na podstawie odpowiedzi (ocen) osób badanych w dwóch, pojawiających się w naprzemiennej kolejności procesach. W pierwszym SNR był zwiększany w krokach 3 dB tak długo, aż badany ocenił prezentowaną serię jako „brak wysiłku". W drugim etapie poziom prezentacji mowy był zmniejszany w krokach 3 dB aż do wybrania przez uczestnika oceny „ekstremalny wysiłek"lub „tylko szum". Pierwsza faza pomiaru trwała dopóki nie zostały określone dolne i górne granice zakresu prezentacji (1 i 13 lub 14 ESCU).

- **II FAZA:** *Oszacowanie SNR dla pięciu kategorii nominalnych zawartych w skali – „bardzo mały wysiłek", „mały wysiłek", „umiarkowany wysiłek","znaczny wysiłek"i „bardzo duży wysiłek"na drodze liniowej interpolacji.*
  Zestawy po 3 zdania zostały zaprezentowane w losowej kolejności przy 5, wyżej zdefiniowanych SNR.

- **III FAZA:** *Dopasowanie prostej do ocen z drugiej fazy za pomocą regresji liniowej oraz ponowne oszacowanie SNR dla kryteriów granicznych („brak wysiłku" i „ekstremalny wysiłek").*

Podobnie jak w przypadku innych języków, dla języka polskiego również występuje wiele odmian testów wykorzystujących materiał słowny, stosowanych m.in. w diagnostyce słuchu. Spośród tych, które wykorzystują najbardziej złożone jednostki językowe wyróżnić można testy składające się ze zdań pochodzących z mowy codziennej, bazujące na testach typu Plompa oraz polską wersję testu typu matrix (Ozimek, 2009). Jak wykazano w literaturze, ale również części eksperymentalnej tej rozprawy (E1), dzięki temu, że polski test zdaniowy matrix cechuje się dużą stromością krzywych psychometrycznych – wynosi ona około 17.1%/dB (podobnie jak w języku niemieckim) i jest znacznie większa niż uzyskana przy testach wykorzystujących listy składające się z pojedynczych wyrazów (Kollmeier i in., 1992; Keidser, 1993), stanowi bardzo czułe i wiarygodne narzędzie wspomagające ocenę czynności narządu słuchu, z uwzględnieniem procesów centralnych, jak i weryfikację ustawień urządzeń wspomagających słyszenie. Z tych powodów w opisywanym eksperymencie wykorzystano właśnie test zdaniowy matrix, zarówno w części dotyczącej wysiłku słuchowego, jak i oceny stopnia zrozumiałości mowy. W przypadku tej drugiej, słuchaczom prezentowano listy składające się z 30 losowo wygenerowanych zdań. Liczba ta jest zgodna z podejściem zaprezentowanym

przez Krueger (2017) oraz pozwala na wyznaczenie z dużą dokładnością nie tylko wartości SRT, ale również nachylenia krzywych psychometrycznych (aby to osiągnąć mierzone określano wartości SNR dla 20% i 80% zrozumiałości mowy). Z uwagi na *efekt treningu*, który jest immanentną cechą testu typu matrix, a który został opisany w rozdziale 6.9 oraz E2, przed częścią dotyczącą zrozumiałości mowy badanym prezentowane były skrócone listy testowe. Jak wcześniej wspomniano, zarówno pomiary zrozumiałości mowy, jak i wysiłku słuchowego odbywały się w obecności sygnałów zakłócających. Ich wybór podyktowany był zróżnicowaniem charakterystyk maskerów oraz faktem zastosowania dokładnie tych samych sygnałów w badaniu dotyczącym klinicznej walidacji polskiej wersji testu matrix (E1) - TSN, IFFM, ICRA5-250 oraz cafeteria. W obu częściach badania masker utrzymywany był na stałym poziomie wynoszącym 65 dB SPL. Ponieważ zdania były prezentowane przy kilku SNR, ogólny poziom sygnału maskującego i użytecznego wynosił 65 dB SPL lub więcej, w zależności od bieżącej wartości SNR, jednak zawsze poniżej wartości mogących wywołać dyskomfort słuchacza. Do pomiarów zrozumiałości mowy, poprzedzonych krótkim wywiadem, otoskopią i audiometrią tonalną, wykorzystano oprogramowanie Oldenburg Measurement Application na PC (Hörzentrum Oldenburg GmbH, Niemcy). W badaniu wykorzystano przetwornik DAC Audinst HUD-mini oraz monitor studyjny Genelec 8010AP umieszczony na statywie. Konfiguracja pomiarowa została skalibrowana do dB SPL w obecności szumu stacjonarnego TSN oraz przy użyciu miernika SVAN 979 w odległości 1.4 m od źródła dźwięku w miejscu zbliżonym do położenia głowy siedzącego słuchacza. Wszystkie badania przeprowadzono w pomieszczeniu realizującym normę ANSI S3.1-1999.

Zgodnie z przedstawionymi powyżej informacjami eksperyment składał się z dwóch etapów. W pierwszym, przy wykorzystaniu metody adaptacyjnej, wyznaczane były progi zrozumiałości mowy SRT w obecności 4 sygnałów maskujących oraz w ciszy. Z uwagi na to, że przebieg pomiaru był bardzo podobny do tego, dotyczącego klinicznej walidacji polskiego testu matrix (E1), nie będzie on w tym miejscu ponownie opisywany. Jedyną istotną różnicą było to, że pomiar odbywał się w polu swobodnym (a nie z wykorzystaniem słuchawek), a prezentowane listy składały się z 30, a nie 20 zdań, choć nadal zachowywały cechy zrównoważenia. W drugiej części, badani proszeni byli o ocenę wysiłku słuchowego w warunkach odsłuchowym identycznych, co w pierwszej części, w oparciu o skalę ACALES. To, która z części eksperymentu była wykonywana jako pierwsza, a także kolejność warunków prezentacji sygnału dobierane było losowo dla każdego słuchacza. Trzy wybrane zdania z polskiego testu matrix (Ozimek, 2009) były prezentowane

badanym przy takim samym SNR przed każdą oceną wysiłku słuchowego. Liczba ta została zaproponowana przez twórców protokołu badania (Krueger i in., 2017) z uzasadnieniem, iż:

- pierwsze zdanie pozwala uzyskać pierwsze wrażenie z sytuacji odsłuchowej,
- drugie zdanie dostarcza wstępnych informacji o postrzeganym wysiłku,
- a dopiero trzecie zdanie umożliwia podjęcie ostatecznej decyzji dotyczącej postrzeganego wysiłku słuchowego.

Zadaniem badanego była odpowiedź na pytanie: „Jak dużo wysiłku sprawia Ci zrozumienie mówcy?", która miała być wyrażona na kategorialnej skali z siedmioma oznaczonymi kategoriami i sześcioma stopniami pośrednimi (jak u Krueger i in., 2017 oraz Luts i in., 2010), uzupełnionej o kategorię „tylko hałas"(nie branej pod uwagę w ostatecznych analizach, jako odpowiedź niebędącą związaną z jakimkolwiek postrzeganym wysiłkiem słuchowym). Poziom SNR zmieniał się adaptacyjnie, z uwzględnieniem wcześniejszych odpowiedzi słuchacza. Przed pomiarem zasadniczym przeprowadzono trening – na tle szumu stacjonarnego TSN zaprezentowano kilka „trójek" zdań wygenerowanych z testu matrix prosząc badanych o ocenę wysiłku słuchowego w oparciu o ACALES. Słuchacze z grupy HI odbywali badanie bez aparatów słuchowych.

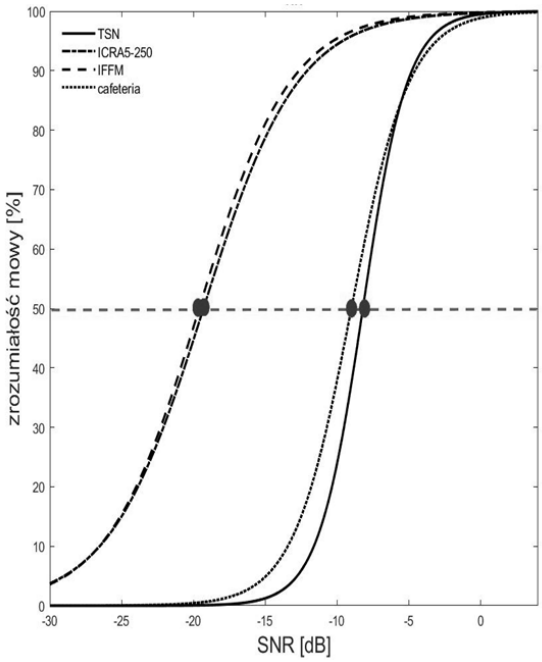
E7.3 Wyniki i ich dyskusja

W Tab. 14 przedstawiono średnie wartości SRT (SNR, przy którym uzyskano 50% zrozumiałość mowy) we wszystkich badanych grupach i warunkach pomiarowych wraz z odchyleniem standardowym. SRT w ciszy wyrażone są w dB SPL.

Tab. 14: *Średnie wartości SRT w trzech badanych grupach: yNH - młodzi prawidłowo słyszący, oNH - starsi prawidłowo słyszący, HI - osoby z symetrycznym odbiorczym niedosłuchem*

		TSN	**ICRA5-250**	**IFFM**	**cafeteria**	**cisza**
yNH	SRT [dB]	-8.3	-19.3	-17.8	-9.0	18.4
	SD	1.8	2.5	8.7	1.4	4.3
oNH	SRT [dB]	-7.2	-16.5	-16	-7.1	20.1
	SD	1.9	3.1	4.0	2.1	4.7
HI	SRT [dB]	-2.7	-7.4	-7.1	-1.9	38.0
	SD	4.7	6.6	7.5	4.9	11.7

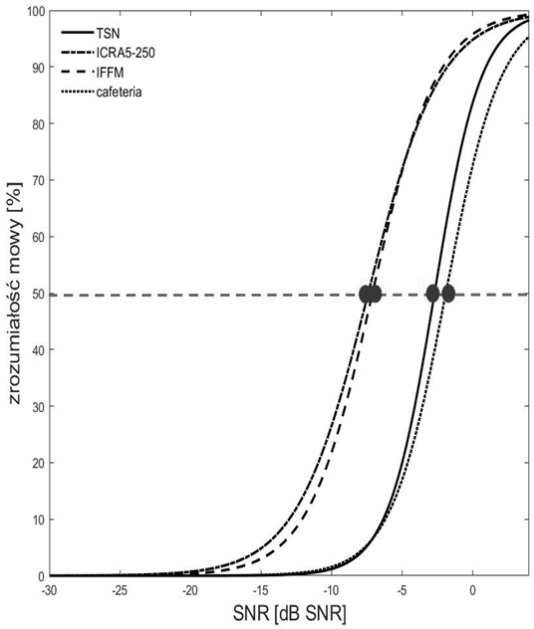
Jak można zauważyć, najbardziej efektywne maskowanie zapewniają szumy stacjonarne – TSN oraz cafeteria – uzyskane SRT są wyższe, niż w przypadku szumów fluktuujących, co może być związane z energetycznym charakterem maskowania, które gwarantują te sygnały. Różnice między średnim SRT w hałasie stacjonarnym i zmiennym są większe w przypadku obu normalnie słyszących grup i sięgają nawet kilkunastu decybeli (np. ponad 11.5 w przypadku progów w TSN i IFFM w grupie yNH).



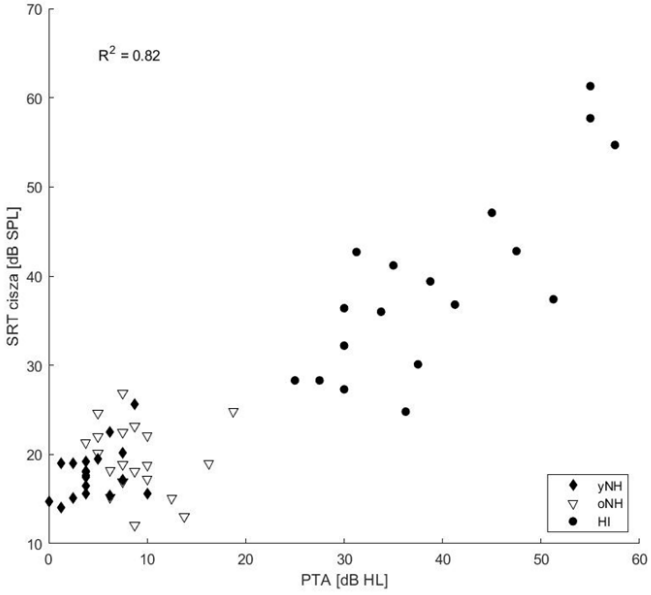
Rys. 40: *Krzywe psychometryczne – grupa młodszych prawidłowo słyszących (yNH)*

Dla porównania, średnie wartości progów zrozumiałości mowy w szumie stacjonarnym i fluktuującym wśród słuchaczy z niedosłuchem różnią się o kilka decybeli, a więc niemal dwukrotnie mniej.

Niskie wartości SRT uzyskane w obecności szumów ICRA5-250 i IFFM mogą być związane ze zjawiskiem *listening in the gaps* (opisanym szczegółowo w części E4). W przypadku osób z ubytkiem słuchu mechanizm ten jest mniej wydajny, ponieważ część informacji zawartych w sygnale mowy, choć prezentowana bez obecności szumu, jest niedostępna, ponieważ znajduje się poniżej indywidualnego progu słyszenia badanego. Niewątpliwie istotny wpływ mają również czynniki pozasłuchowe, opisane m.in. w podrozdziałach 4.9. i 4.10.
Rys. 42. przedstawia liniową korelację między średnimi wartościami PTA i SRT uzyskanymi w ciszy. Jak widać, występuje dość silna zależność ($R^2 = 0.82$), co oznacza, że zrozumiałość mowy w ciszy zależy głównie od progu słyszenia i może być dość dokładnie przewidywana na podstawie jego wartości.



Rys. 41: *Krzywe psychometryczne – grupa badanych z symetrycznym niedosłychem odbiorczym (HI)*

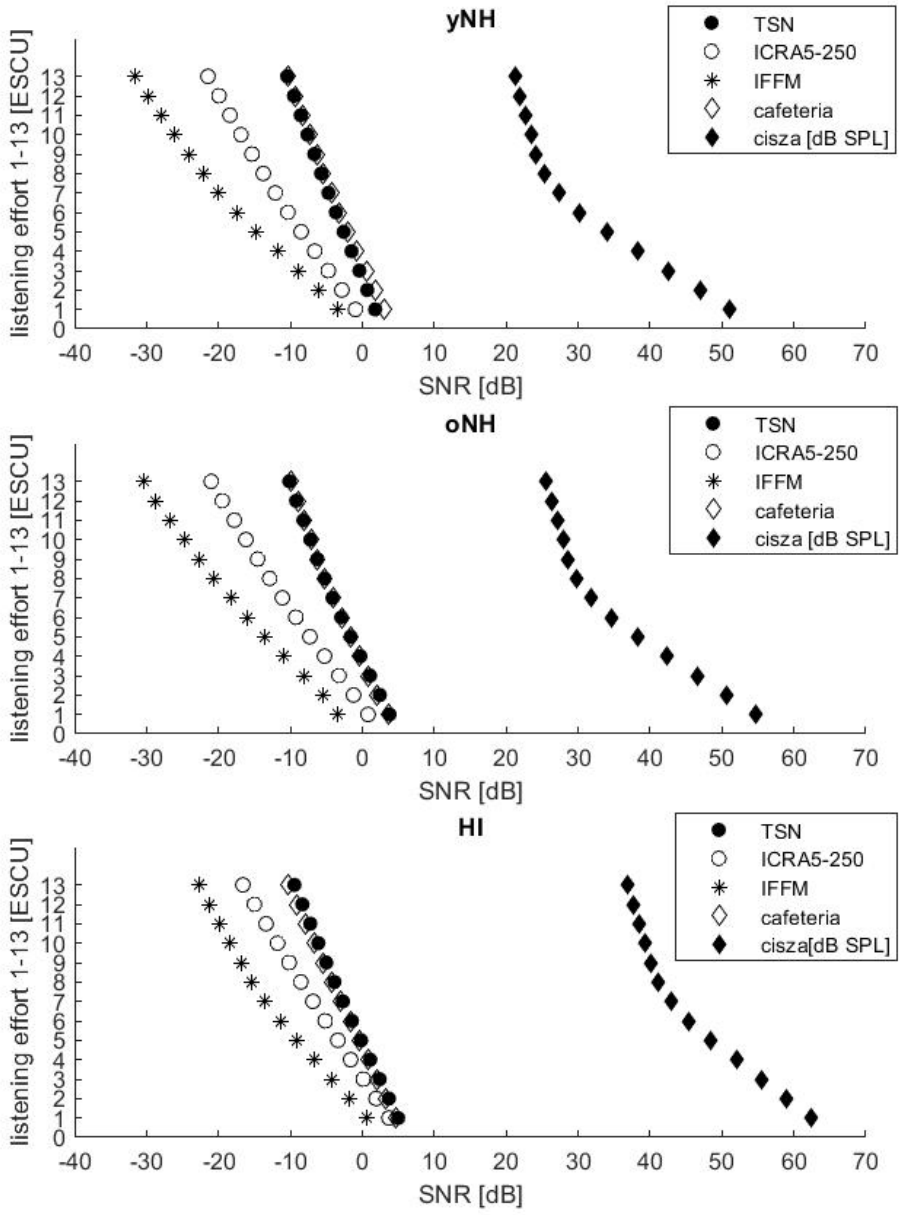


Rys. 42: *Wartości SRT w ciszy w funkcji średnich progów słyszenia (PTA); yNH – młodzi prawidłowo słyszący, oNH – starsi prawidłowo słyszący, HI – osoby z symetrycznym odbiorczym niedosłuchem*

Należy pamiętać, że z punktu widzenia dopasowania aparatów słuchowych, informacje o zrozumiałości mowy w ciszy, które, jak wykazano, można uzyskać na podstawie wyników audiometrii tonalnej, nie są wystarczające do zapewnienia dobrej zrozumiałości mowy w trudnych, rzeczywistych warunkach.W środowiskach akustycznych, w których odbywa się codzienna komunikacja z reguły są obecne różnego rodzaju sygnały zakłócające. Dlatego konieczna jest ocena zrozumiałości mowy w hałasie, co dodatkowo może potwierdzać fakt występowania stosunkowo

słabego związku między progami słyszenia, a zrozumiałością mowy w szumie stacjonarnym (TSN) uzyskanym w tym badaniu współczynnik korelacji wynosi 0.56.

Na Rys. 43 przedstawiono średnie wartości wysiłku słuchowego (wyrażonego w punktach ESCU) w funkcji SNR (lub, w przypadku pomiaru w ciszy – poziomu sygnału testowego) dla 3 grup badanych.



Rys. 43: *Średnia ocena wysiłku słuchowego (listening effort) w funkcji SNR; yNH – młodzi prawidłowo słyszący, oNH – starsi prawidłowo słyszący, HI – osoby z symetrycznym odbiorczym niedosłuchem*

Jak można zauważyć, wykresy funkcji opisujących subiektywnie oceniany wysiłek słuchowy w funkcji stosunku sygnału do szumu mają bardzo podobne przebiegi w przypadku młodszych (yNH) i starszych (oNH) prawidłowo słyszących. Największe różnice występują w przypadku pomiarów w ciszy, choć nie są one statystycznie istotne (jednoczynnikowa analiza ANOVA; F=1.14, p=0.29). Do pewnego poziomu wysiłek słuchowy nieznacznie się zmienia, pomimo stosunkowo dużych zmian poziomu sygnału mowy – krzywe zmian cechują się niewielkim nachyleniem (0.36 w przypadku grupy yNH i 0.37 oNH). Powyżej pewnej wartości krytycznej (około 25 dB SPL w przypadku młodszych i 30 dB SPL starszych) sytuacja ulega zmianie, a wzrost odbywa bardzo dynamicznie. W przedziale 5 dB-owego spadku poziomu, wartość ESCU wzrasta o 6 punktów (więc niemal połowę skali). Podobne tendencje można zaobserwować w przypadku grupy słuchaczy z niedosłuchem. Krzywa określająca wysiłek słuchowy w ciszy jest najmniej stroma (0.43), jednak punkt odcięcia powyżej którego ocena wysiłku dynamicznie wzrasta przy niewielkich zmianach poziomu mowy znajduje się w okolicach 40 dB SPL.

Najbardziej strome przebiegi zmian oceny wysiłku słuchowego w funkcji SNR obserwuje się dla szumów stacjonarnych, we wszystkich grupach słuchaczy (nachylenie krzywych 0.98, 0.85 i 0.83 w szumie TSN kolejno dla grupy yNH, oNH i HI). Oznacza to, że niewielkie zmiany SNR powodują stosunkowo szybki wzrost subiektywnego wysiłku słuchowego. Co więcej, w przypadku szumów stacjonarnych maksymalne wartości skali ESCU (największy wysiłek słuchowy) deklarowane są przy maksymalnie kilkunastodecybelowej różnicy względem wartości 0 ESCU (brak wysiłku) i przy poziomach dużo wyższych niż w przypadku szumów niestacjonarnych. Taka prawidłowość obserwowana jest we wszystkich grupach.

W przypadku szumu fluktuującego IFFM, maksymalny wysiłek słuchowy deklarowano przy poziomach przekraczających -30 dB SNR u yNH oraz oNH, a więc blisko 20 dB niższych niż w przypadku szumów stacjonarnych. Podobnie jest w grupie HI, jednak wartości, przy których deklarowano maksymalny wysiłek słuchowy rzadko przekraczały -20 dB SNR. Inne wartości przyjmuje również nachylenie krzywych. W przypadku prawidłowo słyszących jest ono stosunkowo niewielkie (0.41 i 0.43 dla yNH i oNH), natomiast w grupie HI 0.51.

Podsumowując, w przypadku osób z niedosłuchem, osiągnięcie tych samych wskaźników wysiłku słuchowego w skali ESCU wymaga wyższych poziomów SNR, niż w przypadku grup yNH i oNH. Porównanie poziomów dla 1 ESCU wskazuje na około 3 dB różnicy w przypadku zarówno szumu stacjonarnego, jak i fluktuującego IFFM. Te obserwacje pokrywają się z uzyskanymi przez Krueger at al. (2017) oraz Ibelings (2020).

Na podstawie analizy ocen wysiłku słuchowego można stwierdzić, że postrzegany wysiłek słuchowy wzrasta ze spadkiem SNR (podobnie, jak u Larsby i Arlinger, 2005) i jest zależny od rodzaju sygnału maskującego (jak u Wagener i Brand, 2005) oraz braku bądź obecności niedosłuchu. Nie zaobserwowano jednak tak wyraźnego (jak u Larsby i wsp., 2005) wpływu wieku – wyniki uzyskane dla grupy młodszych i starszych słuchaczy są do siebie bardzo zbliżone. Dla niższych wartości SNR wysiłek słuchowy jest oceniany niżej w przypadku maskerów fluktuujących, w porównaniu z szumami stacjonarnymi, Wraz ze wzrostem SNR natomiast, różnica pomiędzy oceną wysiłku słuchowego dla różnych maskerów maleje, osiągając zaledwie kilku decybelowe (w porównaniu z kilkunasto- lub kilku dziesięcio decybelowymi dla niższych SNR) różnice dla oceny „brak wysiłku". Ta obserwacja jest zgodna z wynikami otrzymanymi przez Krueger i wsp. (2017). W przypadku 10 słuchaczy z grupy yNH przeprowadzono dodatkowe badanie wysiłku słuchowego w trzech najbardziej zróżnicowanych ustawieniach pomiarowych – szumie stacjonarnym TSN, szumie zmodulowanym ICRA5-250 oraz zbliżonym do rzeczywistych warunków (cafeteria), tak aby móc porównać je z wynikami uzyskanymi w wcześniej, w pierwszym badaniu. Odstęp pomiędzy seriami wynosił z reguły 7 dni (maksymalnie 14). Przyjęto założenie, że jeśli wyniki otrzymane przez konkretnych słuchaczy w pierwszej i drugiej sesji byłyby do siebie zbliżone, można stwierdzić, że, w obrębie konkretnej jednostki, wskazania oceny wysiłku są konsekwentne, a pomiary z wykorzystaniem ACALES - wiarygodne i umożliwiające uzyskania powtarzalnych (dla danego badanego) wyników. Przeprowadzono analizę ANOVA z parami pomiarów (test-retest) z dodatkowym czynnikiem czasu (1. sesja, 2. sesja wykonana po 7–14 dniach). Nie stwierdzono istotnej różnicy między wynikami dwóch sesji pomiarowych oraz żadnych istotnych interakcji z czynnikiem czasu (wszystkie $p > 0.05$). Ponadto dokonano analizy współczynnika korelacji wewnątrzklasowej (ICC) dla każdego każdego z trzech maskerów przy minimalnych, maksymalnych oraz środkowych punktach skali ACALES (1, 13 i 7 ESCU; Tab. 15).

Tab. 15: *ICC przy minimalnych, środkowych i maksymalnych punktach skali ACALES*

	1 ESCU	**7 ESCU**	**13 ESCU**	**średnia dla całej skali**
TSN	0.81	0.84	0.76	0.98
ICRA5-250	0.78	0.76	0.69	0.99
cafeteria	0.88	0.78	0.81	0.99

Wartości ICC były wyższe dla niższych ocen wysiłku słuchowego (w tym przypadku 1 i 7 ESCU), które odpowiadają wyższym SNR. Najniższe uśrednione wartości stwierdzono w przypadku górnej granicy skali wysiłku słuchowego określonej w szumie zmodulowanym ICRA5-250 – 0.69. Zgodnie ze statystycznym

standardem Cicchetti (1994), otrzymany wynik świadczy to o wysokim stopniu podobieństwa analizowanych jednostek (w tym przypadku wyników pierwszego i drugiego pomiaru wysiłku słuchowego). Średnie uzyskane dla całej skali we wszystkich badanych warunkach pomiarowych wskazują natomiast na „znakomite" dopasowanie. Międzyosobnicze odchylenia standardowe są natomiast większe niż wewnątrzindywidualne odchylenia standardowe. Oznacza to, że wysiłek postrzegany przez indywidualnego słuchacza może być mierzony z dobrą wiarygodnością, ale istnieją duże indywidualne różnice w tym, jak słuchanie wymagające wysiłku jest postrzegane przez różnych słuchaczy w identycznych sytuacjach.

E7.4 Wnioski z eksperymentu E7

Mimo że wyznaczenie pojedynczej krzywej wysiłku słuchowego trwa kilkanaście minut, a w przypadku potrzeby oceny również w innych warunkach wymagane są dodatkowe pomiary, które mogą doprowadzić do zmęczenia niektórych słuchaczy, procedura ACALES może być uznana za stosunkowo przystępną dla badanych oraz łatwą do wykorzystania praktycznego. ACALES cechuje się wysoką wiarygodnością określoną przez wartość pomiaru test-retest, a także pozwala uchwycić indywidualne różnice z uwzględnieniem np. zmiennych warunków pogłosowych, obecności (i rodzaju) sygnałów zakłócających czy zastosowanie różnych urządzeń wspomagających słyszenie czy metod dopasowania aparatów słuchowych (weryfikacja zysku). Przyszłe prace powinny koncentrować się na opracowaniu modelu wysiłku słuchowego oraz akceptowalnych i łatwych w zarządzaniu metod kwantyfikacji w praktyce klinicznej.

E8 Zrozumiałość mowy przyspieszonej

Rozdzielczość czasowa (bardziej szczegółowo opisana w części teoretycznej rozprawy) zdaje się być bardziej wrażliwa na procesy starzenia się niż zrozumienie mowy w warunkach maskowania widmowego, które warunkowane jest przepustowością filtrów słuchowych (Gehr i Sommers, 1999). Zdaniem Petersa i Moore'a (1992), te ostatnie pozostają niezmienione w zależności od wieku. Z tego powodu zasadne może być wykorzystanie w audiometrii mowy innych sygnałów maskujących niż sam szum stacjonarny. Inną metodą oceny rozdzielczości czasowej w odniesieniu do zrozumiałości mowy jest określenie, do jakiego stopnia może zostać skompresowana w czasie, by pozostała zrozumiała.

Badania dotyczące mowy przyspieszonej sięgają lat 50. XX wieku (Fairbanks i Kodman, 1957). W kolejnych latach została ona włączona do klinicznej oceny uszkodzeń pnia mózgu i kory słuchowej (Beasley i in., 1972), a także podejmowano próby korelowania stopnia kompresji z wiekiem słuchaczy (Gordon-Salant i Fitzgibbons, 1993) lub wielkością ubytku słuchu (Olsen i in., 1975). Obecnie testy wykorzystujące mowę przyspieszoną wykorzystuje się do oceny przetwarzania czasowego u pacjentów z implantem ślimakowym (Fu i in., 2001), a także diagnostyki CAPD wśród dzieci i dorosłych (Stollman i Kapteyn, 1994).

Celem omawianego badania, stanowiącym naturalne uzupełnienie wcześniej omówionych eksperymentów, było określenie progów zrozumiałości mowy SRT50 (50% zrozumiałość) prezentowanej w naturalnym tempie, a także przyspieszonej z wykorzystaniem polskiego testu zdaniowego matrix.

E8.1 Grupa badana

W badaniu wzięło udział 30 słuchaczy ze słuchem prawidłowym, spełniającym kryteria WHO (średnie progi słyszenia nie przekraczające 25 dB HL), których podzielono na dwie grupy – młodszych (yNH) oraz starszych, powyżej 40 r.ż. (oNH). Przyjęto bowiem założenie, że wiek może być zmienną, która w sposób statystycznie istotny wpływa na zrozumiałość mowy przyspieszonej. Szczegółowe dane na temat badanych przedstawiono w Tab. 16.

Tab. 16: *Podstawowe informacje o badanych wraz z odchyleniem standardowym (w nawiasach.*

grupa	**N**	**wiek (SD)**	**PTA (SD)**
yNH	15	25.8 (3.6)	2.5 (1.8)
oNH	15	56.1 (6.2)	10.5 (2.3)

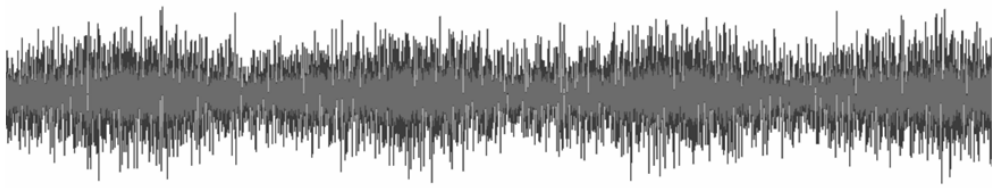
Poza audiometrią tonalną (przewodnictwo kostne i powietrzne) każdy z badanych

został poddany pomiarom otoemisji typu DPOAE (w zakresie częstotliwości: 1–10 kHz). Zgodnie z zapisami literaturowi, chociaż rejestracja otoemisji nie jest badaniem określającym progi słyszenia, pozwala wychwycić nieprawidłowości w obrębie komórek słuchowych zewnętrznych znacznie wcześniej, niż zaczną się one ujawniać w postaci podwyższonych progów słyszenia w podstawowej ocenie audiometrycznej z wykorzystaniem sygnałów tonalnych.

E8.2 Materiał i metody

Badania zrozumiałości mowy przeprowadzono w obecności wykorzystanych i opisanych we wstępie części eksperymentalnej szumów TSN, ICRA5-250 oraz cafeteria, a także dodatkowego sygnału maskującego zawierającego hałas turbiny wiatrowej (ang. *wind turbine*; WT). Wybór nie miał ścisłego uzasadnienia merytorycznego, poza tym, że jest to kolejny, obok ICRA5-250 masker o wyraźnej modulacji amplitudowej. Bez wątpienia jest on maskerem mniej skutecznym, dlatego spodziewany efekt maskowania jest raczej niewielki. Pomysł wykorzystania tego szumu wydał się autorce niniejszej rozprawy ciekawy, z uwagi na osobiste zaangażowanie w projekt polsko-norweskim Hetman[8] dotyczącym dokuczliwości hałasu turbin wiatrowych, a więc, tym samym, łatwy dostęp do próbek zawierających to źródło dźwięku. Idea wykorzystania próbek nagrań rzeczywistych warunków środowiskowych (uwzględniających obecność np. hałasu komunikacyjnego lub lotniczego) nie jest niczym nowym, na co wskazują badania Aniansson (1978). To niezaprzeczalny argument przemawiający za uniwersalnością badań zrozumiałości mowy. Możliwość ich zastosowania nie ogranicza się wyłącznie do audiometrii przeprowadzanej w przypadku diagnostyki słuchu, ale może stanowić narzędzie wspierające ocenę aspektów związanych z codziennym funkcjonowaniem człowieka w danej przestrzeni kształtowanej również przez dźwięk. Wykorzystane nagrania zawierały wyraźną modulację amplitudy o częstotliwości ok. 0.5 Hz i maksymalnej energii w pasmach 200 i 500 Hz. Szczegółowe informacje na temat próbek, z których pozyskano opisywany sygnał maskujący przedstawiono m.in. w publikacji Felcyna (2022). Przebieg czasowy fragmentu sygnału, na którym wyraźnie widoczna jest modulacja amplitudowa przedstawiono na Rys. 44.

[8]Projekt: HETMAN Healthy society - towards optimal management of wind turbines' noise NOR/POLNOR/Hetman/0073/2019-00; szczegółowe informacje na stronie: https://hetman-wind.ios.edu.pl/ [dostęp: 25.10.2024 r.]



Rys. 44: *Przebieg czasowy sygnału zawierającego szum turbiny wiatrowej (WT)*

Jako że wykorzystywanym do oceny zrozumiałości mowy materiałem językowym był polski test zdaniowy matrix, zasadniczą część pomiarową poprzedzone prezentacją dwóch dwudziesto zdaniowych list, aby zniwelować wpływ efektu treningu (szczegółowo opisanego w E2), a samo badanie odbywało się w przedstawionej poniżej konfiguracji (w kolejności losowej dla każdego z uczestników pomiaru).
Poza mową prezentowaną w tempie „normalnym" w badaniu wykorzystano również mowę przyspieszoną (w stopniu określonym przez *acceleration factor*; AF) dwukrotnie oraz 3.7–krotnie. Wybór takich wartości podyktowany był wynikami badań przeprowadzonych przez Niemca i Kocińskiego (2006), w których wartość 3.7-krotnego przyspieszenia mowy pochodzącej z polskiego testu zdaniowego matrix stanowiła progową wartość TCT (opisanego w podrozdziale 5.2.4.):

- mowa „normalna" (współczynnik AF:1)
  - szum stacjonarny TSN o stałym poziomie 65 dB SPL,
  - szum zmodulowany ICRA5-250 o stałym poziomie 65 dB SPL,
  - szum zmodulowany WT o stałym poziomie 65 dB SPL,
- mowa przyspieszona dwukrotnie (współczynnik AF: 2):
  - szum stacjonarny TSN o stałym poziomie 65 dB SPL,
  - szum zmodulowany ICRA5-250 o stałym poziomie 65 dB SPL,
  - szum zmodulowany WT o stałym poziomie 65 dB SPL,
- mowa przyspieszona 3.7–krotnie (zgodnie z wynikami uzyskanymi w badaniach Kocińskiego i in., 2016):
  - szum stacjonarny TSN o stałym poziomie 65 dB SPL,
  - szum zmodulowany ICRA5-250 o stałym poziomie 65 dB SPL,
  - szum zmodulowany WT o stałym poziomie 65 dB SPL.

E8.3 Wyniki i ich dyskusja

Wartością wyznaczoną w sposób adaptacyjny (jak w m.in. E1 czy E4) był próg zrozumiałości mowy SRT określający stosunek sygnału do szumu, przy którym badany prawidłowo powtórzył 50% zaprezentowanego materiału językowego. Uśrednione wartości SRT wraz z odchyleniem standardowym (SD) dla wszystkich warunków pomiarowych przedstawiono w Tab. 17 (a–d).

Tab. 17: *Uśrednione wartości SRT wraz z odchyleniem standardowym (SD) dla wszystkich warunków pomiarowych (a-TSN, b-ICRA5-250, c-cafeteria, d-wind turbine)*

		AF = 1	**AF = 2**	**AF = 3.6**
		TSN		
yNH	SRT [dB]	-7.4	-1.6	6.3
	SD	1.6	2.3	3.4
oNH	SRT [dB]	-6.2	-1.3	7.6
	SD	2.7	3.2	5.4
		ICRA5-250		
yNH	SRT [dB]	-16.3	-6.1	2.9
	SD	2.8	3.2	2.6
oNH	SRT [dB]	-13.6	-4.1	7.9
	SD	4.1	5.5	4.6
		cafeteria		
yNH	SRT [dB]	-7.5	1.8	8.9
	SD	3.0	1.9	2.2
oNH	SRT [dB]	-6.4	2.5	10.2
	SD	1.8	2.6	4.5
		WT		
yNH	SRT [dB]	-15.0	-8.7	-3.9
	SD	2.1	2.8	3.9
oNH	SRT [dB]	-13.3	-6.1	-0.6
	SD	2.6	6.3	5.5

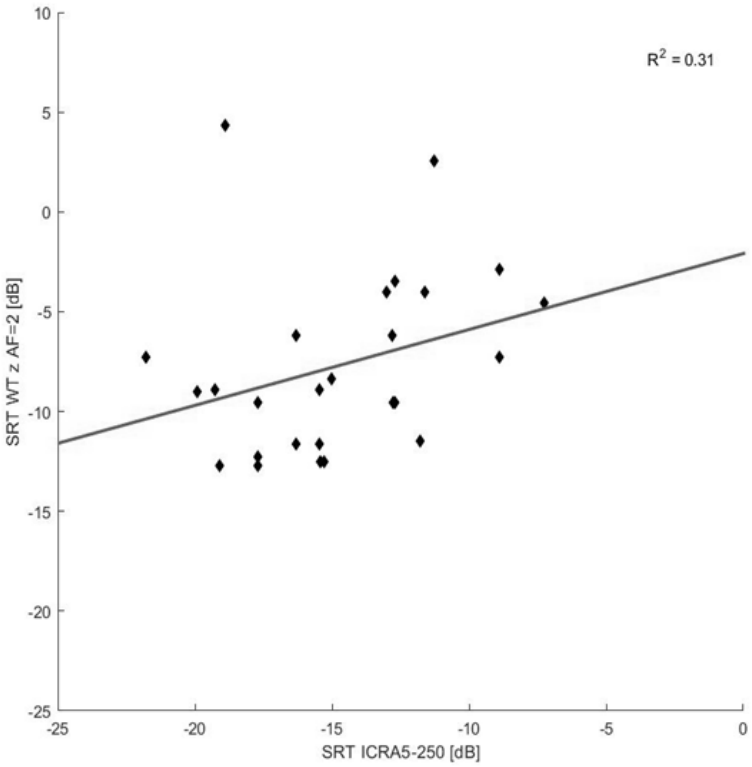
Wartości SRT otrzymane przy mowie prezentowanej w tempie wyjściowym/naturalnym (AF = 1) w ogólności pokrywają się z tymi, które otrzymano w badaniu walidacyjnym (E1) oraz dotyczącym wysiłku słuchowego (E7). Najbardziej skuteczne maskowanie zapewniają szumy stacjonarne –cafeteria oraz TSN. W przypadku szumu ICRA5-250, zarówno w grupie młodszych, jak i starszych słuchaczy progi zrozumiałości mowy osiągają wartości ujemne, rzędu kilkunastu decybeli, co, jak już wykazano (E4), związane jest z efektem *masking release*. Podobnie wygląda sytuacja w obecności wykorzystanego po raz pierwszy szumu WT o charakterystycznym, zmodulowanym charakterze.

Gdy mowa ulega dwukrotnemu przyspieszeniu, jej zrozumiałość ulega osłabieniu, choć dla wszystkich maskerów poza cafeteria odnotowano ujemne wartości SRT.

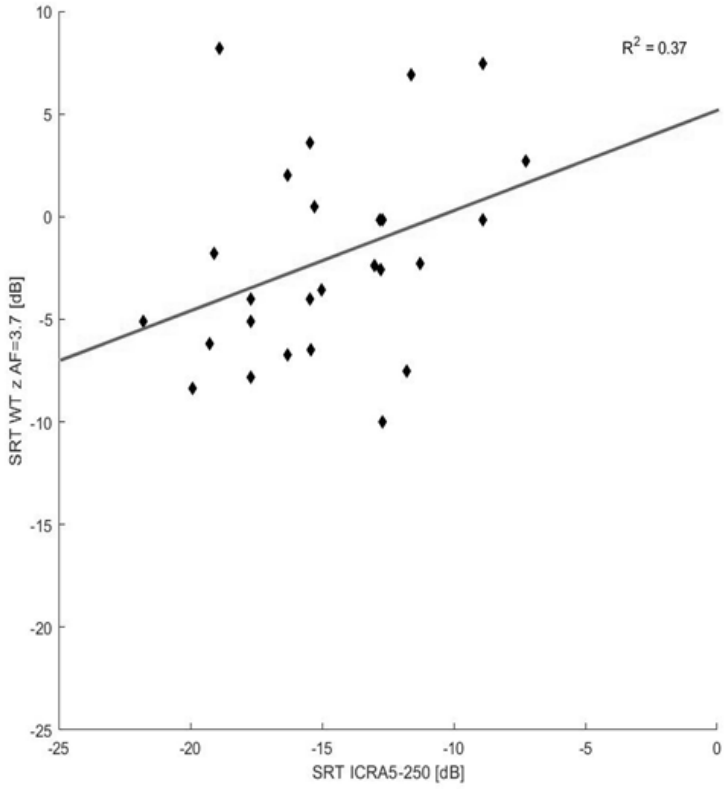
Zmiana AF na 3.6 jeszcze bardziej ogranicza efektywność percepcji. W przypadku szumów stacjonarnych TSN i cafeteria, różnica SRT osiąga około 15 dB (AF = 1 *vs* AF = 3.7) w obydwu grupach wiekowych.

W przypadku szumu ICRA5-250 wpływ wieku wydaje się być silniej zaznaczony. Chociaż średnia redukcja SRT w grupie młodszych badanych (blisko 20 dB) i starszych (21.5dB) jest podobna, istotna jest także wzajemna relacja tych wielkości. U yNH progi kształtują się na poziomie ok. 3 dB, natomiast w grupie oNH ponad dwa razy więcej (8 dB). Szum WT jako jedyny maskuje mowę na tyle nieefektywnie, że otrzymane SRT są ujemne. Prawdopodobnie wynika to z odległości pomiędzy kolejnymi modulacjami. W mowie przyspieszonej, zdania pochodzące z testu w całości mieszczą się w lukach czasowych maskera. Stąd prezentowane są niemal wyłącznie w ciszy. Obniżenie SRT o kilkanaście dB względem wyjściowego tempa prezentacji mowy związane jest więc przede wszystkim ze zbyt dużym elementów słownych (sylab, wyrazów) przypadających na jednostkę czasu, a nie udziałem sygnału maskującego *per se*.

Ciekawym aspektem, który całkiem niedawno pojawił się w literaturze jest powiązanie mowy przyspieszonej z szumami fluktuującymi. Badania prowadzone przez Gransiera i wsp. (2022) wskazują, że osoby, które dobrze radzą sobie ze słuchaniem mowy prezentowanej w szumie zmodulowanym osiągają wysokie wartości zrozumiałości mowy przyspieszonej. Takie spojrzenie na ocenę zrozumiałości mowy nie było brane pod uwagę na etapie konstruowania protokołu pomiarowego. Mimo to przeprowadzono analizę potencjalnego związku wartości SRT w szumie zmodulowanym ICRA5-250 przy naturalnym tempie prezentacji sygnału mowy (AF = 1) z SRT uzyskanymi w obecności szumu WT (ze względu na brak pomiaru w ciszy) po zastosowaniu kompresji czasowej (AF = 2 oraz AF = 3.7). Niestety, analiza korelacji nie wykazała jej obecności (wsp. korelacji Pearsona R^2 wynosi kolejno 0.31 oraz 0.37), chociaż można zaobserwować trend, zgodnie z którym niskim wartościom SRT zmierzonym w szumie stacjonarnym ICRA5-250 towarzyszy lepsza zrozumiałość mowy przyspieszonej. Związek ten jest wyraźniejszy dla wyżej wartości AF. Szczegółowe dane przedstawiono na Rys. 45 i 46 (grupy yNH oraz oNH zostały potraktowane łącznie, N = 30).



Rys. 45: *Związek progów SRT w szumie zmodulowanym ICRA5-250 (AF = 1) z SRT w szumie turbiny wiatrowej WT (AF = 2)*



Rys. 46: *Związek progów SRT w szumie zmodulowanym ICRA5-250 (AF = 1) z SRT w szumie turbiny wiatrowej WT (AF = 3.7)*

Porównania między grupami yNH oraz oNH dla wszystkich rodzajów maskerów oraz wartości współczynnika AF przeprowadzone za pomocą testu Manna-Whitneya nie wykazały statystycznie istotnych różnic między grupami, z wyjątkiem rezultatów uzyskanych w szumie ICRA przy tempie mowy 3.7 (najwyższym badanym). Wynik ten jest szczególnie interesujący, gdyż, wbrew pierwotnym założeniom, nie zaobserwowano znaczącego wpływu wieku na poziom zrozumiałości mowy, nawet w warunkach, w których mowa była nie tylko częściowo zamaskowana, ale i poddana kompresji czasowej. Pomimo powszechnie przyjmowanej tezy o pogarszaniu się zdolności percepcyjnych wraz z wiekiem, której stawianie uzasadnia wiele badań przytoczonych w niniejszej rozprawie, w tym konkretnym eksperymencie nie potwierdzono takiego związku.
Ważnym aspektem, który należy wziąć pod uwagę analizując uzyskane wyniki są wielkości odchylenia standardowego progów zrozumiałości mowy, szczególnie w odniesieniu do mowy przyspieszonej. W grupie starszych prawidłowo słyszących wartość SD była wyższa w porównaniu z grupą młodszych badanych, co sugeruje większy wpływ indywidualnych czynników skutkujący znaczną zmiennością wyników. W przypadku tempa mowy o AF równym 3.7, wartość odchylenia standardowego przekroczyła 4.5 dB, co stanowi ważne wskazanie na zróżnicowanie zdolności percepcyjnych w grupie starszych słuchaczy. Taka heterogeniczność może wynikać z wielu czynników, zarówno poznawczych, jak i fizjologicznych. Ma także istotne znaczenie kliniczne oraz praktyczne.

E8.4 Wnioski z eksperymentu E8

Wzorem wcześniej przytoczonych badań z części eksperymentalnej niniejszej tezy, uzyskane w eksperymencie E8 wyniki stanowią niepodważalny dowód na konieczność uzupełniania diagnostyki słuchu odbywającej się przy wykorzystaniu tonów, również mową, najlepiej prezentowaną w hałasie.
Po drugie, rezultaty te wskazują na fakt ogromnej złożoności ludzkiej percepcji słuchowej, zarówno na etapie peryferyjnym, jak i centralnym. Mimo rozwijających się metod oceny psychofizjologicznej, uwzględniającej również neuroobrazowanie, dokładne mechanizmy leżące u podstaw obserwowanej zmienności międzyosobniczej nie zostały jednoznacznie zdefiniowane.
Po trzecie, obserwowana zmienność wyników między badanymi starszymi słuchaczami może stanowić wczesny wskaźnik trudności percepcyjnych, wymagających monitorowania. Choć wniosek ten może wydawać się pochopny, to mając na uwadze chociażby treści i wyniki przedstawione w części eksperymentalnej

oraz wiedząc, jak istotny wpływ na funkcjonowanie jednostki w szeroko rozumianym środowisku mają zaburzenia poznawcze, przedstawione przez autorkę niniejszej rozprawy sugestie warte są dokładnego zgłębienia, również w postaci dalszych badań, które powinny objąć swoim zasięgiem większą grupę badanych.

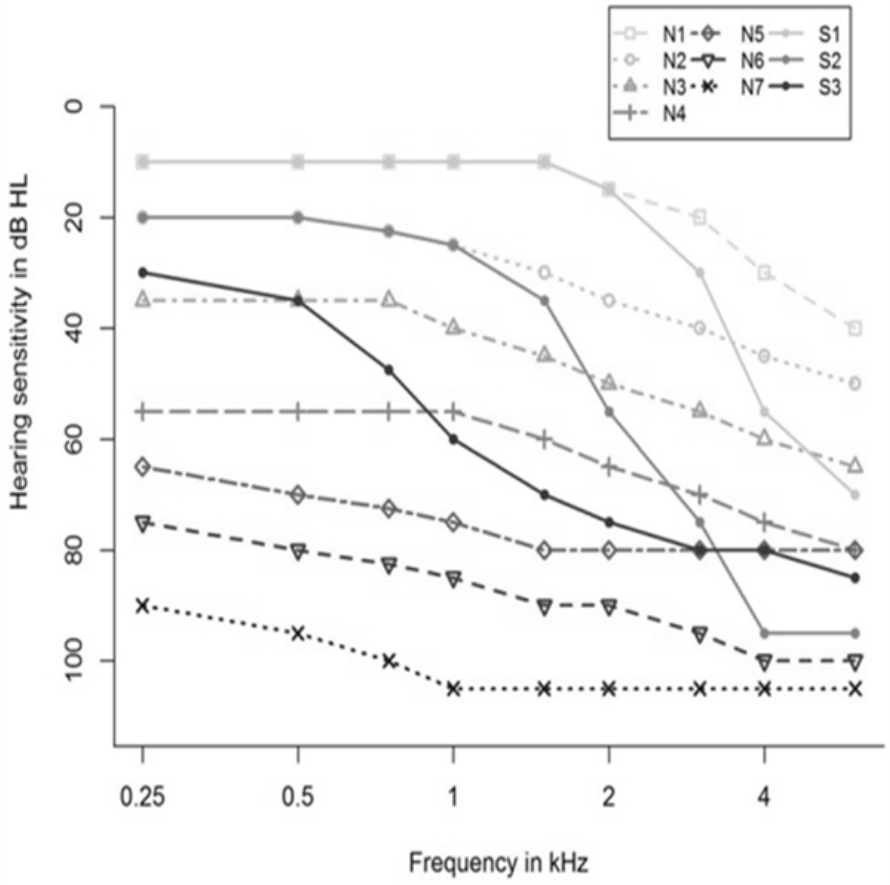
E9 Weryfikacja dopasowania uzyskanych wyników do klasyfikacji Bisgaarda

Jak wspomniano we wcześniejszych częściach rozprawy, w praktyce audiologicznej standardem pozwalającym na klasyfikację stopnia niedosłuchu są kryteria zaproponowane przez WHO, zgodnie z którymi ubytki słuchu można podzielić, ze względu na średnią wartość progu słyszenia (liczoną jako średnia arytmetyczna dla częstotliwości 500, 1000, 2000 i 4000 Hz), na:

- lekki (średni próg słuchu: 21–40 dB HL),
- umiarkowany (średni próg słuchu 41–70 dB HL),
- znaczny (średni próg słuchu: 71–90 dB HL),
- głęboki (średni próg słuchu powyżej 91 dB HL).

Jednoliczbowy wskaźnik, choć często wykorzystywany w regulacjach z zakresu opieki zdrowotnej, w praktyce jest mało informatywny. Więcej danych dostarcza audiogram – znajomość jego przebiegu dla konkretnych częstotliwości pozwala precyzyjniej określać np. wpływ ubytku na procesy przetwarzania informacji słuchowej. Przykładem potwierdzającym to stwierdzenie są choćby tzw. hałasowe uszkodzenia słuchu, które, w początkowej fazie, objawia się podwyższeniem progów słyszenia, najczęściej w zakresie 4–6 kHZ. Pomysł wykorzystania ustandaryzowanych audiogramów został po raz pierwszy zasugerowany przez Nordic Cooperation on Disability (NSH), a służyć miał bardziej dokładnej ocenie działania aparatów słuchowych i zysku percepcyjnego wynikającego z ich wykorzystania, również na etapie ich prognozowania, przed faktycznym dopasowaniem aparatów pacjentowi. Weryfikacja zaproponowanych w 2003 r. pięciu klas audiogramów okazała się być niepomyślna – w analizowanej próbie składającej się z 15 000 audiogramów zarejestrowanych w jednym ze sztokholmskich szpitali, do modelowych przebiegów udało się dopasować jedynie 26% zebranych danych. Poszukiwanego innego podejścia, które byłoby, z jednej strony było praktyczne, a więc uwzględniało najczęściej występujące w praktyce klinicznej przypadki, ale z drugiej, pokrywało jak najszerszy zakres ubytków słuchu. Zdecydowano się zastosować procedurę kwantyzacji wektorowej (*Vector Quantization*) polegającej na rozpoznawania struktur (wzorów) wejściowych i aproksymowania ich za pomocą jednego z wcześniej zdefiniowanych wzorców. Danymi wejściowymi uczyniono zbiór blisko 30 tys. audiogramów zgromadzonych w latach 2001–2004 w Stockholm South Hospital. Na

drodze skomplikowanych analiz statystycznych opisanych szczegółowo w artykule Bisgaarda, Vlaminga i Dahlquista (2010) opracowano model (Rys. 47), składający się z 10 najbardziej typowych przebiegów progów słuchu określonych w audiometrii tonalnej dla przewodnictwa powietrznego, z podziałem na ubytki o płaskim przebiegu audiogramu (klasy N1–N7) oraz stromo opadające krzywe (typowe dla ubytków wysokoczęstotliwościowych; klasy S1–S3). Dzięki nim możliwe jest dość precyzyjne dopasowanie do indywidualnego ubytku słuchu konkretnego badanego. Tak zaproponowany podział pozwolił zwiększyć możliwość dopasowania modelu do faktycznie rejestrowanych przebiegów progu słyszenia o 20 punktów procentowych względem sugestii NSH, prowadząc do zgodności rzędu 46%.



Rys. 47: *Klasy audiogramów zaproponowane przez Bisgaarda* [*za* : *Bisgaardiin.*, 2010]*; na osi x znajduje się częstotliwość (w kHZ), natomiast y – wartość progu dla danej częstotliwości (w dB HL)*

E9.1 Grupa badana

W przedstawionej poniżej analizie zweryfikowano skuteczność dopasowania wykorzystując część wyników uzyskanych w ramach realizacji pracy doktorskiej. Pod uwagę wzięto jedynie osoby z niedosłuchem (symetrycznym, odbiorczym),

a wszystkie dane musiały pochodzić z badań przeprowadzonych z wykorzystaniem dokładnie takiej samej metody (w oparciu o ten sam protokół badania). W analizach nie ma grup z największymi ubytkami słuchu, tj. S3 oraz N5, bo takowe nie brały udziału w eksperymentach. Uwzględniając wymienione założenia, w analizie wzięto pod uwagę 78 słuchaczy (dane w Tab. 18)

Tab. 18: *Liczba słuchaczy przypisanych do odpowiednich grup wg klasyfikacji Bisgaarda*

grupa	**N1**	**N2**	**N3**	**N4**	**S1**	**S2**
liczba osób	10	14	19	10	10	15

E9.2 Materiał i metody

W analizie wzięto pod uwagę wyniki pomiarów realizowanych w ramach wyżej wymienionych eksperymentów w zakresie progów słyszenia określonych w audiometrii tonalnej oraz zrozumiałości mowy w szumach i ciszy. Wybrano tylko osoby z niedosłuchem (klasyfikując je do odpowiedniej grupy wg wytycznych Bisgaarda) biorące udział w badaniach prowadzonych przy wykorzystaniu tej samej metodologii:

- zrozumiałość mowy określana w pomiarze słuchawkowym w tym samym pomieszczeniu,
- mowa prezentowana na poziomie zmieniającym się w sposób adaptacyjny na tle szumu o stałym poziomie 65 dB SPL – wybrano maskery SRT, ICRA5-250 oraz cafeteria, a także badanie w ciszy,
- pomiar zasadniczy poprzedzony dwoma listami pomiarowymi (z uwagi na efekt treningu),
- formuła *closed-set*.

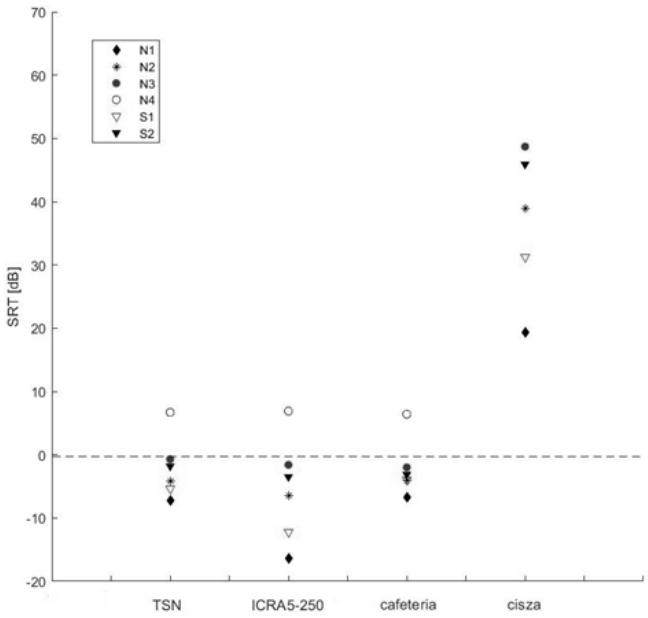
Klasy Bisgaarda zestawiono ze średnimi wartościami SRT we wszystkich badanych warunkach maskowania aby ocenić, czy zastosowanie klasyfikacji bazującej na progach słyszenia umożliwia grupowanie pacjentów również w kontekście stopnia zrozumienia mowy.

E9.3 Wyniki i ich dyskusja

Szczegółowe wyniki (wartości średnie SRT wraz odchyleniem standardowym) uzyskane dla wszystkich klas niedosłuchu i warunków pomiarowych przedstawiono w Tabeli 19 oraz na Rys. 48.

Tab. 19: *Uśrednione wartości SRT (wraz z odchyleniem standardowym) w analizowanych klasach (N1, N2, N3, N4 oraz S1, S2) i warunkach pomiarowych*

		SRT	SD
TSN	**N1**	-7.2	2.1
	N2	-4.2	2.1
	N3	-0.7	5.7
	N4	6.7	6.5
	S1	-5.3	1.4
	S2	-1.8	2.5
ICRA5-250	**N1**	-16.4	3.3
	N2	-6.5	4.6
	N3	-1.6	7.4
	N4	6.9	5.7
	S1	-12.2	2.9
	S2	-3.5	4.7
cafeteria	**N1**	-6.7	1.6
	N2	-4	3.2
	N3	-2	4.3
	N4	6.4	6.1
	S1	-3.9	3.3
	S2	-3.1	2.5
cisza	**N1**	19.4	4.3
	N2	38.9	10.2
	N3	48.7	12.3
	N4	70.4	10.2
	S1	31.2	5.9
	S2	45.9	15.8



Rys. 48: *Średnie wartości SRT per grupa z klasyfikacji Bisgaarda w ocenianych warunkach maskowania*

Wykorzystując test Bonferroniego przeprowadzono analizę *post hoc* porównując w parach poszczególne klasy ubytków słuchu i wartości średnie progów zrozumiałości mowy (SRT) wyznaczonych w pięciu szumach maskujących oraz w ciszy. Za hipotezę zerową uznano brak różnic między wziętymi pod uwagę klasami Bisgaarda w kolejnych pomiarach z obecnością szumu maskujacego lub ciszy. Uzyskane wyniki przedstawiono w Tab. 20, w której 1 oznacza istotność statystyczną różnic pomiędzy wynikami zrozumiałości mowy w danej parze klas ubytków słuchu (p <0.005), natomiast 0 – brak istotności.

TAB. 20: *Analiza post hoc – istotność statystyczna różnic SRT w poszczególnych szumach maskujących w danej parze klas ubytku słuchu (1 – istotność statystycza różnicy, 0 – brak istotności)*

para klas	**TSN**	**ICRA5-250**	**cafeteria**	**cisza**
N1 N3	1	1	1	1
N2 N3	0	0	0	0
N2 N1	0	1	0	1
S2 N3	0	0	0	0
S2 N1	1	1	0	1
S2 N2	0	0	0	0
N4 N3	1	1	1	1
N4 N1	1	1	1	1
N4 N2	1	1	1	1
N4 S2	1	1	1	1
S1 N3	0	1	0	1
S1 N1	0	0	0	0
S1 N2	0	0	0	0
S1 S2	0	1	0	1
S1 N4	1	1	1	1

Przeprowadzona analiza post hoc wykazała, że dla niektórych par klas ubytków słuchu, różnice między nimi nie są istotne statystycznie, bez względu na warunki prezentacji sygnału mowy (rodzaj zastosowanego maskera). Dotyczy to par klas:
N2-N3, S2-N3, S2-N2, N1-S1 oraz S1-N2. W ich przypadku klasyfikacja zaproponowana przez Bisgaarda wydaje się być szczegółowa – podobne wartości SRT określające stopień zrozumiałości mowy otrzymały osoby o różniących się przebiegach progów słyszenia.
W przypadku par: N1-N2, N3-N4, N4-N1, N4-N2, N4-S2 oraz S1-N4 odnotowane różnice były istotne statystycznie dla wszystkich badanych warunków. Co ciekawe wśród tego zbioru są pary skrajnie różniące się przebiegami audiogramów (tj. N1-N4 oraz S1-N4), w których rezultaty takie wydają się być oczywiste, ale również pary klas ze sobą sąsiadujących (N1-N2 oraz N3-N4). Istotne statystycznie różnice zaobserwowano także dla tych samych par klas w szumach TSN i cafeteria

oraz ICRA5-250 i ciszy – jeżeli odnotowano je w szumie stacjonarnym TSN, obecne były także w przypadku szumu cafeteria. W parach, w których różnice istotne statystycznie wystąpiły w szumie zmodulowanym ICRA5-250, pojawiły się one również w ciszy. Warto zauważyć, że różnice te nie zawsze występowały między skrajnie różniącymi się (stopniem niedosłuchu) klasami. To bez wątpienia interesujący aspekt, która warto zgłębić po rozszerzeniu bazy zbadanych osób.

E9.4 Wnioski z eksperymentu E9

W autorskich badaniach przedstawionych w części eksperymentalnej rozprawy udowodniono, iż głównym czynnikiem wpływającym na efektywność percepcji słuchowej w zakresie zrozumiałości mowy, zwłaszcza prezentowanej w obecności sygnałów maskujących, jest występowanie ubytku słuchu. W analizach dużych zbiorów danych w pełni zasadne wydaje się być wykorzystanie grupowania pacjentów z zależności od głębokości niedosłuchu. Owocem tej strategii jest np. klasyfikacja zaproponowana przez WHO. Szerszej informacji dostarcza jednak wiedza o przebiegu całego audiogramu, a nie jedynie jednoliczbowy wskaźnik określany jako średnia arytmetyczna progów zmierzonych dla czterech częstotliwości. Naprzeciw tym potrzebom wyszedł Bisgaard definiując kilkanaście klas audiogramów z, co ważne, uwzględnieniem ich stromości, tworząc najbardziej typowych „profile" słuchaczy badań psychofizycznych lub pacjentów audiologicznych. W przypadku analizowanych danych pochodzących z pomiarów na blisko 80 osobach z symetrycznym niedosłuchem odbiorczym stwierdzono, że systematyzacja dzieląca grupę badawczą na klasy w zależności od przebiegu progów słyszenia określonych w audiometrii tonalnej okazała się być zbyt szczegółowa – obserwowane pomiędzy klasami różnice nierzadko były niewielkie (nieistotne statystycznie), więc podobne rezultaty były osiągane przez badanych zakwalifikowanych do różnych klas. Są to wyniki zbieżne z uzyskanymi w pracy magisterskiej autorki niniejszej rozprawy. Oczywiście nie jest to jednoznaczny argument przemawiający za niewystarczającym dopasowaniem modelu do faktycznych cech warunkujących właściwą percepcję w zakresie zrozumiałości mowy. Na uwagę zasługuje fakt, iż klasyfikacja Bisgaarda jest pierwszą, która bierze pod uwagę kształt audiogramu, a więc, pośrednio, wpływ ubytków wysokoczęstotliwościowych, których obecność wielokrotnie „ginie", jeśli progi słyszenia definiuje się wyłącznie poprzez jednoliczbowy wskaźnik PTA. Być może potrzebne są dalsze badania, niekoniecznie stawiające za cel walidację zaproponowanej klasyfikacji, a raczej uwzględnienie jej w dyskusji dotyczącej metod oceny zrozumiałości mowy,

E10 Wysiłek słuchowy (*listening effort*) w salach o zmiennych cechach akustycznych

Wśród czynników wpływających na zrozumiałość mowy wymienionych w części teoretycznej rozprawy, a z których kilka poddano analizie w wyżej opisanych eksperymentach, znajdują się warunki propagacji fali akustycznej. W odniesieniu do rzeczywistych sytuacji komunikacyjnych chodzi przede wszystkim o szeroko rozumianą akustykę wnętrz. To właśnie ona, wpływając na fizyczne cechy dźwięku, pełni niebagatelną rolę w kształtowaniu efektywności percepcji słuchowej (często współuczestniczącej z aspektami wizualnymi, np. w salach teatralnych). Nie bez znaczenia pozostaje również wysiłek poznawczy, który jej towarzyszy i razem z nią składa się na pełne doświadczenie słuchowe. W części E7 przedstawiono badanie z wykorzystaniem przetłumaczonej przez autorkę niniejszej rozprawy i wykorzystanej po raz pierwszy dla języka polskiego skali ACALES. Choć świadomość współobecności wysiłku słuchowego nie jest niczym nowym (mimo że został niejednoznacznie zdefiniowany dopiero kilka lat temu – szczegółowe informacje w rozdziale E7), potencjał badań, które go dotyczą nadal nie jest w pełni wykorzystany. Szczególnie, że poza płaszczyzną audiologiczną i uwzględnieniem jego oceny np. w dopasowaniu aparatów słuchowych, pomiary *listening effort* można rozszerzyć do ogólniej pojmowanego słyszenia, a więc np. do ewaluacji metod transmisji sygnału czy cech akustycznych pomieszczenia, w którym się on propaguje. To właśnie drugie zagadnienie ma istotne znaczenie w badaniu opisywanym w niniejszym rozdziale.

W badaniu realizowanym przez autorkę rozprawy doktorskiej z Błasińskim i Kocińskim, skupiono się na ocenie zrozumiałości mowy w różnych warunkach akustycznych i zestawienie uzyskanych rezultatów z obiektywnymi parametrami charakteryzującymi akustykę pomieszczeń.. Analiza miała na celu zidentyfikowanie ewentualnych związków między subiektywną oceną a miarami fizycznymi, w tym indeksem transmisji mowy (STI), klarownością (C50), czasem pogłosu (RT) i czasem wczesnego zaniku (EDT).

Pierwszy z parametrów, STI (ang. *Speech Transmission Index*) pochodzi z wyników pomiarów funkcji przeniesienia modulacji (Houtgast & Steeneken, 1973) i określa ilościowo jakość sygnału w skali od 0 (słaba zrozumiałość) do 1 (doskonała zrozumiałość). W przypadku pomieszczeń analizowanych w niniejszym badaniu, STI został sklasyfikowany jako „dostateczny" bądź „dobry".

Klarowność, C50, miara wprowadzona przez Marshalla (1994), reprezentuje logarytmiczny stosunek wczesnej do późnej energii dźwięku, gdzie „wczesna" odnosi

się do początkowych 50 (lub 80) milisekund, a „późna" oznacza czas trwania po tym okresie (Kociński i Ozimek, 2017). Próg 50 milisekund odgrywa kluczową rolę w odróżnianiu korzystnych odbić od destrukcyjnych i ma zasadnicze znaczenie dla oceny zdatności pomieszczenia do właściwej percepcji mowy.

Rozpowszechniony w literaturze czas pogłosu (*Reverberation Time*; RT) jest definiowany jako czas potrzebny do obniżenia poziomu ciśnienia akustycznego (SPL) określonego źródła dźwięku o 60 dB po jego nagłym wyłączeniu (Sabine, 1922). Ze względu na ograniczenia techniczne (osiągnięcie zakresu pomiarowego przekraczającego 60 dB jest często niemożliwe), zwykle stosuje się pomiary RT20 i RT30 z punktem początkowym 5 dB poniżej poziomu energii w stanie ustalonym (ISO 3382-2, 2010). W takim przypadku RT20 jest trzykrotnością spadku SPL o 20 dB, a RT30 dwukrotnością spadku SPL o 30 dB). W przypadku wydarzeń akustycznych, w których pojawia się zarówno muzyka, jak i tekst zaleca się RT około 1 sekundy (Sakai i in., 1998; Ando, 2007), a w przypadku muzyki bez tekstu preferowane są dłuższe RT (Kuhl, 1954). W teatrach sugerowany RT może wynosić do 1.6 s. Prawidłowa zrozumiałość mowy w audytoriach, dużych salach lekcyjnych i innych pomieszczeniach, w których mowa pozostaje głównym sygnałem, wymaga niższego RT, począwszy od 0.5–0.7 s do 1 s (Everest, 2001). Wartości w tym zakresie zapobiegają nakładaniu się odbitych dźwięków na sygnał bezpośredni, co jest częstym problemem w przypadku zbyt długich RT. Subiektywne postrzeganie pogłosu zależy od poziomu wzbudzenia i hałasu, co jest bardziej zbliżone do czasu wczesnego zaniku (EDT), który odnosi się do początkowej i najwyższego poziomu części zanikającej energii, a nie samego RT (Ahnert i Tennhardt, 2008).

Ciekawym aspektem realizowanych pomiarów było wykorzystanie odpowiedzi impulsowych zarejestrowanych w salach z systemem RES (*Reverberation Enhancement System*), które zostały splecione z materiałem odsłuchowym. Systemy RES, na drodze cyfrowego przetwarzania sygnału, mają zdolność do zwiększania wczesnych odbić i modulowania RT, przy jednoczesnym zachowaniu właściwości akustycznych pomieszczenia (Bakker & Gillian, 2014). Jako rozwiązania gwarantujące kontrolę i precyzję większe niż tradycyjne metody pasywne (np. panele pochłaniające), w ostatnich latach RES zyskują na popularności.

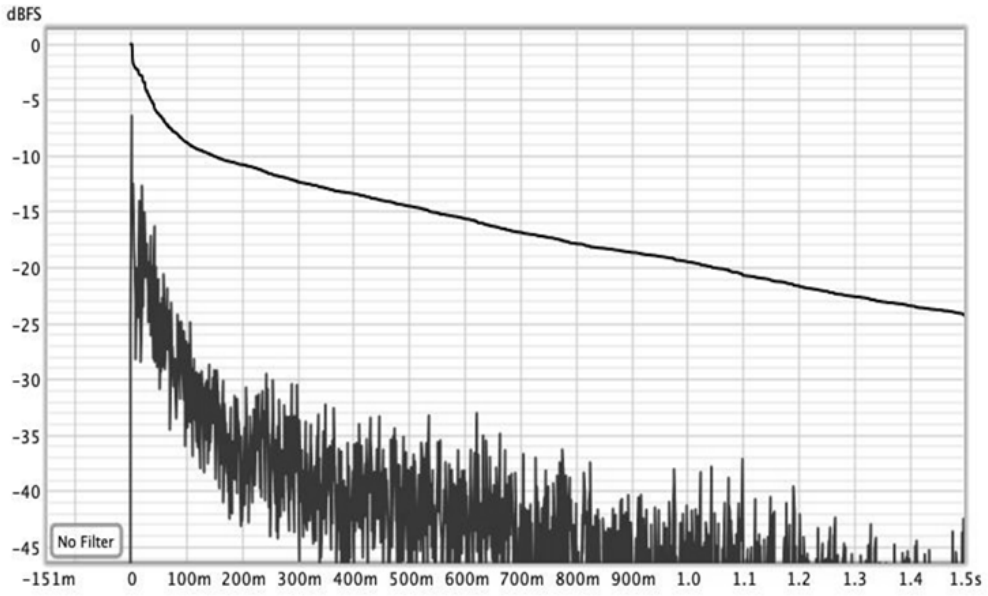
Do tej pory nie przeprowadzano jednak w ich kontekście badań w zakresie zrozumiałości mowy, zatem opisywane badanie stanowi *novum*.

E10.1 Grupa badana

W pomiarach wzięły udział trzy grupy po 20 pełnoletnich osób (średni wiek: 23.5). Każda z nich przed badaniem była poddana ocenie otoskopowej audiometrii tonalnej. Wybrano wyłącznie słuchaczy ze słuchem prawidłowym (zgodnie z kryteriami WHO).

E10.2 Materiał i metody

Badanym przedstawiono 50-elementowe (jak zalecają Howard & Angus, 2017), zrównoważone fonematycznie listy logatomowe w języku polskim (Brachmański i Staroniewicz, 1999) splecione z odpowiedziami impulsowymi 3 sal z zainstalowanym RES (z każdej z 3 ustawieniami systemu). Wybór materiału językowego podyktowany był chęcią oceny wyłącznie odbioru sygnału mowy, bez jej dalszej analizy, czy interpretacji przez wykorzystanie np. kontekstu. Sygnał mowy prezentowano przez słuchawki Sennheiser HDA201 (z przedwzmacniaczem SR46OH DOD) w pomieszczeniu spełniającym normę ANSI/ASA S3.1-1999. Układ pomiarowy skalibrowano z wykorzystaniem miernika Brüel & Kjaer 2203 oraz sztucznego ucha Brüel & Kjaer 4152. W pierwszej części eksperymentu (prowadzonej i opisanej przez Błasińskiego, (2023)) skupiono się głównie na pomiarach parametrów obiektywnych sal z RES oraz zestawieniu ich z miarą zrozumiałości mowy (dokładniej: wyrazistością logatomową). Okazało się, że gdy ustawienia RES generują krótki czas pogłosu, zrozumiałość mowy kształtuje się w sposób tożsamy z pomieszczeniami z naturalnym pogłosem, co zostało opisane m.in. przez Houtgasta i Steenekena (1985). Jeśli jednak ustawienia RES wydłużają czas pogłosu to powyżej mniej więcej 2.5s nie obserwuje się znacznego pogorszenia zrozumiałości mowy wraz ze wzrostem RT, jak ma to miejsce w salach z naturalnie występującym długim pogłosem. Istnieje wiele potencjalnych wyjaśnień takiej prawidłowości. Jednym z nich może być nieliniowy charakter krzywej zaniku rejestrowany dla pomieszczeń z RES (tzw. krzywa *double-slope*; Rys. 49).



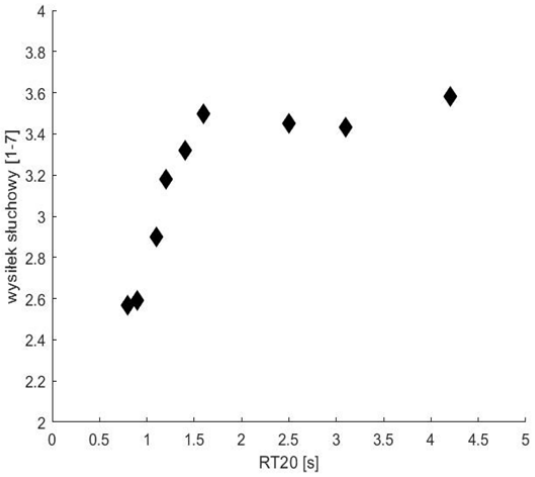
Rys. 49: *Przykładowa krzywa zaniku o podwójnym nachyleniu – amplituda [dBFS] w funkcji czasu* (na podst.: Błasiński, 2023)

Zasadniczym celem części badania omawianej w tej rozprawie jest jednak *listening effort*. Oprócz testu zrozumiałości mowy, uczestnicy oceniali więc również subiektywnie postrzegany wysiłek słuchowy używając 7-stopniowej skali Likerta (Likert, 1932), w zakresie od 1 (brak wysiłku) do 7 (bardzo duży wysiłek). Jest to narzędzie dużo mniej dokładne niż ACALES (E7), ale pozwala szybko i w prosty sposób otrzymać wiarygodne wyniki, jak np. w badaniu Johnson i in, 2015.
W Tab. 18 przedstawiono uśrednione wartości wysiłku słuchowego (wraz z odchyleniem standardowym – w nawiasie) oraz zmierzone wartości RT.

Tab. 21: *Uśrednione wartości wysiłku słuchowego (LE) oraz zmierzone wartości czasu pogłosu (RT)*

	RT	**LE [1–7]**
1	0.8	2.6 (1.2)
	0.9	2.6 (1.2)
	1.1	2.9 (1.2)
2	1.2	3.2 (1.4)
	1.4	3.3 (1.5)
	1.6	3.5 (1.5)
3	2.5	3.3 (1.5)
	3.1	3.4 (1.5)
	4.2	3.6 (1.6)

Wprawdzie można zaobserwować trend, zgodnie z którym zgłaszane wartości wysiłku słuchowego rosną wraz ze zwiększaniem się RT, jednak współczynnik korelacji Pearsona nie wskazuję silnego związku ($R^2 = 0.55$) – Rys 50. Podobnie jest w przypadku EDT ($R^2 = 0.56$) oraz C50 ($R^2 = 0.46$).

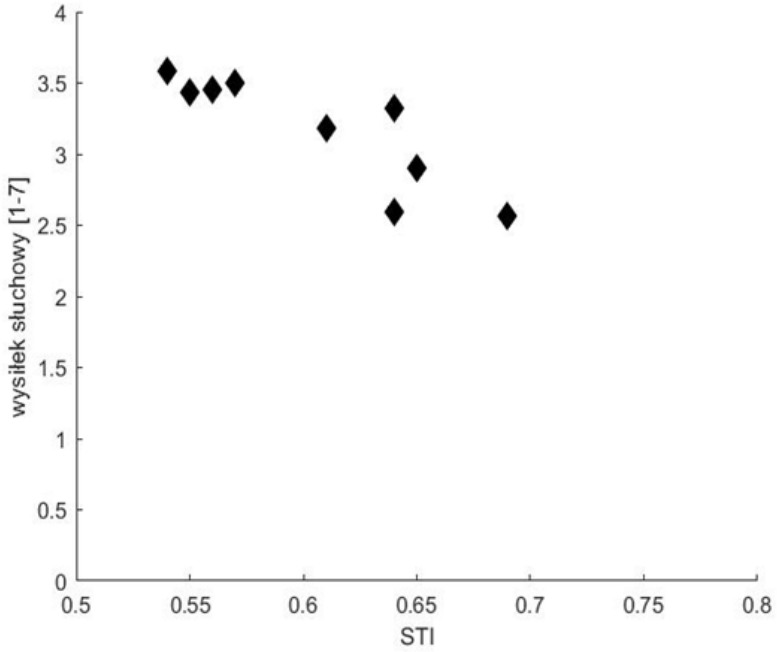


Rys. 50: *Związek RT20 z wysiłkiem słuchowym w skali Likerta (1–7)*

Średnie oceny wysiłku słuchowego zestawiono, podobnie jak w badaniu E7 z miarami zrozumiałości mowy. W tym przypadku były one dwojakiej natury – wyznaczona obliczeniowo (STI) oraz określona w testach odsłuchowych z listami logatomowymi (% zrozumiałość mowy). Porównano obiektywną charakterystykę pomieszczenia z wysiłkiem słuchowym związanym ze zrozumieniem mowy prezentowanej w warunkach określonych przez ustawienia RES. Biorąc pod uwagę, że:

- w teście logatomowym bez znaczenia jest zasób leksykalny badanego (Giovannone, 2021),
- maksymalnie wyeliminowana jest zależność zrozumiałości mowy od kontekstu wypowiedzi,
- cały pomiar trwał mniej niż 20 minut, minimalizując skutki zmęczenia (zgodnie ze wskazaniami Brachmańskiego i Dobruckiego, 2021),
- wszyscy badani byli w podobnym wieku (średnia 23.5 lat) i mieli prawidłowy słuch,

przyjęto założenie, że głównym czynnikiem wpływającym na stopień zrozumiałość mowy są cechy ścieżki propagacji fali akustycznej (akustyka pomieszczenia) z potencjalnym udziałem innych, pozasłuchowych czynników percepcyjnych. Jak się okazało, wysiłek słuchowy silniej koreluje ze wskaźnikiem transmisji mowy niż z RT. ($R^2 = 0.74$; Rys 51).



Rys. 51: *Powiązanie STI ze średnim wysiłkiem słuchowym (1–7)*

Ze względu na zastosowanie stosunkowo niewielkiej liczby ustawień RES (łącznie 9 różnych RT), nie jest uzasadnione wyciąganie kategorycznych wniosków. Jednym z powodów tej obserwacji może być fakt, że STI jest bardziej bezpośrednią miarą zrozumiałości, która uwzględnia szerszy zakres czynników wpływających na rozumienie sygnału mowy, takich jak zniekształcenia czy modulację sygnału, co zapewnia bardziej kompleksową ocenę w porównaniu do samego czasu utrzymywania się dźwięku w pomieszczeniu po wyłączeniu źródła. Można przypuszczać, że słuchacze subiektywnie oceniają swój wysiłek słuchania w oparciu o ogólną zrozumiałość mowy, metrykę lepiej reprezentowaną przez STI niż przez RT. Wprawdzie, jak wspomniano zaobserwowano tendencję, zgodnie z którą średnie wartości oceny wysiłku słuchowego są większe przy wyższych RT, a zmniejszonej zrozumiałości logatomów towarzyszyła wyższa ocena wysiłku słuchowego. Nie stwierdzono jednak istotności statystycznej ($p > 0.05$). Jest wysoce prawdopodobne, że badani przecenili swoje możliwości poznawcze pomimo niższych rzeczywistych poziomów zrozumiałości. Tendencja ta jest szczególnie wyraźna w przypadku używania słów nonsensownych, gdzie słuchaczom trudniej jest ocenić dokładność napisanych słów w porównaniu z semantycznie znaczącym materiałem językowym. Innym potencjalnym wyjaśnieniem tego braku istotności, a także ograniczeniem badania, jest to, że 7-punktowa skala zastosowana do pomiaru wysiłku słuchowego może nie być wystarczająco czuła, aby wykryć różnice w całym zakresie wartości RT. W przypadku trzech najkrótszych RT deklarowany wysiłek spada nieco poniżej wyniku 3 (Rys 50.), podczas gdy dla wszystkich innych czasów nieznacznie

przekracza 3 w siedmiostopniowej skali. Sugeruje to, że wysiłek wymagany do zrozumienia logatomów pozostaje względnie spójny przy różnych RT.

E10.3 Wnioski z eksperymentu E10

Wysiłek słuchowy może być krytyczny w ogólnej ocenie akustyki pomieszczenia, podobnie jak subiektywne preferencje użytkowników aparatów słuchowych, które często determinują ich użycie użycie pomimo obiektywnie zmierzonych i dostosowanych parametrów wskazujących na poprawę „wydajności" słuchowej (E5). Przestrzeń może mieć dobrą zrozumiałość i parametry akustyczne wskazujące na wysoką jakość, ale nadal może być postrzegana słabo ze względu na duży wysiłek wymagany od słuchaczy. Chociaż w przytoczonym badaniu analizowane były tylko podstawowe dane, wyniki jasno dowodzą, że obiektywne parametry, których użycie jest standardem w akustyce wnętrz, nie są w stanie oszacować komfortu słuchacza w trakcie próby zrozumienia tekstu mówionego, a ten ma niebagatelne znaczenie w całościowym doświadczeniu słuchowym (lub audiowizualnym). Wprawdzie pierwsze próby uwzględnienia subiektywnych odczuć słuchaczy, wykraczające poza fizyczne pomiary i modele, zostały podjęte przez Beranka (2004) i Ando (2007), jednak nadal nie jest to podejście standardowe. Uzasadnione wydaje się być zatem powszechne uwzględnienie wskaźników opisujących wysiłek poznawczy w charakterystyce akustyki pomieszczenia, szczególnie w przypadku prezentacji sygnałów mowy. Oczywiście wymaga to wcześniejszych rozszerzonych (np. o inny rodzaj materiału testowego) badań.

E11 Zrozumiałość mowy u osób w fazie prodromalnej (bezobjawowej) choroby Alzheimera – wstępne wyniki z projektu *Alzheimer Prediction Project*

Jednym z czynników wpływających na przebieg i skuteczność percepcji słuchowej w zakresie zrozumiałości mowy, które zostały wymienione w rozdziale 4. części teoretycznej rozprawy jest kompensacja poznawcza. Rozszerzyć ją można do bardziej ogólnego procesu, a więc zdolności prawidłowego przetwarzania pochodzących ze środowiska informacji, potocznie określanej mianem sprawności umysłowej. Istnieje wiele zmiennych warunkujących, jak długo zostanie ona utrzymana w cyklu życiowym jednostki – aspekty genetyczne, dieta, aktywność fizyczna oraz, co podkreślane coraz częściej, uczenie się, doświadczanie i nabywanie nowych umiejętności. Profesor Vetulani powtarzał wielokrotnie w humorystyczny sposób, że „w mózgu nie ma zasiłku dla bezrobotnych". Układ nerwowy musi być bodźcowany, aby w procesie odbioru pobudzeń, ich interpretacji oraz reagowania nań, stymulować określone ośrodki mózgu. Badania wykazują, że umysłowa i fizyczna aktywność mogą redukować ryzyko demencji (m.in. Verghese i in., 2003) – zespołu otępiennego objawiającego się utratą funkcji poznawczych. W przypadku pierwszej z nich jest to nawet o 26% mniejsza podatność na neurodegeneracyjne choroby mózgu (Carlson i wsp., 2015). Tego rodzaju doniesienia są bardzo istotne, zwłaszcza, że demencja staje się chorobą cywilizacyjną. WHO szacuje, że dotyka ona ponad 55 milionów osób, a liczba ta wzrośnie do 78 milionów w 2030 roku i do 139 milionów w 2050 roku (WHO, 2021).

Pod pojęciem demencji kryje się wachlarz schorzeń. Najczęstszym z nich jest choroba Alzheimera (ang. *Alzheimer's disease*; AD), stanowiąca 50–60% przypadków demencji[9]. Dokładna etiologia nie została jednoznacznie zdefiniowana, ale przypuszcza się, że jedną z przyczyn neurodegeneracji oraz AD są procesy mikrozapalne. Kiedy z nieprawidłowego białka zwanego beta-amyloidem tworzy się tzw. płytka starcza, osadzają się na niej komórki mikrogleju generujące tlenek azotu powodując powstanie odczynu zapalnego (Vetulani, 2007). Do głównych objawów AD należą utrata pamięci, trudności w znalezieniu właściwych słów lub rozumieniu mowy, a także zmiany osobowości i nastroju Chociaż dynamika rozwoju choroby jest kwestią indywidualną, demencja jest główną przyczyną niepełnosprawności oraz braku samodzielności osób starszych, często prowadzącą do przedwczesnej śmierci.

W obliczu przedstawionych informacji szczególnie istotne i w pełni uzasadnione jest prowadzenie dalszych badań dotyczących demencji. Im większa wiedza

[9]źr.: https://www.alzint.org/about/, [dostęp: 13.10.2024 r.]

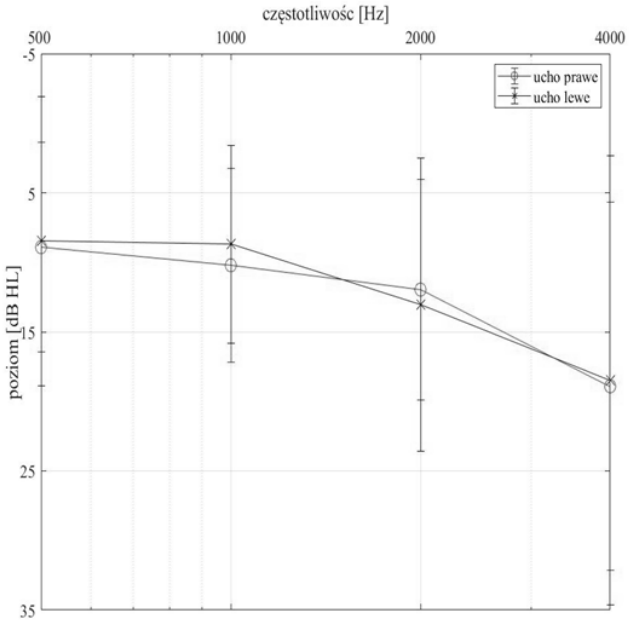
na temat przyczyn występowania dysfunkcji mózgu oraz ich przebiegu, tym większa możliwości coraz bardziej skutecznej prewencji, a w przyszłości, być może, opracowania metod ich leczenia. Postawione na wczesnym etapie choroby rozpoznanie ma fundamentalne znaczenie dla rokowania. Obecne metody diagnostyczne są jednak bardzo kosztowne, inwazyjne i trudno dostępne.
Wychodząc naprzeciw tym potrzebom Uniwersytet im. Adama Mickiewicza w Poznaniu oraz Politechnika Poznańska prowadzą projekt badawczy Alzheimer Prediction Project (APP), którego celem jest opracowanie nowej, bezinwazyjnej metody diagnozowania choroby Alzheimera na jej bardzo wczesnym etapie. Wykorzystanie obiektywnych miar psychofizycznych miar, których wdrożenie wiąże się z dużo mniejszymi kosztami niż środki konwencjonalne, potencjalnie pozwoliłoby wstępnie zidentyfikować osoby wymagające dalszej specjalistycznej diagnostyki. W przypadku osób już zdiagnozowanych, obserwacja przebiegu choroby, śledzenia jej dynamiki, a także reakcji na wdrożone leczenie i procedury medyczne mogłyby natomiast znacznie ułatwić skuteczne zarządzanie objawami choroby, prowadzić bardziej efektywną rehabilitację, poprawiając komfort życia pacjentów i ich bliskich. W literaturze pojawiły się doniesienia na temat zmian w percepcji wzrokowej i słuchowej, które towarzyszą demencji, w tym chorobie Alzheimera. Stąd kluczowa jest analiza postrzegania bodźców za pomocą właśnie tych dwóch modalności, co ma miejsce w przypadku opisywanego projektu badawczego. Z uwagi na zakres niniejszej rozprawy, przytoczone zostaną jedynie wstępne wyniki dotyczące zrozumiałości mowy. Dokładny opis protokołu eksperymentalnego można znaleźć na stronie internetowej projektu Alzheimer Prediction Project[10].

E11.1 Osoby badane

W badaniach wzięło udział 30 osób, z których wybrano 22 ze słuchem prawidłowym zgodnie z wytycznymi WHO (średnio po 12.6 dB HL w uchu prawym i lewym), tak, aby uniknąć bezpośredniego wpływu ewentualnego niedosłuchu na stopień zrozumiałości mowy. Uśrednione audiogramy dla ucha prawego (kółka) i lewego (krzyżyki) dla czterech z ośmiu badanych częstotliwości przedstawiono na Rys. 52.

Wszyscy słuchacze (w wieku od 40 do 75 roku życia; średnia 59.4) byli w prodromalnej (bezobjawowej) fazie choroby Alzheimera potwierdzonej konwencjonalną diagnostyką szpitalną lub z jej podejrzeniem (na podstawie wywiadu rodzinnego) i zostali skierowani do udziału w badaniach przez psychiatrę lub neurologa biorących udział w projekcie.

[10]Strona internetowa APP: https://alz.put.poznan.pl/, [dostęp: 13.10.2024 r.]



Rys. 52: *Uśrednione progi słyszenia biorących udział w badaniu (N=22)*

Wyznaczając średnie progi słyszenia bierze się pod uwagę wyłącznie 4 częstotliwości (500, 1000, 2000, 4000 Hz) bez zachowania informacji o przebiegu audiogramu. Zdarza się zatem, że u pacjentów stwierdza się znaczne podwyższenie progów dla najwyższych częstotliwości, jednak z uwagi na to, że detekcja tonów o częstotliwości 500 czy 1000 Hz odbywa się przy poziomie 0–10 dB HL, obliczona średnia mieści się w granicach normy. Chcąc uzyskać dodatkowe informacje o słyszeniu uczestników badania, zmierzone zostały progi dla szerszego zakresu częstotliwości (125–8000 Hz), a otrzymane audiogramy uszeregowano zgodnie z klasyfikacją Bisgaarda (opisaną szczegółowo w rodziale E10 części eksperymentalnej rozprawy). Poza jedną badaną osobą, u wszystkich przebiegi progów słyszenia określono jako typu N1. Mimo prawidłowej detekcji tonów zaledwie u 5 słuchaczy zarejestrowano otoemisję typu DPOAE dla wszystkich częstotliwości w przypadku pozostałej części nie była ona obecna obustronnie dla częstotliwości 8kHz, a w niektórych przypadkach również 6 kHz. Nie jest to sytuacja niecodzienna, co jest związane z większą czułością badania DPOAE. Wiele doniesień literaturowych wskazuje na to, że wprawdzie nie jest to metoda służąca wyznaczeniu progów słyszenia, jednak stanowi narzędzie dużo bardziej czułe niż audiometria tonalna. Rozproszone uszkodzenie nawet 30% komórek słuchowych zewnętrznych może nie znajdować odzwierciedlenia w badaniu z wykorzystaniem tonów (Bartnik i in., 2002).

E11.2 Materiał i metody

Jednym z badań, poza pomiarami psychoakustycznymi (m.in. wyznaczanie

progów detekcji modulacji amplitudowej oraz BMLD [11] z zastosowaniem pakietu Psychoacoustics [12]) oraz diagnostyką obiektywną i elektrofizjologiczną (audiometria impedancyjna, otoemisja typu DPOAE w zakresie 1–8 kHz oraz rejestracja potencjałów wywołanych z pnia mózgu ABR), była ocena zrozumiałości mowy przy wykorzystaniu polskiego testu zdaniowego (opisanego w podrozdziale 6.7.1.). Poziom maskującego szumu mowopodobnego utworzonego dla języka polskiego z maksimum w okolicy 6–7 kHz, wynikającego z występowania dużej liczby głosek zwarto–trących i trących (tzw. babble noise) utrzymywany był na stałym poziomie 65 dB SPL, natomiast poziom mowy zmieniał się w sposób adaptacyjny (metoda adaptacyjna została opisana w podrozdziale 5.2.5.). W badaniu wykorzystano słuchawki Sennheiser HD600 w układzie z przetwornikiem cyfrowo–analogowym Tucker-Davis Technology (TDT) RP2.
Każda lista pomiarowa składała się z 25 zdań. Pierwsze trzy pomiary (w tym jeden próbny) odbywały się w konfiguracji S0N0, w której sygnał mowy i szum docierały do słuchacza z tego samego kierunku. W kolejnych dwóch kierunek sygnału zakłócającego zmieniał się o 90° w prawo – dochodziło do przestrzennego rozseparowania mowy i maskera (S0N90), co zilustrowano na Rys. 12 w podrozdziale 4.6 dotyczącym słyszenia binauralnego, z zastrzeżeniem, że, jak wspomniano, pomiar odbywał się z wykorzystaniem słuchawek. Zadaniem badanego było powtórzenie usłyszanego materiału słownego.

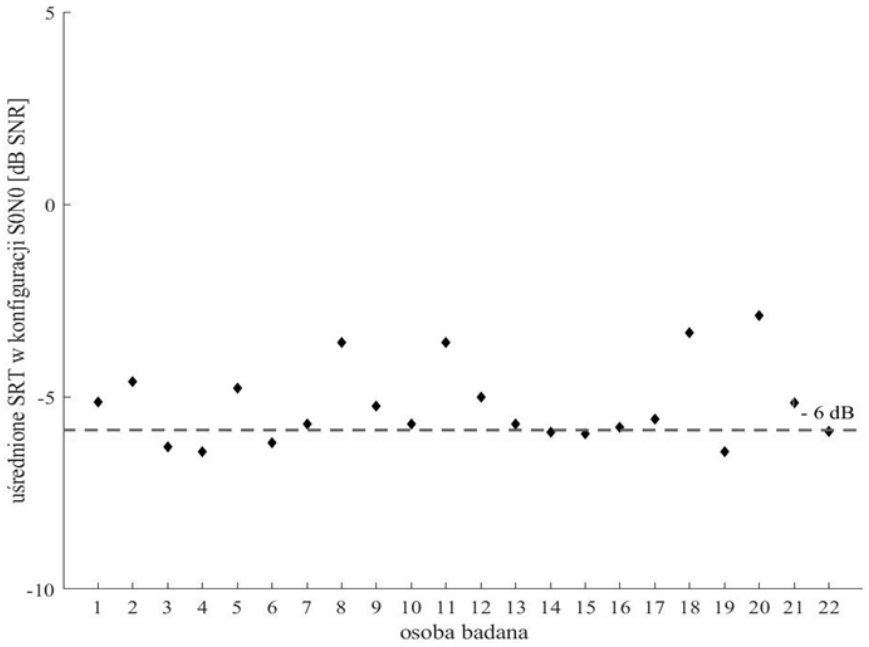
E11.3 Wyniki i ich dyskusja

Zmierzone wartości SRT dla konfiguracji S0N0 mieszą się w granicach wartości typowej dla grupy referencyjnej wynoszącej -6dB (średnia -5.2, mediana -5.6). W niektórych przypadkach odstają one od chmury danych, jednak odchylenie standardowe oszacowano na zaledwie 1 dB.

W przypadku pomiarów SRT w konfiguracji S0N90 (rozseparowanie przestrzenne maskera i mowy) obserwowane jest większe rozproszenie wyników – odchylenie standardowe wynosi blisko 3 dB. Wyznaczona wartość średnia osiągnęła -10.8, przy

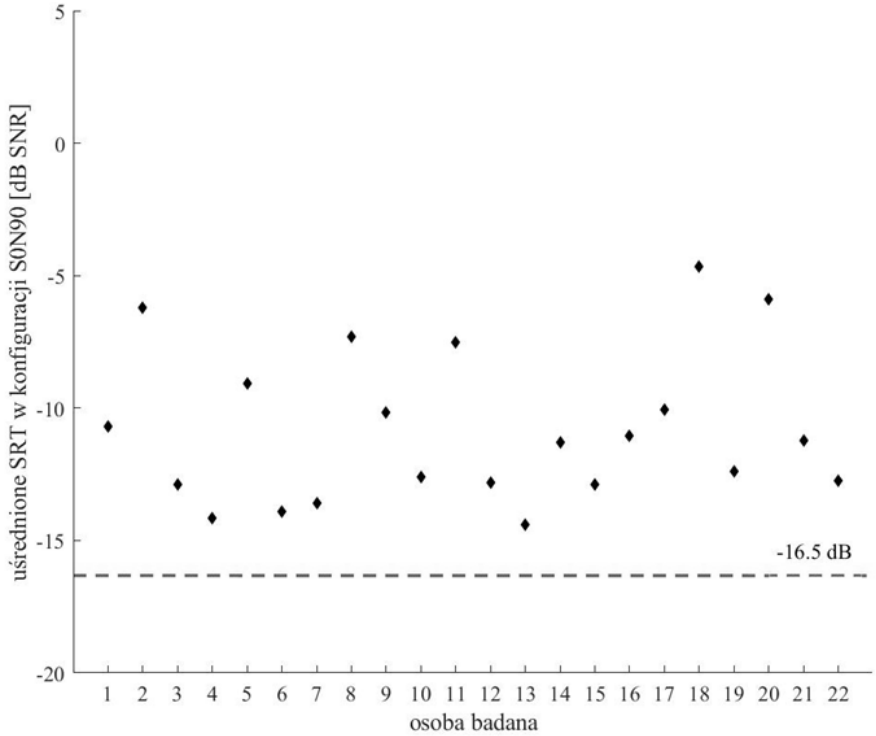
[11] Badanie dotyczące BMLD (ang. *Binaural Masking Level Difference*) polega na wyznaczeniu minimalnego słyszalnego poziomu tonu przedstawionego na tle szumu (poziom odmaskowania sygnału). Porównuje się warunki, w którym tony prezentowane są do obojga uszu z tą samą fazą (S0N0) oraz gdy do jednego z uszu podawany jest sygnał tonalny przesunięty w fazie o π (SπN0). BMLD u prawidłowo słyszących osób dla parametrów użytych w opisywanym projekcie badawczym wynosi około 8 dB (np. Ignatious i wsp., 2023).

[12] Pakiet aplikacji na PC, który umożliwia konfigurowanie i prowadzenie szerokiej gamy eksperymentów psychoakustycznych przez prof. A. Sęka (UAM) oraz B.C.J. Moore'a (University of Cambridge); Sęk i Moore (2021), Guide to PSYCHOACOUSTICS, Adam Mickiewicz University Press.



Rys. 53: *Średnie SRT polskiego testu zdaniowego (PTZ) w konfiguracji S0N0*

medianie -11.26 dB. Jest ona znacznie niższa niż referencyjna określona (dla tego samego materiału językowego) na -16.5 dB (Ozimek i in., 2013).



Rys. 54: *Średnie SRT polskiego testu zdaniowego (PTZ) w konfiguracji S0N90*

W badanej grupie, średnia różnica między S0N0 a S0N90 wynosiła -5.57 dB (odch. standardowe 2.02, mediana -5.78), a więc blisko dwa razy mniej niż

w grupie referencyjnej przedstawionej np. cytowanej już w publikacji Ozimka (2013). Wyniki otrzymane w konfiguracji S0N0 są ściśle skorelowane z SRT w S0N90, o czym świadczy wysoka wartość współczynnika Pearsona wynosząca 0.78. Słuchacze, którzy osiągali stosunkowo wysokie progi zrozumiałości mowy w S0N90 (słaba zrozumiałość mowy) nie byli w stanie wykorzystać percepcyjnie separacji przestrzennej sygnału mowy i szumu maskującego, co skutkowało relatywnie podwyższonymi wartościami progowymi.

Oczywiście, fakt iż obserwowana korzyść jest dwa razy mniejsza niż w literaturze może być związana z wiekiem uczestników opisywanego badania, jednak nie bez wpływu pozostają tutaj również zmiany o podłożu poznawczym. Należy mieć na uwadze, że odchylenie uzyskanych wyników od wartości referencyjnych może wynikać z innych, współistniejących stworzeń lub naturalnego procesu starzenia się (choroba Alzheimera jest najbardziej powszechna w populacji powyżej 65 r.ż.), w których ubytek słuchu nie zawsze musi manifestować się podwyższeniem progu słyszenia. Najbardziej prawdopodobne scenariusze tłumaczące otrzymane wyniki sprowadzają się do trzech zasadniczych aspektów:

- **Ograniczonej wydajności procesów przetwarzania informacji dźwiękowej, zwłaszcza bardziej złożonej (np. nacechowanej semantycznie).** Jak wykazały badania Strouse'a (1995) czy Iliadou (2003), wartości określające m.in. skuteczność dyskryminacji mowy czy barwy dźwięku są u osób z demencją znacznie niższe niż u grupy kontrolnej.

- **Deficytów w przetwarzaniu binauralnym.** Badania Li i zespołu (2023) wykorzystujące testy mowy przerzucanej (zasada ich tworzenia opisana w podrozdziale 6.8.) potwierdzają, że obserwuje się zaburzenia w łączeniu informacji prezentowanej w układzie dychotycznym nie tylko wśród osób z AD, ale również przy niewielkich deficytach (ang. *mild cognitive impairment*). Co więcej, adaptacja neuronalna do deficytu przetwarzania obuusznego może nasilać zaburzenia poznawcze (Wang i wsp., 2022), stąd kompensacja niedosłuchu, gdy występuje, lub rehabilitacja słuchowa mają tak istotne znaczenie.

- **Ograniczeń pamięci roboczej.** Jak wspomniano, separacja przestrzenna sygnałów opiera się nie tylko na podstawowym przetwarzaniu informacji akustycznej, ale także na zdolnościach poznawczych, w które niezwykle silnie zaangażowane są uwaga i pamięć robocza. Wyniki badań Zokaei i Husaina (2019) sugerują, że błędy poznawcze wynikają w dużej mierze z charakterystycznego dla AD nieprawidłowego powiązania otrzymanej

informacji, z wzorcem przechowywanym w pamięci. Są to wnioski zgodne z koncepcjami neuroanatomicznymi, które wskazują na istotną rolę przyśrodkowych płatów skroniowych, zwłaszcza hipokampa (ograniczenie czynności którego jest kluczową oznaką AD) w utrzymywaniu wiązań cech, niezależnie od tego, jak długo trwa przechowywania informacji.

E11.4 Wnioski z eksperymentu E11

Jak wynika z przedstawionych danych, czynników i zmiennych jest wiele. Często nakładają się nań inne, wynikające z wieku lub chorób przewlekłych. Z tego względu utrzymane rezultaty, również te wynikające z badań współprowadzonych przez autorkę niniejszej rozprawy (i opisywanych w tym rozdziale) należy, zgodnie z zasadą *cross-check*, zestawić z danymi pochodzącymi z innych pomiarów, w szczególności takich, które nie wymagają bezpośredniego zaangażowania osoby badanej w procedurę pomiarową (badania obiektywne). Dużą wartością byłoby również śledzenie, jak efektywność słyszenia mierzona stopniem zrozumiałości mowy (wyrażona poprzez SRT) zmienia się wraz z upływem czasu.
W tym aspekcie projekt APP może mieć niebagatelne znaczenie w rozszerzaniu wiedzy na temat zmian efektywności percepcji w przebiegu choroby Alzheimera, a, potencjalnie, może pomóc w stworzeniu narzędzia przesiewowego oraz służącego do monitorowania przebiegu AD u pacjentów już zdiagnozowanych. Zarówno w badaniach dotyczących zrozumiałości mowy w salach ze zmieniającym się czasem pogłosu (E10), jak i wyżej przytoczonych pomiarach psychoakustycznych z udziałem osób w prodromalnej fazie choroby Alzheimera (E11) nie wykorzystano polskiego testu zdaniowego matrix, którego, w zasadniczej części, dotyczy niniejsza rozprawa. Obecność krótkich opisów przyjętej metodologii i wyników uzyskanych w tych eksperymentach jest jednak w pełni uzasadniona. Ich treść stanowi bowiem dodatkowe wzmocnienie stanowiska klarowanego przez całą zawartość pracy a przemawiającego za tym, że ocena wydajności percepcji słuchowej nie może ograniczać się do powszechnie przyjmowanych standardów audiologicznych. Te zakładają często realizację jedynie podstawowej oceny słuchu, która, w skrajnych przypadkach, oznacza audiometrię tonalną, wzbogaconą niekiedy prostym testem wykorzystującym mowę przeprowadzanym w ciszy. Konwencjonalna diagnostyka skupia się zatem w zasadzie wyłącznie na czynności peryferyjnej części układu słuchowego.
Na percepcję składają się jednak, co wielokrotnie podkreślano w części eksperymentalnej rozprawy, aspekty uwzględniające zaangażowanie wyższych pięter

drogi słuchowej, czy nawet pozasłuchowe.

Po pierwsze, wymienić można te związane z szeroko rozumianymi warunkami propagacji sygnału. Właśnie z tego powodu w badaniach opisanych w części eksperymentalnej wykorzystywano różnego rodzaju szumy maskujące (mające charakter zakłócający). Zasygnalizowano również wpływ akustyki pomieszczenia, w którym mowa jest prezentowana (eksperyment E10).

Po drugie – cechy osobnicze. Z jednej strony indywidualny leksykon, kontekst i uwaga (stąd wybór testu zdaniowego, a nie np. sylab czy wyłącznie logatomów), z drugiej zaburzenia bardziej precyzyjnego przetwarzania informacji akustycznej (rozdzielczość czasowa, częstotliwościowa – *masking release*; E4) czy nawet nieprawidłowości w obrębie ośrodkowego układu nerwowego (próba wykorzystania testów zrozumiałości mowy w ocenie przebiegu choroby neurodegeneracyjnej; E11).

Choć wyniki opisywanych eskperymentów nie wyczerpują tematu w pełni, a zdarza się, że są niejednoznaczne, zebrane w pomiarach E1–E8 doświadczenie pokazuje szerszy kontekst badań wykorzystujących sygnał mowy. Przedłużeniem tego założenia jest przedstawienie potencjalnych, dalszych zastosowań materiału językowego w ocenie wieloaspektowej percepcji słuchowej i towarzyszącego jej wysiłku poznawczego (E10 i E11). Dopiero gdy gruntowana, budowana na przestrzeni wielu dekad wiedza kliniczna zostanie połączona z nowatorskim podejściem, często nieoczywistym i zakrawającym o inne niż wyłącznie akustyka i audiologia dyscypliny, możliwe stanie się zrozumienie mechanizmów rządzących ludzką percepcją (zwłaszcza związanych z najważniejszym dla człowieka sygnałem – mową), opracowanie bardziej precyzyjnych metod służących do oceny ilościowej i jakościowej, a także rozwiązań sprzyjających kompensacji jej zaburzeń.

Podsumowanie

Jak wykazano na podstawie przytoczonej literatury przedmiotu, a także wyników badań przeprowadzonych przez autorkę niniejszej rozprawy, układ słuchowy jest zdolny do przetwarzania mowy nawet w bardzo niesprzyjających warunkach – w obecności hałasu (Vanthornhout i in., 2018) lub kiedy sam sygnał mowy jest zniekształcony (Smith i in., 2002) czy przyspieszony (Meng i wsp., 2019). Efektywność percepcji jest kształtowana nie tylko przez wydolność części peryferyjnej, ale także tzw. wyższych pięter drogi słuchowej, na odpowiednich ośrodkach korowych mózgu kończąc (Versfeld i Dreschler, 2002). Należy pamiętać, że problemy ze zrozumieniem mowy nie ograniczają się do obecności ubytków słuchu wyrażających się podwyższonymi progami słyszenia wyznaczonymi w audiometrii tonalnej. Słuchaczy bez stwierdzonego niedosłuchu również dotyczy stosunkowo duża zmienność międzyosobnicza, która może wynikać z różnic w skuteczności analizy i interpretacji pobudzenia akustycznego.

W rozprawie, poza wynikami badań ściśle dotyczących tematu, jakim jest zrozumiałość mowy i ocena wysiłku słuchowego z wykorzystaniem testu typu matrix, umieszczono również wnioski z pomiarów opartych na innym materiale językowym. Punktem wspólnym pozostaje jednak zrozumiałość mowy i wskazanie czynników, które mogą mieć nań wpływ. Uwzględnienie tych wyników w niniejszej pracy wydaje się więc nie tylko uzasadnione, ale także istotne, gdyż jednocześnie pokazuje dalsze kierunki możliwych badań z wykorzystaniem testów zrozumiałości mowy.

Oczywiście, testy wykorzystujące mowę, nawet poddane skrupulatnej walidacji, nie są pozbawione wad. Najdokładniej przygotowane listy odsłuchowe nie są w stanie w pełni odzwierciedlić rzeczywistych warunków komunikacyjnych, w tym stopnia skomplikowania sygnału mowy. Nie mogą również oddać dynamiki i złożoności interakcji osób mówiących i słuchaczy, w której niebagatelną rolę odgrywają wskazówki pozaakustyczne, często związane z innymi niż słuch modalnościami. Testy wykorzystujące mowę pozwalają jednak uzyskać dużo więcej informacji niż audiometria tonalna, która nadal bywa jedynym badaniem przeprowadzanym

w gabinetach protetyki słuchu. Są narzędziami niezwykle czułymi, pozwalającymi określić wpływ nawet niewielkich zmian poziomu sygnału mowy (lub SNR) na jej zrozumiałość. Wiele z nich jest prostych w użyciu, nawet w warunkach pozaklinicznych, a pomiar z ich zastosowaniem nie zajmuje dużo czasu. Testy oparte o materiał językowy umożliwiają utrzymanie kontrolowanych i odtwarzalnych warunków badania, co również zwiększa ich wiarygodność.

Z niniejszej rozprawy płynie również kilka praktycznych wniosków, które mogą stanowić propozycję „dobrych praktyk" dla protetyków słuchu oraz innych specjalistów zaangażowanych w tematykę percepcji mowy, oceny jej wydajności, a także poszukiwania metod ją wspierających.

1. Z uwagi na możliwość wykorzystania wskazówek pochodzących z innych modalności, np. wzrokowych (co zostało szerzej opisane w podrozdziale 3.1.) przy ocenie wydajności percepcyjnej w zakresie zrozumiałości mowy, w części diagnostycznej warto uniemożliwić pacjentowi czytanie z ruchu warg, np. poprzez zasłanianie ust ręką lub unikanie kierowania twarzy „na wprost" badanego. Z drugiej strony, w momencie pierwszego kontaktu lub w trakcie tłumaczenia procedury diagnostycznej pacjentowi z niedosłuchem (a więc w momentach dla niego ważnych i stresujących), zapewnienie pacjentowi możliwości obserwacji ruchu warg jest wysoce pożądane.

2. Badania z zakresu audiometrii mowy powinny być wykonywane w obecności sygnałów maskujących. Nawet w ogólnej, przesiewowej ocenie, w której zamiast zwalidowanych narzędzi używa się wyłącznie własnego głosu (np. prowadząc z pacjentem rozmowę w trakcie zmieniania ustawień aparatu słuchowego, by zdefiniować te, które będą najbardziej komfortowe) warto skorzystać z naturalnych metod zakłócania sygnału mowy np. przez szum wody z kranu, szuranie butami po podłodze. Dzięki temu możliwa będzie ocena faktycznej wydajności układu słyszenia, jedynie w minimalnym stopniu wspomaganej kontekstem czy innymi aspektami czysto poznawczymi (kojarzeniowymi).

3. Należy pamiętać, że pacjenci funkcjonują w różnych środowiskach. W gabinetach protetyki słuchu lub innych pomieszczeniach, w których prowadzi się diagnostykę słuchu, z reguły panują warunki ograniczonego pogłosu i ciszy. Badania, jak chociażby to przytoczone w niniejszej tezie (Błasiński, 2023), wskazują jednak, że obecność wydłużonego ma niebagatelny

wpływ na zrozumiałość mowy oraz wysiłek słuchowy (a więc pośrednio – komfort pacjenta).

4. Audiometria mowy jest świetną metodą uzupełniającą diagnostykę obiektywną (np. ABR – ocenę potencjałów wywołanych z pnia mózgu), pozwalającą na lokalizację niedosłuchu. Korzystając z różnego rodzaju narzędzi (testów o nacechowaniu semantycznym bądź nie) i warunków prezentacji sygnału testowego (brak szumu, szum stacjonarny lub *speech-on-speech*), możliwa jest ocena, czy zgłaszany przez pacjenta problem ze zrozumieniem mowy wynika wyłącznie z ograniczeń peryferyjnych (wywołanych np. przewodzeniowym ubytkiem słuchu), czy raczej pobudzenie akustyczne odbierane jest właściwe, ale na dalszych etapach jego analizy lub interpretacji dochodzi do zaburzenia właściwej percepcji (jak np. w CAPD).

5. Audiometria mowy to nie tylko bezpośrednia metoda oceny wydajności układu słuchowego, ale, pośrednio, potencjalny marker szerzej rozumianych zmian neurodegeneracyjncyh. W jednym z rozdziałów pracy opisany jest projekt dotyczący opracowania narzędzia pozwalającego na ocenę i monitorowanie pacjentów z demencją, w tym prodromalnej fazie choroby Alzheimera. Testy wykorzystujące mowę wydają się być jego fundamentem.

6. Zrozumiałość mowy jest podstawową składową efektywności percepcji słuchowej. Warto mieć na uwadze również osobnicze warunki, które jej towarzyszą. Dobrym przykładem może być wysiłek słuchowy. Przy dopasowaniu aparatów słuchowych możliwe jest uzyskanie identycznego stopnia zrozumiałości mowy dla dwóch ustawień. Wybór najbardziej optymalnego może być w takiej sytuacji podyktowany wielkością wysiłku słuchowego.

7. Zrozumiałość mowy i ocena wysiłku słuchowego to miary przydatne nie tylko w audiologii, ale i akustyce wnętrz. Uzupełnienie powszechnie stosowanych fizycznych (obiektywnych) miar takich jak STI czy C50 o ocenę subiektywną (*listening effort*) i faktycznie zmierzoną (a nie jedynie wyliczoną lub modelowaną) zrozumiałość mowy może być rozwiązaniem o ogromnym potencjale do wykorzystania przy projektowaniu akustyki pomieszczeń użytkowych, czy zmianie ustawień obsługujących ich systemów, zwłaszcza w dobie rosnącego zainteresowania salami wielofunkcyjnymi.

8. Myśląc o mowie i jej rozumieniu nie można zapomnieć o najważniejszej roli, jaką pełni protetyka słuchu. Choć zasadniczym jej celem jest kompensowanie

niedosłuchu w największym możliwym stopniu, tak naprawdę chodzi o zapewnienie pacjentom wysokiej jakości życia. Z tego względu obowiązkowe jest zachowanie szczególnej troski i dbałości o wykonywane procedury diagnostyczne. Konieczne jest korzystanie ze zwalidowanych narzędzi o zweryfikowanej skuteczności i wysokim współczynniku *test-retest*. Zgodnie z zasadą *cross-check*, jakiekolwiek działania pomiarowe winny być uzupełniane innymi (np. testami obiektywnymi), a otrzymane eksperymentalnie wyniki konfrontowane z subiektywnymi odczuciami badanych (zgłaszanym wysiłkiem słuchowym czy komfortem).

Dalsze kierunki badań

W świetle przeprowadzonych i opisanych eksperymentów warto podkreślić potencjał dalszej eksploracji omawianych zagadnień, bazując na wiedzy i obserwacjach poczynionych przez autorkę niniejszej rozprawy oraz współbadaczy.
Przykładem mogą tu być badania realizowane przez autorkę we współpracy naukowej z Łukaszem Błasińskim i prof. Jędrzejem Kocińskim w zakresie powiązania obiektywnych miar akustyki wnętrz z nie tylko zrozumiałością mowy, ale także, co nadal nieoczywiste – wysiłkiem słuchowym. A to właśnie on może mieć krytyczne znaczenie w ogólnej ocenie pomieszczenia, analogicznie jak, w dyscyplinie nieco bliższej autorce, subiektywne preferencje użytkowników aparatów słuchowych i innych urządzeń wspomagających słyszenie często determinują ich używanie. Sala może być świetnie zaprojektowana pod kątem pożądanych wartości wskaźników obiektywnych, niemniej ostateczna ocena należy do słuchaczy, co stanowi ciekawe, multidyscyplinarne wyzwanie, w ramach którego poszukiwana jest odpowiedź na pytanie m.in. o metody walidacji akustyki pomieszczeń.
Innym obszarem wymagającym dalszych badań są działania grupy skupionej wokół Alzheimer Prediction Project. To zupełnie nowe podejście wspomagające wczesną diagnostykę i monitorowanie przebiegu chorób neurodegeneracyjnych, którego niewątpliwą zaletą jest połączenie wiedzy i doświadczenia ekspertów-praktyków z wielu dziedzin. Pierwsze wyniki badań pilotażowych są obiecujące i wskazują na dalsze możliwe kierunki rozwoju.
Warto również zwrócić uwagę na kwestie poruszone w rozprawie w części teoretycznej a dotyczące wielomodalnego odbioru mowy. Pandemia COVID-19 i związane z nią noszenie maseczek znacząco ograniczyło dostęp do informacji wizualnych, co uwypukliło rolę wzroku w procesie rozumienia mowy. W tym kontekście interesujące wydaje się być zweryfikowanie, w jakim stopniu zakrywanie

twarzy wpływa na zrozumiałość mowy, zarówno u osób prawidłowo słyszących (młodszych i starszych), jak i z niedosłuchem, w różnych warunkach akustycznych (np. w ciszy i w szumie). Z tego względu naturalną kontynuacją badań weryfikujących skuteczność polskiego testu zdaniowego matrix (oraz innych narzędzi w języku polskim) oraz obszarów, w których mógłby być zastosowany wydaje się być również krok poczyniony przez zespoły pracujące nad jego odpowiednikami dla innych języków, a więc testy audio-wizualne. Opracowanie, zweryfikowanie i kliniczne zwalidowanie takiego narzędzia byłoby niewątpliwie olbrzymim osiągnięciem z zakresu akustyki mowy polskiej i mogłoby otworzyć nowe możliwości zarówno w diagnostyce i rehabilitacji słuchu, akustyce wnętrz (projektowanie i ocena empiryczna), jak i szeroko rozumianej psychoakustyce oraz roli modalności wzrokowej w percepcji słuchowej, ze szczególnym uwzględnieniem mowy, która, jak wielokrotnie wspomniano, jest z punktu widzenia ewolucji i funkcjonowania człowieka w środowisku sygnałem najważniejszym.

Spis rysunków

1 *„Banan mowy" dla języka angielskiego* (źr.: Northern i Downs, 2014) 3

2 *Schematyczne przedstawienie tonotopowej struktury ślimaka* (źr.: Pruszewicz, 2010) 4

3 *Budowa narządu Cortiego* (oprac. własne na podst.: Goldstein, 2013) 5

4 *Zależność poziomu wejściowego (oś y) potrzebnego do wytworzenia jednakowej prędkości drgań błony podstawnej w zależności od częstotliwości stymulacji (oś x). Krzywa oznaczona kółkami uzyskana przy prawidłowym stanie fizjologicznym świnki morskiej, krzywa strojenia oznaczona czarnymi kwadratami pochodzi z pomiarów po śmierci zwierzęcia – post-mortem*(źr.: Selick i in. 1982) 6

5 *Schematycznie przedstawiony przebieg F1 i F2 zgłosek „ba" i „da"* 7

6 *Przykładowe spektrogramy nagrań zawierających nieprawidłowe dźwięki oddechowe – świsty wydechowe (panel górny) oraz rzężenia wdechowe (panel dolny)* 13

7 *Spektrogram graficznie reprezentujący wypowiedź „piesek"* 14

8 *Wrażenie i percepcja na przykładzie wzroku (oprac. własne na podst.: Goldstein, 2013)* 18

9 *Schematyczny przebieg procesu percepcyjnego (oprac. własne na podst.: Goldstein, 2013)* 19

10 *Kolejne etapy percepcji słuchowej prowadzące do zrozumienia sygnału mowy* . . 21

11 *Uproszczona wizualna analogia mechanizmu maskowania, w którym obiektem podstawowym jest figurka w białym nakryciu głowy: panel lewy – bez maskowania, panel środkowy – maskowanie energetyczne, panel prawy – maskowanie informacyjne* 29

12 *Binauralne konfiguracje przestrzenne pomiaru SRT – S0N0 oraz S0N90* 34

13 *Schematyczny przebieg procesów oddolnych i odgórnych w percepcji słuchowej* . 36

14 *Schematyczne przedstawienie możliwości wykorzystania informacji kontekstowej (zdanie: „Ala ma kota")* 39

15 *Przykładowa krzywa psychometryczna z zaznaczoną wartością progową* 46

16 *Przykładowy przebieg wyznaczania progu z wykorzystaniem procedury adaptacyjnej (próg wyznaczony po 20 prezentacjach: -1 dB)* 49

17 *Test zdaniowy matrix w języku polskim* 58

18 *Przykładowy ekran odpowiedzi w polskim teście matrycowym dla dzieci* 59

19 *Wysiłek słuchowy jako przykład procesu równoważenia zysku i kosztów percepcyjnych* 65

20 *Schematyczny podział metod oceny wysiłku słuchowego* 69

21 *Schemat odpowiedzi dynamicznej leżącej u podstaw pomiaru fNIRS* 72

22 *Fragmenty przebiegów czasowych wykorzystanych sygnałów maskujących – stacjonarnych TSN i cafeteria oraz zmodulowanych ICRA5-250 i IFFM* 85

23 *Przebieg audiogramów osób biorących udział w badaniu walidacyjnym (grupy słuchaczy: yNH – młodzi prawidłowo słyszący, oNH – starsi prawidłowo słyszący,*
M – z lekkim niedosłuchem, MO – z umiarkowanym niedosłuchem, S – ze znacznym niedosłuchem) 89

24 *Progi zrozumiałości mowy SRT w szumie stacjonarnym TSN a średnie progi słyszenia PTA (grupy słuchaczy: yNH – młodzi prawidłowo słyszący, oNH – starsi prawidłowo słyszący, M – z lekkim niedosłuchem, MO – z umiarkowanym niedosłuchem, S – ze znacznym niedosłuchem)* 91

25 *Progi zrozumiałości mowy SRT w szumie stacjonarnym TSN a średnie progi słyszenia PTA (grupy słuchaczy: yNH – młodzi prawidłowo słyszący, oNH – starsi prawidłowo słyszący, M – z lekkim niedosłuchem, MO – z umiarkowanym niedosłuchem, S – ze znacznym niedosłuchem)* 92

26 *Zależność SRT w szumach ICRA5-250 i TSN (grupy słuchaczy: yNH – młodzi prawidłowo słyszący, oNH – starsi prawidłowo słyszący, M – z lekkim niedosłuchem, MO – z umiarkowanym niedosłuchem, S – ze znacznym niedosłuchem)* 93

27 *Zmiany w wartości SRT w kolejnych pomiarach (P1–P5) w szumie stacjonarnym TSN [N=14]* 96

28 *Krzywe psychometryczne z wartościami SRT (czerwone punkty) dla grupy młodych, prawidłowo słyszących badanych (yNH)* 100

29 *Krzywe psychometryczne z wartościami SRT (czerwone punkty) dla grupy badanych z niedosłuchem (HI)* 101

30 *Zdanie „Tomasz wygra kilka czarnych okien" (szary) zamaskowane szumem stacjonarnym TSN (górny wykres) oraz szumem zmodulowanym IFFM (dolny wykres)* 105

31 *Wielkość uwolnienia od maskowania (masking release) w funkcji średnich progów słyszenia (PTA) w pięciu grupach badanych: yNH – młodzi prawidłowo słyszący, oNH – starsi prawidłowo słyszący, M – z lekkim niedosłuchem, MO – z umiarkowanym niedosłuchem, S – ze znacznym niedosłuchem* 108
32 *Zależność wielkości masking release od wieku w pięciu grupach badanych: yNH – młodzi prawidłowo słyszący, oNH – starsi prawidłowo słyszący, M – z lekkim niedosłuchem, MO – z umiarkowanym niedosłuchem, S – ze znacznym niedosłuchem* 109
33 *Procentowy udział potrzeb (usprawnień) zgłaszanych przez użytkowników aparatów słuchowych, kolejno: poprawa zrozumiałości mowy w szumie, lepsze powiązanie z telefonem, poprawa zrozumiałości mowy w ciszy, maskowanie szumów usznych, lepsze powiązanie z telefonem komórkowym, wyższa jakość dźwięków muzycznych (źr.: Kochkin, 2002)* 111
34 *Średnie progi słyszenia (z odchyleniem standardowym)* 113
35 *SRT w ciszy (panel lewy) i szumie stacjonarnym (prawy) w funkcji średnich progów słyszenia PTA* 116
36 *Związek PTA z SRT w szumach TSN i ICRA5-250* 117
37 *Fragment formularza odpowiedzi zmodyfikowanego formularza SSQ przetłumaczonego na język polski* 123
38 *Uśrednione przebiegi audiogramów uczestników badania z ACALES; kółka – ucho prawe, krzyżyki ucho lewe* 129
39 *Polska wersja skali ACALES (opracowanie własne na podst. Krueger i in., 2017)* 130
40 *Krzywe psychometryczne – grupa młodszych prawidłowo słyszących (yNH)* . . . 134
41 *Krzywe psychometryczne – grupa badanych z symetrycznym niedosłychem odbiorczym (HI)* 135
42 *Wartości SRT w ciszy w funkcji średnich progów słyszenia (PTA); yNH – młodzi prawidłowo słyszący, oNH – starsi prawidłowo słyszący, HI – osoby z symetrycznym odbiorczym niedosłuchem* 135
43 *Średnia ocena wysiłku słuchowego (listening effort) w funkcji SNR; yNH – młodzi prawidłowo słyszący, oNH – starsi prawidłowo słyszący, HI – osoby z symetrycznym odbiorczym niedosłuchem* 136
44 *Przebieg czasowy sygnału zawierającego szum turbiny wiatrowej (WT)* 143
45 *Związek progów SRT w szumie zmodulowanym ICRA5-250 (AF = 1) z SRT w szumie turbiny wiatrowej WT (AF = 2)* 146
46 *Związek progów SRT w szumie zmodulowanym ICRA5-250 (AF = 1) z SRT w szumie turbiny wiatrowej WT (AF = 3.7)* 146
47 *Klasy audiogramów zaproponowane przez Bisgaarda* [*za* : *Bisgaardiin.*, 2010]*; na osi x znajduje się częstotliwość (w kHZ), natomiast y – wartość progu dla danej częstotliwości (w dB HL)* 150

48 *Średnie wartości SRT per grupa z klasyfikacji Bisgaarda w ocenianych warunkach maskowania* 152
49 *Przykładowa krzywa zaniku o podwójnym nachyleniu – amplituda [dBFS] w funkcji czasu (na podst.: Błasiński, 2023)* 158
50 *Związek RT20 z wysiłkiem słuchowym w skali Likerta (1–7)* 159
51 *Powiązanie STI ze średnim wysiłkiem słuchowym (1–7)* 160
52 *Uśrednione progi słyszenia biorących udział w badaniu (N=22)* 165
53 *Średnie SRT polskiego testu zdaniowego (PTZ) w konfiguracji S0N0* 167
54 *Średnie SRT polskiego testu zdaniowego (PTZ) w konfiguracji S0N90* 167

Spis tabel

1 *Osoby biorące udział w walidacji polskiego testu zdaniowego matrix (grupy słuchaczy: yNH – młodzi prawidłowo słyszący, oNH – starsi prawidłowo słyszący, M – z lekkim niedosłuchem, MO – z umiarkowanym niedosłuchem, S – ze znacznym niedosłuchem)* 88

2 *Wartości SRT w szumach maskujących (TSN, ICRA5-250, IFFM, cafeteria) i ciszy wraz z SD (grupy słuchaczy: yNH – młodzi prawidłowo słyszący, oNH – starsi prawidłowo słyszący,*
M – z lekkim niedosłuchem, MO – z umiarkowanym niedosłuchem, S – ze znacznym niedosłuchem) 90

3 *Czułość i swoistość polskiego testu zdaniowego matrix w ciszy i szumie stacjonarnym TSN* 94

4 *Charakterystyka słuchaczy biorących udział w badaniu nachylenia krzywych psychometrycznych (w nawiasach wartości odchylenia standardowego); yNH – młodzi prawidłowo słyszący, oNH – starsi prawidłowo słyszący, HI – osoby z symetrycznym odbiorczym niedosłuchem* 99

5 *Nachylenie krzywych psychometrycznych [%/dB] wraz z odchyleniem stand. (SD);*
yNH – młodzi prawidłowo słyszący, oNH – starsi prawidłowo słyszący, HI – osoby z symetrycznym odbiorczym niedosłuchem 102

6 *Grupy badanych (liczebność, wiek, próg słyszenia) z odch. standardowym w nawiasie; yNH – młodzi prawidłowo słyszący, oNH – starsi prawidłowo słyszący, M – badani z lekkim niedosłuchem, MO – z umiarkowanym niedosłuchem, S – ze znacznym niedosłuchem* 107

7 *Średnie wartości SRT w szumie TSN, ICRA5-250 i ciszy w pięciu grupach badanych: yNH – młodzi prawidłowo słyszący, oNH – starsi prawidłowo słyszący, M – z lekkim niedosłuchem, MO – z umiarkowanym niedosłuchem, S – ze znacznym niedosłuchem* 107

8 *Średnie wartości SRT bez kompensacji, z własnym aparatem słuchowym (HA) oraz MHA* 116

9 *Zysk w zrozumiałości mowy przy kompensacji (M)HA* 118

10 *Wybrane i przetłumaczone na język polski pytania z kwestionariusza SSQ (numeracja odnosi się do tej, zastosowanej w pełnej wersji SSQ)* 122

11 *Średnie oceny w kategoriach formularza SSQ* 123
12 *Współczynnik korelacji między średnią oceną SSQ a SRT* 124
13 *Informacje na temat uczestników badania z wykorzystaniem ACALES - yNH i oNH kolejno: młodsi i starsi (≥5 0 r.ż.) prawidłowo słyszący, HI – osoby z symetrycznym niedosłuchem odbiorczym; w nawiasach wartości odchylenia standardowego* 128
14 *Średnie wartości SRT w trzech badanych grupach: yNH - młodzi prawidłowo słyszący, oNH - starsi prawidłowo słyszący, HI - osoby z symetrycznym odbiorczym niedosłuchem* 133
15 *ICC przy minimalnych, środkowych i maksymalnych punktach skali ACALES* . 138
16 *Podstawowe informacje o badanych wraz z odchyleniem standardowym (w nawiasach.* 141
17 *Uśrednione wartości SRT wraz z odchyleniem standardowym (SD) dla wszystkich warunków pomiarowych (a-TSN, b-ICRA5-250, c-cafeteria, d-wind turbine)* 144
18 *Liczba słuchaczy przypisanych do odpowiednich grup wg klasyfikacji Bisgaarda* 151
19 *Uśrednione wartości SRT (wraz z odchyleniem standardowym) w analizowanych klasach (N1, N2, N3, N4 oraz S1, S2) i warunkach pomiarowych* 152
20 *Analiza post hoc – istotność statystyczna różnic SRT w poszczególnych szumach maskujących w danej parze klas ubytku słuchu (1 – istotność statystycza różnicy, 0 – brak istotności)* 153
21 *Uśrednione wartości wysiłku słuchowego (LE) oraz zmierzone wartości czasu pogłosu (RT)* 158

Bibliografia

Ahnert W., Tennhardt H.P. (2008), Acoustics for Auditoriums and Concert Halls, [w:] *Handbook for Sound Engineers*, red. G.M. Ballou G.M., Focal Press, Waltham.

Alhanbali S., Dawes P., Millman R.E., Munro K.J. (2019), Measures of Listening Effort Are Multidimensional, *Ear Hear*, 40(5), s. 1084–1097. doi: 10.1097/AUD.0000000000000697.

Aparicio M., Peigneux P., Charlier B., Baleriaux D., Kavec M., Leybart J. (2017), The neural basis of speech perception through lipreading and manual cues: evidence from deaf native users of cued speech, *Front. Psychol*, 8, 426. doi: 10.3389/fpsyg.2017.00426.

Anderson Gosselin P., Gagné J.P. (2011), Older adults expend more listening effort than young adults recognizing speech in noise, *J Speech Lang Hear Res*, 54(3), s. 944–958. doi: 10.1044/1092-4388(2010/10-0069).

Ando Y. (2007), Concert hall acoustics based on subjective preference theory, [w:] *Springer handbook of acoustics*, ed. T. Rossing, Springer, Nowy Jork, s. 351–386.

Aniansson G. (1978), Speech Intelligibility in and Speech Interference Levels of Traffic Noise in doi: 10.3109/00016487809123488.

Arbogast T.L., Mason C.R., Kidd G. (2005), The effect of spatial separation on informational masking of speech in normal-hearing and hearing-impaired listeners, *J Acoust Soc Am*, 117(4), s. 2169–2180. doi: 10.1121/1.1861598.

Atcherson S.R., Mendel L.L., Baltimore W.J., Patro C., Lee S., Pousson M., Spann M.J. (2017), The Effect of Conventional and Transparent Surgical Masks on Speech Understanding in Individuals with and without Hearing Loss, *J Am Acad Audiol*, 28(1), s. 58–67. doi: 10.3766/jaaa.15151.

Bacon S.P., Opie J.M., Montoya D.Y. (1998), The effects of hearing loss and noise masking on the masking release for speech in temporally complex backgrounds, *J*

Speech Lang Hear Res, 41, s. 549–563. doi: 10.1044/jslhr.4103.549.

Bakker R., Gillian S. (2014), The history of Active Acoustic Enhancement Systems, *Proceedings of the Institute of Acoustics*, 36(2), s. 56–65.

Beasley D.S., Schwimmer S., Rintelmann W.F. (1972), Intelligibility of time-compressed CNC monosyllables, *J Speech Hear Res*, 15(2), s. 340–350. doi: 10.1044/jshr.1502.340.

von Békésy G.(1956), Current status of theories of hearing, *Science*, 123, s. 779–783. doi: 10.1126/science.123.3201.779

Von Békésy G. (1959), Synchronism of neural discharges and their demultiplication in pitch perception on the skin and in hearing, *Journal of the Acoustical Society of America*, 31, s. 338–349. doi: 10.1121/1.1907722

von Békésy G.V. (1960), *Experiments in hearing*, McGraw–Hill, Londyn.

Bennett R.J., Meyer C.J., Eikelboom R.H., Atlas M.D. (2018), Evaluating Hearing Aid Management: Development of the Hearing Aid Skills and Knowledge Inventory (HASKI), Am J Audiol, 27(3), s. 333–348. doi: 10.1044/2018_AJA-18-0050.

Beranek L.L. (2004), Concert Halls and Opera Houses: Music, [w:] *Acoustics, and Architecture*, Springer, Nowy Jork.

Bernarding C., Strauss D.J., Hannemann R., Corona-Strauss F.I. (2012), Quantification of listening effort correlates in the oscillatory EEG activity: a feasibility study, *Annu Int Conf IEEE Eng Med Biol Soc*, s. 4615-4618. doi: 10.1109/EMBC.2012.6346995.

Best V., Boyd A.D., Sen K. (2023), An Effect of Gaze Direction in Cocktail Party Listening, *Trends Hear*, 27. doi: 10.1177/23312165231152356.

Beutelmann R., Brand T., Kollmeier B. (2010), Revision, extension, and evaluation of a binaural speech intelligibility model, *J Acoust Soc Am*, 127(4), s. 2479–2497. doi: 10.1121/1.3295575.

Bharadwaj H.M., Masud S., Mehraei G., Verhulst S., Shinn-Cunningham B.G. (2015) Individual differences reveal correlates of hidden hearing deficits, *J Neurosci*, 35, s. 2161–2172. doi: 10.1523/JNEUROSCI.3915-14.2015.

Bisgaard N., Vlaming M.S., Dahlquist M. (2010), Standard audiograms for the IEC 60118-15 measurement procedure, *Trends Amplif*, 14(2), s. 113–120. doi: 10.1177/1084713810379609.

Błasinski Ł. (2023), Objective evaluation of multipurpose enclosures equipped with active acoustic enhancement systems, *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, 265, s. 3502–3512. doi: 10.3397/IN_2022_0497.

Bogert B.P., Healy M.J.R., Tukey J.W. (1963), The Quefrency Analysis of Time Series for Echoes: Cepstrum, Pseudo-Autocovariance, Cross-cepstrum and Saphe Cracking, [w:] *Proceedings of the Symposium on Time Series Analysis*, red. M. Rosenblat, Wiley, Nowy Jork, s. 209–243.

Bonetti LV., Hassan S.A., Lau S.T., Melo L.T., Tanaka T., Patterson K.K., Reid W.D. (2019), Oxyhemoglobin changes in the prefrontal cortex in response to cognitive tasks: a systematic review, *The International Journal of Neuroscience*, 129(2), s. 195–203. doi: 10.1080/00207454.2018.1518906.

Bottalico P., Murgia S., Puglisi G.E., Astolfi A., Kirk K.I. (2020), Effect of masks on speech intelligibility in auralized classrooms, *J Acoust. Soc Am*, 148, a. 2878–2884. doi: 10.1121/10.0002450.

Boucsein W., Ottmann W. (1996), Psychophysiological stress effects from the combination of night-shift work and noise, *Biological Psychology*, 42(3), s. 301–322, doi: 10.1016/0301-0511(95)05164-3.

Brachmanski S., Dobrucki A. (2021), Impact of the level of noise and echo on the reaction time of listeners in the perception of logatoms, *Vibrations in Physical Systems*, 32(2), doi: 10.21008/j.0860-6897.2021.2.15.

Brachmański S., Staroniewicz P. (1999), Fonetyczna struktura materiału testowego stosowanego w subiektywnych pomiarach jakości mowy, *Speech and Language Technology*, 3, s. 71–80.

Bradley J.S. (1986), Predictors of speech intelligibility in rooms, J Acoust Soc Am, 80(3), s. 837–845. doi: 10.1121/1.393907.

Brand T., Kollmeier B. (2002), Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests, *J Acoust Soc Am*, 111(6), s. 2801–2810. doi: 10.1121/1.1479152.

Bregman A.S. (1990), *Auditory scene analysis: The perceptual organization of sound*, The MIT Press, Cambridge MA.

Broadbent D.E. (1958), Perception and communication. Pergamon Press, Oxford. doi: 10.1037/10037-000.

Bronkhorst A.W. (2000), The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions, *Acta Acoustica united with Acoustica*, 86(1), s. 117–128. doi: 10.3758/s13414-015-0882-9.

Bronkhorst A.W., Plomp R. (1989), Binaural speech intelligibility in noise for hearing-impaired listeners, *J Acoust Soc Am*, 86(4), s. 1374–1783. doi: 10.1121/1.398697

Brouwer S., Van Engen K.J., Calandruccio L., Bradlow A.R. (2012), Linguistic contributions to speech-on-speech masking for native and non-native listeners: language familiarity and semantic content, *J Acoust Soc Am*, 131(2), s. 1449–1464. doi: 10.1121/1.3675943.

Brown J.A., Bidelman G.M. (2022), Familiarity of Background Music Modulates the Cortical Tracking of Target Speech at the "Cocktail Party", *Brain Sci*, 12(10), 1320. doi: 10.3390/brainsci12101320.

Brungart D.S. (2001), Informational and energetic masking effects in the perception of two simultaneous talkers, *J Acoust Soc Am,* 109(3), s. 1101–1109. doi: 10.1121/1.1345696.

Brungart D.S., Simpson B.D. (2002), Within-ear and across-ear interference in a cocktail-party listening task, *J Acoust Soc Am*, 112(6), s. 2985–2995. doi: 10.1121/1.1512703.

Buss E., Hodge S.E., Calandruccio L., Leibold L.J., Grose J.H. (2019), Masked Sentence Recognition in Children, Young Adults, and Older Adults: Age-Dependent Effects of Semantic Context and Masker Type, *Ear Hear*, 40(5), s. 1117–1126. doi: 10.1097/AUD.0000000000000692.

Calandruccio L., Dhar S., Bradlow A.R. (2010), Speech-on-speech masking with variable access to the linguistic content of the masker speech, *J Acoust Soc Am*, 128(2), 860–869. doi: 10.1121/1.3458857.

Campbell G.A. (1910), Telephonic intelligibility, *Philosophical Magazine*, 19, s. 152–159.

Carhart R., Tillman T.W. (1970), Interaction of competing speech signals with hearing losses, *Arch Otolaryngol*, 91(3), s. 273–279. doi: 10.1001/archotol.1970.00770040379010.

Carlson M.C., Helms M.J., Steffens D.C., Burke J.R., Potter G.G., Plassman B.L. (2008), Midlife activity predicts risk of dementia in older male twin pairs, *Alzheimers*

Dement, 4(5), s. 324–331. doi: 10.1016/j.jalz.2008.07.002.

Carr K., Kihm J. (2022), MarkeTrak – Tracking the Pulse of the Hearing Aid Market, *Semin Hear*, 43(4), s. 277–288. doi: 10.1055/s-0042-1758380.

Carraturo S., Chen S., Van, Engen K. (2023), The cognitive demands of adverse listening conditions for monolingual and bilingual listeners: A pupillometry study, *Berkeley Program in Law and Economics*, 153(3), A341–A341.

Chermak G., Musiek F. (1997), *Central auditory processing disorders: New perspectives*, Singular, San Diego.

Chisolm T.H., Johnson C.E., Danhauer J.L., Portz L.J., Abrams H.B., Lesner S., McCarthy P.A., Newman C.W. (2007), A systematic review of health-related quality of life and hearing aids: final report of the American Academy of Audiology Task Force On the Health-Related Quality of Life Benefits of Amplification in Adults, *J Am Acad Audiol*, 18(2), s. 151–83. doi: 10.3766/jaaa.18.2.7.

Christiansen C., Dau T. (2012), Relationship between masking release in fluctuating maskers and speech reception thresholds in stationary noise, *J Acoust Soc Am*, 132(3), s. 1655–1666. doi: 10.1121/1.4742732.

Chua M.H., Cheng W., Goh S.S. i in. (2020), Face Masks in the New COVID-19 Normal: Materials, Testing, and Perspectives, *Research (Wash DC)*, 7286735. doi: 10.34133/2020/7286735.

Clinard C.G., Tremblay K.L. (2013), Aging degrades the neural encoding of simple and complex sounds in the human brainstem, *J Am Acad Audiol*, 24(7), s. 590–599. doi: 10.3766/jaaa.24.7.7.

Cooke M. (2006), A glimpsing model of speech perception in noise, *J Acoust Soc Am*, 119(3), s. 1562–1573. doi: 10.1121/1.2166600.

Corey R.M., Jones U., Singer A.C. (2020), Acoustic effects of medical, cloth, and transparent face masks on speech signals, *J Acoust Soc Am*, 148(4), s. 2371–2375. doi: 10.1121/10.0002279.

Cueva R.A. (2004), Auditory brainstem response versus magnetic resonance imaging for the evaluation of asymmetric sensorineural hearing loss, *Laryngoscope*, 114(10), 1686–1692. doi: 10.1097/00005537-200410000-00003.

Cytowic R.E., Eagleman D.M. (2009), *Wednesday is indigo blue: Discovering the brain of synesthesia*, Boston Review.

Davis J. F., Mundl W.J. (1964), A dual modality tracking system, *Psychophysiology*, 1(2), s. 183–191. doi: 10.1111/j.1469-8986.1964.tb03233.

Davis M.H., Johnsrude I.S. (2007), Hearing speech sounds: top-down influences on the interface between audition and speech perception, *Hear Res,* 229(1-2), s. 132–147. doi: 10.1016/j.heares.2007.01.014.

Deacon T.W. (1997), *The Symbolic Species: The Co-evolution of Language and the Brain*, W.W. Norton & Company.

Demenko G. (2011), Percepcja mowy w zarysie, [w:] *Wybrane zagadnienia z audiometrii mowy*, red. A. Obrębowski, Wydawnictwo Naukowe UM w Poznaniu, s. 33–57.

Deniz B., Gülmez Z.D., Kara H., Kara E. (2024), Effect of Digital Noise Reduction in Hearing Aids on Speech Intelligibility in Both Quiet and Noisy Environments, *Noise Health*, 26(121), s. 220–225. doi: 10.4103/nah.nah_67_23.

Desjardins J.L., Doherty K.A. (2014), The effect of hearing aid noise reduction on listening effort in hearing-impaired adults, *Ear Hear*, 35(6), s. 600-610. doi: 10.1097/AUD.0000000000000028.

Dietrich S., Hertrich I., Ackermann H. (2013), Ultra-fast speech comprehension in blind subjects engages primary visual cortex, fusiform gyrus, and pulvinar – a functional magnetic resonance imaging (fMRI) study, *BMC Neurosci*, 14, s. 4–74 doi: 10.1186/1471-2202-14-74.

Dirks D.D., Bower D.R. (1969), Masking effects of speech competing messages, *J Speech Hear Res*, 12(2), s. 229–245. doi: 10.1044/jshr.1202.229.

Dreschler W.A., Boymans M., Goverts S.T., Kramer, S. (2001), The benefit of binaural hearing aid fittings: results of a retrospective study, *Clinical Otolaryngology & Allied Sciences*, 26, s. 346–346. doi: 10.1046/j.1365-2273.2001.00479-20.x.

Dronkers N.F., Plaisant O., Iba-Zizen M.T., Cabanis E.A. (2007), Paul Broca's historic cases: high resolution MR imaging of the brains of Leborgne and Lelong, *Brain*, 130(5), s. 1432–1441. doi: 10.1093/brain/awm042.

Drullman R., Festen J.M., Plomp R.(1994), Effect of temporal envelope smearing on speech reception, *J Acoust. Soc Am*, 95, s. 1053–1064. doi: 10.1121/1.408467

Dubno J.R., Ahlstrom J.B., Horwitz A.R. (2002), Spectral contributions to the benefit from spatial separation of speech and noise, *J Speech Lang Hear Res*, 45(6),

s. 1297–1310. doi: 10.1044/1092-4388(2002/104).

Dupoux E., Green K. (1997), Perceptual adjustment to highly compressed speech: Effects of talker and rate changes, *J Exp Psychol Hum Percept Perform*, 23(3), s, 914–927. doi: 10.1037//0096-1523.23.3.914

Duquesnoy A.J., Plomp R. (1983), The effect of a hearing aid on the speech-reception threshold of hearing-impaired listeners in quiet and in noise, *J Acoust Soc Am*, 73(6), s. 2166–2173. doi: 10.1121/1.389540.

Eco U. (1996), *Nieobecna Struktura*, Wyd. KR, Warszawa, s.62.

Eisenberg L.S., Dirks D.D., Bell T.S. (1995), Speech recognition in amplitude-modulated noise of listeners with normal and listeners with impaired hearing, *J Speech Hear Res*, 38(1), s. 222–233. doi: 10.1044/jshr.3801.222.

Elliott L.L., Connors S., Kills E., Levin S. (1979), Children's understanding of monosyllabic nouns in quiet and in noise, *J Acoust Soc Am*, 66, s. 12–21. doi: 10.1121/1.383065.

Everest F.A. (2001), *The Master Handbook of Acoustics* (4 ed.), Mc Graw Hill, Nowy Jork, s. 153.

Ezzatian P., Li L., Pichora-Fuller K., Schneider B.A. (2012), The effect of energetic and informational masking on the time-course of stream segregation: Evidence that streaming depends on vocal fine structure cues, *Lang Cogn Process*, 27, s. 1056–1088. doi: 10.1080/01690965.2011.591934.

Fant G. (1960), *Acoustic theory of speech production*, Mouton, Haga.

Fairbanks G., Kodman F. (1957), Word Intelligibility as a Function of Time Compression, *J Acoust Soc Am*, 29 (5), s. 636–641. doi: 10.1121/1.1908992.

Fairbanks G. (1958), Test of phonemic differentiation: The rhyme test, *J Acoust Soc Am*, 30(7), s. 596–600.

Felcyn J., Preis A., Gogol R. (2022), Evaluation of Annoyance Due to Wind Turbine Noise Based on Pre-learned Patterns, Vibrations in Physical Systems, 33(3), s. 1–5. doi: 10.21008/j.0860-6897.2022.3.10.

Ferrari M., Quaresima V. (2012), A brief review on the history of human functional near-infrared spectroscopy (fNIRS) development and fields of application, *Neuroimage*, 63(2), s. 921–935. doi: 10.1016/j.neuroimage.2012.03.049.

Ferrari M., Mottola L., Quaresima V. (2004), Principles, techniques, and limitations of near infrared spectroscop, *Canadian Journal of Applied Physiology*, 29(4), s. 463–487. doi: 10.1139/h04-031.

Festen J.M., Plomp R. (1986), Speech-reception threshold in noise with one and two hearing aids, *J Acoust Soc Am*, 79(2), s. 465-471. doi: 10.1121/1.393534.

Festen J.M., Plomp R. (1983), Relations between auditory functions in impaired hearing, *J Acoust Soc Am*, 73(2), s. 652–662. doi: 10.1121/1.388957.

Festen J.M., Plomp R. (1990), Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing, *J Acoust Soc Am*, 88, s. 1725–1736. doi: 10.1121/1.400247

Gordon-Salant S., Fitzgibbons P.J. (1993), Temporal factors and speech recognition performance in young and elderly listeners, *J Speech Hear Res*, 36(6), s. 1276–1285. doi: 10.1044/jshr.3606.1276.

Fletcher H., Galt R.H. (1950), The perception of speech and its relation to telephony, *J Acoust Soc Am*, 22, s. 89–151. doi: 10.1121/1.1906605

Francart T., Lenssen A., Wouters J. (2011), Enhancement of interaural level differences improves sound localization in bimodal hearing, *J Acoust Soc Am*, 2130(5), s. 2817–2826. doi: 10.1121/1.3641414.

Frisina R.D. (1997), Speech recognition in noise and presbycusis: Relations to possible neural mechanisms, *Hear Res*, 106, s. 95–104. doi: 10.1016/S0378-5955(97)00006-3.

Fu Q.J., Galvin J.J., Wang X. (2001), Recognition of time-distorted sentences by normal-hearing and cochlear-implant listeners, *J Acoust Soc Am*, 109(1), s. 379–384. doi: 10.1121/1.1327578.

Füllgrabe C. Moore B.C.J., Stone M.A. (2014), Age-group differences in speech identification despite matched audiometrically normal hearing: Contributions from auditory temporal processing and cognition, *Front Aging Neurosc*i, 6, 347. doi: 10.3389/fnagi.2014.00347.

Ganesh G., Srinivasan V.S., Krishnamurthi S. (2016), A model to demonstrate the place theory of hearing, *Adv Physiol Educ*, 40(2), s. 191–193. doi: 10.1152/advan.00128.2015.

Garlińska U., Michalak P., Pawłowski S., Popielarczyk T. (2015), Pomiary

zrozumiałości mowy dźwiękowych systemów ostrzegawczych, *Bezpieczeństwo i Technika Pożarnicza*, 3, s. 161–171.

Gatehouse S., Noble W. (2004), The Speech, Spatial and Qualities of Hearing Scale (SSQ), *Int J Audiol*, 43, s. 85–89. doi: 10.1080/14992020400050014.

Gehr S.E., Sommers M.S. (1999), Age differences in backward masking, *J Acoust Soc Am*, 106(5), s. 2793–2799. doi: 10.1121/1.428104.

Giovannone N., Theodore R.M. (2020), Individual Differences in Lexical Contributions to Speech Perception, *J Speech Lang Hear Res*, 64(3), s. 707–724. doi: 10.1044/2020_JSLHR-20-00283.

Glasberg B.R., Moore B.C.J. (1986) Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments, *J Acoust Soc Am*, 79(4), s. 1020–1033. doi: 10.1121/1.393374.

Glasberg B.R., Moore B.C.J., Bacon S.P. (1987), Gap detection and masking in hearing-impaired and normal-hearing subjects, *J Acoust Soc Am*, 81(5), s. 1546–1556. doi: 10.1121/1.394507.

Gleason J.B., Ratner N.B. (2005), *Psycholingwistyka*, Gdańskie Wydawnictwo Psychologiczne, Gdańsk.

Glover G.H. (2011), Overview of functional magnetic resonance imaging, *Neurosurgery Clinics of North America*, 22(2), s. 133–139, doi: 10.1016/j.nec.2010.11.001.

Goldstein E. (2013), *Sensation and perception* (9 ed.), Cengage Learning, Boston MA.

Gosselin P.A., Gagné J.P. (2010), Use of a dual–task paradigm to measure listening effort, *Canadian Journal of Speech–Language Pathology and Audiology*, 34(1), s. 43–51.

Gosselin P.A., Gagné J.P. (2011), Older adults expend more listening effort than young adults recognizing speech in noise, *J Speech Lang Hear Res*, 54(3), s. 944–958. doi: 10.1044/1092-4388(2010/10-0069).

Gransier R., van Wieringen A., Wouters J. (2022), The intelligibility of time-compressed speech is correlated with the ability to listen in modulated noise, *J Assoc Res Otolaryngol*, 23, s. 413–426. doi: 10.1007/s10162-021-00832-0.

Greenspan S.L., Bennett R.W., Syrdal A.K. (1998), An evaluation of the diagnostic

rhyme test, *Int J Speech Technol*, 2, s. 201–214. doi: 10.1007/BF02111208.

Hagerman B. (1982), Sentences for testing speech intelligibility in noise, *Scand Audiol*, 11(2), s. 79–87. doi: 10.3109/01050398209076203.

Hampton T., Crunkhorn R., Lowe N., Bhat J., Hogg E., Afifi W. (2020), The negative impact of wearing personal protective equipment on communication during coronavirus disease, *J Laryngol Otol*, 134, s. 577–581. doi: 10.1017/S0022215120001437.

Hart S., Staveland L. (1988), Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research, [w:] *Human Mental Workload*, red. P. Hancock, N. Meshkati, North Holland, Amsterdam, s. 139–183. doi: 10.1016/S0166-4115(08)62386-9.

Hétu R., Riverin L., Lalande N., Getty L., St-Cyr C. (1988), Qualitative analysis of the handicap associated with occupational hearing loss, *Br J Audiol*, 22(4), s. 251–64. doi: 10.3109/03005368809076462.

Helfer KS, Freyman RL, The role of visual speech cues in reducing energetic and informational masking, *J Acoust Soc Am*, 117(2), s. 842–849. doi: 10.1121/1.1836832.

Hick C.B., Tharpe A.M. (2002), Listening effort and fatigue in school-age children with and without hearing loss, *J Speech Lang Hear Res*, 45(3), s. 573–584. doi: 10.1044/1092-4388(2002/046).

Hillis C.L., Uchanski R.M., Davidson L.S. (2023), *Common Sounds Audiograms: Quantitative Analyses and Recommendations*, Semin Hear, 44: 49–63. doi: 10.1055/s-0043-1764128.

Hockett C. (1979), Zagadnienie uniwersaliów w języku, [w:] *Językoznawstwo strukturalne. Wybór tekstów*, red. H. Kurkowska i A. Weinsberg, Wydawnictwo Naukowe PWN, Warszawa, s. 209–228.

Holube I., Fredelake S., Vlaming M., Kollmeier B. (2010), Development and analysis of an International Speech Test Signal (ISTS), *Int J Audiol*, 49(12), s. 891–903. doi: 10.3109/14992027.2010.506889.

Holube I., von Gablenz P. (2011), A Representative Study of Hearing Ability in North West Germany, *Audiol Res*, 1(1), e3. doi: 10.4081/audiores.2011.e3.

Holube I., Haeder K., Imbery C., Weber R. (2016), Subjective Listening Effort and Electrodermal Activity in Listening Situations with Reverberation and Noise,

Trends Hear, 20, 2331216516667734. doi: 10.1177/2331216516667734.

House A.S., Williams C.E., Hecker M.H.L., Kryter K.D. (1965), Articulation-testing methods: Consonantal differentiation with a closed-response set, *J Acoust Soc Am*, 37, s. 158–166.

Houtgast T., Steeneken H.J.M. (1973), The Modulation Transfer Function in Room Acoustics as a Predictor of Speech Intelligibility, *The Journal of the Acoustical Society of America*, 54, s. 557–557.

Houtgast T., Steeneken H.J.M. (1985), A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria, *J Acoust Soc Am*, 77(3), s. 1069–1077. doi: 10.1121/1.392224.

Howard D.M., Angus J.A.S. (2017), *Acoustics and Psychoacoustics* (5 ed.), Routledge, Oxfordshire.

Humes L.E., Kidd G.R., Lentz J.J. (2013), Auditory and cognitive factors underlying individual differences in aided speech-understanding among older adults, *Front Syst Neurosci*, 1, s. 7–55. doi: 10.3389/fnsys.2013.00055.

Hunter C., Humes L. (2022), Predictive Sentence Context Reduces Listening Effort in Older Adults With and Without Hearing Loss and With High and Low Working Memory Capacity, *Ear Hear,* 43(4), s. 1164–1177. doi: 10.1097/AUD.0000000000001192.

Ibelings S., Brand T., Holube I. (2022), Speech Recognition and Listening Effort of Meaningful Sentences Using Synthetic Speech, *Trends in Hearing*, 26. doi:10.1177/23312165221130656.

Idrizbegovic E., Hederstierna C., Dahlquist M. (2013), Short-term longitudinal study of central auditory function in Alzheimer's disease and mild cognitive impairment, *Dement Geriatr Cogn Dis Extra*, 3, s. 468–471. doi: 10.1159/000355371.

Ignatious E., Azam S., Jonkman M., De Boer F. (2023), Binaural masking level difference for pure tone signals, *J Otol*, 18(3), s. 160–167. doi: 10.1016/j.joto.2023.06.001.

Illg A., Amann E., Koinig K.A., Anderson I., Lenarz T., Billinger-Finke M. (2023), A holistic perspective on hearing loss: first quality-of-life questionnaire (HL-QOL) for people with hearing loss based on the international classification of functioning, disability, and health, *Front Audiol Otol*, 1, 1207220. doi: 10.3389/fauot.2023.1207220.

Iliadou V., Kaprinis S. (2003), Clinical psychoacoustics in Alzheimer's disease central auditory processing disorders and speech deterioration, *Ann Gen Hosp Psychiatry*, 2, 12. doi: 10.1186/1475-2832-2-12.

Iwankiewicz S. (1961), Audiometria mowy, *Otolaryng Po*l, 15(1), s. 277.

Jansen S., Luts H., Wagener K.C., Kollmeier B., Del Rio M., Dauman R., James C., Fraysse B., Vormès E., Frachet B., Wouters J., van Wieringen A. (2012), Comparison of three types of French speech-in-noise tests: a multi-center study, *Int J Audiol*, 51(3), s. 164–173. doi: 10.3109/14992027.2011.633568.

Jassem W. (1973), *Podstawy fonetyki akustycznej*, Państwowe Wydawn. Naukowe, Warszawa.

Jijo P.M., Asha Y., 2013, Audiological findings and aided performance in individuals with auditory neuropathy spectrum disorders (ANSD) – a retrospective study, *J Hear Sc*, 3, s. 18–26. doi: 10.17430/883990.

Johnson J., Xu J., Cox R., Pendergraft P. (2015), A Comparison of Two Methods for Measuring Listening Effort As Part of an Audiological Test Battery, *Am J Audiol*, 24(3), 419–431. doi: 10.1044/2015_AJA-14-0058.

Jorasz U. (1998), *Wykłady z psychoakustyki*, Wydawnictwo Naukowe UAM, Poznań.

Kahneman D., Tversky A. (1973), On the psychology of prediction, *Psychological Review,* 80(4), s. 237–251. doi: 10.1037/h0034747.

Kalikow D.N., Stevens K.N., Elliott LL. (1977), Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability, *J Acoust Soc Am*, 61(5), s. 1337–1351. doi: 10.1121/1.381436.

Kandel E.R., Schwartz J.H., Jessell T.M., Siegelbaum S.A., Hudspeth A.J. (2012), *Principles of neural science* (5 ed.), McGraw Hill Professional, Nowy Jork.

Kayser H., Ewert S.D., Anemüller J., Rohdenburg T., Hohmann V., Kollmeier B. (2009), Database of multichannel in-ear and behind-the-ear head-related and binaural room impulse responses, *EURASIP Journal of Advanced Signal Processing*, 298605(1),s. 1–10.

Keidser G., Dillon H., Flax M., Ching T., Brewer S. (2011), The NAL-NL2 Prescription Procedure, *Audiology Research*, 1(1), e24. doi: 10.4081/audiores.2011.e24.

Kent R.D. (2004), *The MIT Encyclopedia of Communication Disorders*, MIT Press.

Kaplan Neeman R., Roziner I., Muchnik C. (2022), A Clinical Paradigm for Listening Effort Assessment in Middle-Aged Listeners, *Frontiers in Psychology*, 13, 820227. doi: 10.3389/fpsyg.2022.820227.

Kerlin J.R., Shahin A.J., Miller L.M. (2010), Attentional gain control of ongoing cortical speech representations in a "cocktail party", *J Neurosci*, 30(2), s. 620–628. doi: 10.1523/JNEUROSCI.3631-09.2010.

Kiessling J., Pichora-Fuller M.K., Gatehouse S., Stephens D., Arlinger S., Chisolm T., Davis A.C., Erber N.P., Hickson L., Holmes A., Rosenhall U., von Wedel H. (2003), Candidature for and delivery of audiological services: special needs of older people, *Int J Audiol*, 42(2), s. 92–101. doi: 10.3109/14992020309074650.

King S. (2014), Measuring a decade of progress in text-to-speech, *Loquens*, 1(1), e006. doi: 10.3989/loquens.2014.006.

Klink K.B., Meis M., Schulte M. (2012a), Measuring listening effort in the field of audiology—A literature review of methods, part 1, *Zeitschrift für Audiologie – Audiological Acoustics*, 51(2), s. 60–67.

Klink K.B., Meis M., Schulte M. (2012b), Measuring listening effort in the field of audiology—A literature review of methods, part 2, *Zeitschrift für Audiologie – Audiological Acoustics*, 51(3), s. 95–105.

Kociński J., Sęk A. (2005), Speech intelligibility in various spatial configurations of background noise, *Archives of Acoustics*, 30(2), s. 173–191.

Kociński J. (2007), Speech Intelligibility Improvement Using Convolutive Blind Source Separation Assisted by Denoising Algorithms, *Speech Communication*, 50(1), s. 29–37. doi: 10.1016/j.specom.2007.06.003.

Kociński J., Niemiec D. (2016) Time-compressed speech intelligibility in different reverberant conditions, *Applied Acoustics*, 113, s. 58–63. doi: 10.1016/j.apacoust.2016.06.009.

Kociński J., Ozimek E. (2015), Speech Intelligibility in Rooms with and without an Induction Loop for Hearing Aid Users, *Archives of Acoustics*, 40(1), s. 51–58. doi: 10.1515/aoa-2015-0007.

Kociński J., Ozimek E. (2017), Logatome and Sentence Recognition Related to Acoustic Parameters of Enclosures, *Archives of Acoustics*, 42(3), s. 385–394.

Koelewijn T., Zekveld A.A., Festen J.M., Kramer S.E. (2012), Pupil dilation

uncovers extra listening effort in the presence of a single-talker masker, *Ear Hear*, 33(2), s. 291–300. doi: 10.1097/AUD.0b013e3182310019.

Kollmeier B., Peissig J., Hohmann V. (1993), Real-time multiband dynamic compression and noise reduction for binaural hearing aids, *J Rehabil Res Dev*, 30(1), s. 82–94.

Kollmeier B., Wesselkamp M. (1997), Development and evaluation of a sentence test for objective and subjective speech intelligibility assessment, *J Acoust Soc Am*, 102(4), s. 1085–1099. doi: 10.1121/1.419624.

Kollmeier B., Warzybok A., Hochmuth S., Zokoll M.A., Uslar V., Brand T., Wagener K.C. (2015), The multilingual matrix test: Principles, applications, and comparison across languages: A review, *Int J Audio*l, 54(2), s. 3–16. doi: 10.3109/14992027.2015.1020971.

Keidser G. (1993), Normative data in quiet and in noise for "DANTALE": A Danish speech material, *Scandinavian Audiology*, 22(4), s. 231–236. doi: 10.3109/01050399309047474.

Kramer S.E., Kapteyn T.S., Houtgast T. (2006), Occupational performance: comparing normally-hearing and hearing-impaired employees using the Amsterdam Checklist for Hearing and Work, *Int J Audiol*, 45(9), s. 503–512. doi: 10.1080/14992020600754583.

Kressner A.A., Jensen-Rico K.M., Kizach J. i in. (2024), A corpus of audio-visual recordings of linguistically balanced, Danish sentences for speech-in-noise experiments,

Speech Communication, 165, 103141. doi: 10.1016/j.specom.2024.103141.

Krueger M., Schulte M., Zokoll M.A., Wagener K.C., Meis M., Brand T., Holube I. (2017), Relation Between Listening Effort and Speech Intelligibility in Noise, *Am J Audiol*, 26(3S), s. 378–392. doi: 10.1044/2017_AJA-16-0136.

Kuchinsky S.E., Razeghi N., Pandža N.B. (2023), Auditory, Lexical, and Multitasking Demands Interactively Impact Listening Effort, *J Speech Lang Hear Res*, 66(10), s. 4066–4082. doi: 10.1044/2023_JSLHR-22-00548.

Kuhl W. (1985), Über versuche zur ermittlung der günstigsten nachhallzeit großer musikstudios (About trying to find the best reverberation time for big music studios), *Acta Acustica United with Acustica*, 4(5), s. 618–634.

Kujawa S.G., Liberman M.C. (2009), Adding insult to injury: cochlear nerve degeneration after "temporary"noise-induced hearing loss, *J Neurosci*, 29(45), s. 14077–14085. doi: 10.1523/JNEUROSCI.2845-09.2009.

Olsen W.O., Noffsinger D., Kurdziel S. (1975), Speech discrimination in quiet and in white noise by patients with peripheral and central lesions, *Acta Otolaryngol*, 180(5–6), s. 375–82. doi: 10.3109/00016487509121339.

Kwak C., Han W. (2021), Age-Related Difficulty of Listening Effort in Elderly, *Int J Environ Res Public Health*, 18(16), 8845. doi: 10.3390/ijerph18168845.

Lang A. (2003), *Analiza krzywych normowych w audiometrii mowy dla języka polskiego*, praca magisterska, UAM, Poznań.

Larsby B., Hällgren M., Lyxell B., Arlinger S. (2005), Cognitive performance and perceived effort in speech processing tasks: effects of different noise backgrounds in normal-hearing and hearing-impaired subjects, *Int J Audiol*, 44(3), s. 131–143. doi: 10.1080/14992020500057244.

Lauterbur P.C. (1973), Image formation by induced local interactions. Examples employing nuclear magnetic resonance, *Nature*, 242, s.190–191. doi: 10.1038/242190a0.

Lavie N. (2005), Distracted and confused?: Selective attention under load, *Trends in Cognitive Sciences*, 9(2), s. 75–82. doi: 10.1016/j.tics.2004.12.004.

Lawrence B.J., Jayakody D.M., Bennett R.J., Eikelboom R.H., Gasson N., Friedland P.L. (2020), Hearing loss and depression in older adults: a systematic review and meta-analysis, *Gerontol*, 60(3), s. e137–e154. doi: 10.1093/geront/gnz009.

Lemke U., Besser J. (2016), Cognitive Load and Listening Effort: Concepts and Age-Related Considerations, *Ear Hear*, 37(1), 77S–84S. doi: 10.1097/AUD.0000000000000304.

Le Rhun L., Llorach G., Delmas T., Suied C., Arnal L.H., Lazard D.S. (2024), A standardised test to evaluate audio-visual speech intelligibility in French, *Heliyon*, 10(2), e24750. doi: 10.1016/j.heliyon.2024.e24750.

Levitt H. (1971), Transformed up–down methods in psychoacoustics, *Journal of the Acoustical Society of America*, 49, s. 467–477.

Li J., Zhang X., Gesang M., Luo B. (2023), The changes of central auditory processing function and its electroencephalogram in patients with

mild cognitive impairment and early Alzheimer's disease, *Heliyon*, 9(5), e15641. doi: 10.1016/j.heliyon.2023.e15641.

Liberman A.M., Mattingly I.G. (1985), The motor theory of speech perception revised, *Cognition*, 21(1), s. 1–36. doi: 10.1016/0010-0277(85)90021-6.

Lieberman P. (2006), *Toward an Evolutionary Biology of Language,* Harvard University Press, Cambridge MA.

Lieberman M.D., Eisenberger N.I., Crockett M.J., Tom S.M., Pfeifer J.H., Way B.M. (2007), Putting Feelings Into Words, *Psychological Science*, 18(5), s. 421–428. doi: 10.1111/j.1467-9280.2007.01916.

Llorach G., Kirschner F., Grimm G., Zokoll M.A., Wagener K.C., Hohmann V. (2022), Development and evaluation of video recordings for the OLSA matrix sentence test, *Int J Audiol*, 61(4), s. 311–321. doi: 10.1080/14992027.2021.1930205.

Luts H., Eneman K., Wouters J., Schulte M., Vormann M., Buechler M., Dillier N., Houben R., Dreschler W.A., Froehlich M., Puder H., Grimm G., Hohmann V., Leijon A., Lombard A., Mauler D., Spriet A. (2010), Multicenter evaluation of signal enhancement algorithms for hearing aids, *J Acoust Soc Am*, 127(3), s.1491–1505. doi: 10.1121/1.3299168.

MacPherson A., Akeroyd M.A. (2014), Variations in the slope of the psychometric functions for speech intelligibility: a systematic survey, *Trends in hearing*, 18, 2331216514537722. doi: 10.1177/2331216514537722.

Mackersie C.L., Cones H. (2011), Subjective and psychophysiological indexes of listening effort in a competing-talker task, *J Am Acad Audiol*, 22, s. 113–122. doi: 10.3766/jaaa.22.2.6.

Makarewicz R., (2023), *Dźwięki i fale* (6 ed.), Wydawnictwo Naukowe UAM, Poznań.

Manchaiah V., Picou E.M., Bailey A., Rodrigo H. (2021), Consumer Ratings of the Most Desirable Hearing Aid Attributes, *J Am Acad Audiol*, 32(8), s. 537–546. doi: 10.1055/s-0041-1732442.

Manning L., Thomas-Antérion C. (2011), Marc Dax and the discovery of the lateralisation of language in the left cerebral hemisphere, *Rev Neurol*, 167(12), s. 868–872.
doi: 10.1016/j.neurol.2010.10.017.

Marshall L.G. (1994), An acoustic measurement program for evaluating auditoriums based on the early/late sound energy ratio, *Journal of Acoustical Society of America*, 96(4), s. 2251–2261. doi: 10.1121/1.410097.

Marslen-Wilson W.D. (1987), Functional parallelism in spoken word-recognition, *Cognition*, 25(1–2), s. 71–102. doi: 10.1016/0010-0277(87)90005-9.

Marrone N., Mason C.R., Kidd G. (2008), Tuning in the spatial dimension: Evidence from a masked speech identification task, *J Acoust Soc Am*, 124, s. 1146–1158. doi:10.1121/1.2945710.

Martin R.L., McAnally K.I., Bolia R.S. (2012), Spatial release from speech-on-speech masking in the median sagittal plane, *J Acoust Soc Am*, 131, s. 378–385. doi: 10.1121/1.3669994.

Martini F., Bartholomew E. (2003), *Essentials of Anatomy and Physiology*, Benjamin Cummings, s. 267.

Mattingly I.G., Liberman A.M., Syrdal A.K. Halwes T.G. (1971), Discrimination in speech and nonspeech modes, *Cognitive Psychology*, 2, s, 131–157. doi: 10.1016/0010-0285(71)90006-5.

Mattys S.L., Davis M.H., Bradlow A.R., Scott S.K. (2012), Speech recognition in adverse conditions: A review, *Language and Cognitive Processes*, 27(7–8), s. 953–978. doi: 10.1080/01690965.2012.705006.

McCormack A., Fortnum H. (2013), Why do people fitted with hearing aids not wear them? *Int J Audiol*, 52(5), s. 360–368. doi: 10.3109/14992027.2013.769066.

McGarrigle R., Munro K.J., Dawes P., Stewart A.J., Moore D.R., Barry J.G., Amitay S. (2014), Listening effort and fatigue: what exactly are we measuring? A British Society of Audiology Cognition in Hearing Special Interest Group 'white paper', *Int J Audiol*, 53(7), s. 433–440. doi: 10.3109/14992027.2014.890296.

McGurk H., MacDonald J. (1976), Hearing lips and seeing voices, *Nature*, 264, s. 746–748. doi: 10.1038/264746a0.

Meister H., Koelewijn T., Başkent D.,Abdellatif K. (2023), Real-time speech segregation and listening effort during speech-on-speech masking - examination based on a visual world paradigm, [w:] *Forum Acusticum 2023 – 10th Convention of the European Acoustics Association, EAA 2023 (Proceedings of Forum Acusticum)*, European Acoustics Association, EAA.

Mener D.J., Betz J., Genther D.J., Chen D., Lin F.R. (2013), Hearing loss and depression in older adults, *J Amer Ger Soc*, 61(9), 16–27. doi: 10.1111/jgs.12429.

Miller G.A. (1947), The masking of speech, *Psychological Bulletin*, 44(2), s. 105–129. doi: 10.1037/h0055960.

Miller G.A., Licklider J.C.R. (1950), The intelligibility of interrupted speech, *Journal of the Acoustical Society of America*, 22, s. 167–173. doi:10.1121/1.1906584.

Mishra J., Anguera J.A., Ziegler D.A., Gazzaley A. (2013), A cognitive framework for understanding and improving interference resolution in the brain, *Prog Brain Res*, 207, s. 351–77. doi: 10.1016/B978-0-444-63327-9.00013-8.

Mizuno K., Tanaka M., Yamaguti K., Kajimoto O., Kuratsune H., Watanabe Y. (2011), Mental fatigue caused by prolonged cognitive load associated with sympathetic hyperactivity, *Behav Brain Funct*, 7, 17. doi: 10.1186/1744-9081-7-17.

Moore B.C.J. (1999), *Wprowadzenie do psychologii słyszenia*, PWN, Poznań.

Moore B.C.J., Glasberg B.R. (1989), Mechanisms underlying the frequency discrimination of pulsed tones and the detection of frequency modulation, *J Acoust Soc Am*, 86(5), s. 1722–1732. doi: 10.1121/1.398603.

Moore J.K., Linthicum F.H. (2007), The human auditory system: A timeline of development, *Int J Audiol*, 46, s. 460–478. doi: 10.1080/14992020701383019.

Moser T., Starr A. (2016), Auditory neuropathy--neural and synaptic mechanisms, *Nat Rev Neurol*, 12(3), s. 135–149. doi: 10.1038/nrneurol.2016.10.

Moulin A., Richard C. (2016), Sources of variability of speech, spatial, and qualities of hearing scale (SSQ) scores in normal-hearing and hearing-impaired populations, *Int J Audiol*, 55(2), s. 101–109. doi: 10.3109/14992027.2015.1104734.

Murman D.L. (2015), The Impact of Age on Cognition, Semin Hear, 36(3), s. 111–121. doi: 10.1055/s-0035-1555115.

Musiek F.E., Gollegly K.M., Kibbe K.S., Verkest-Lenz S.B. (1991), Proposed screening test for central auditory disorders: follow-up on the dichotic digits test, Am J Otol, 12(2), s. 109–113.

Nachtegaal J., Smit J.H., Smits C., Bezemer P.D., van Beek J.H., Festen J.M., Kramer S.E. (2009), The association between hearing status and psychosocial health before the age of 70 years: results from an internet-based national survey on hearing, *Ear Hear,* 30(3), s. 302–312. doi: 10.1097/AUD.0b013e31819c6e01.

Neely S.T., Kim D.O. (1986), A model for active elements in cochlear biomechanics, *The Journal of the Acoustical Society of America*, 79(5), s. 1472–1480. doi: 10.1121/1.393674.

Nguyen D.D., McCabe P., Thomas D. (2021), Acoustic voice characteristics with and without wearing a facemask, *Sci Rep*, 11, 5651. doi: 10.1038/s41598-021-85130-8.

Nilsson M., Soli S.D., Sullivan J.A. (1994), Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise, *J Acoust Soc Am*, 95, s. 1085–1099. doi: 10.1121/1.408469.

Oates P.A., Kurtzberg D., Stapells D.R. (2002), Effects of sensorineural hearing loss on cortical event-related potential and behavioral measures of speech-sound processing, Ear Hear, 23, s. 399–415. doi: 10.1097/00003446-200210000-00002.

Ohlenforst B., Zekveld A.A., Jansma E.P., Wang Y., Naylor G., Lorens A., Lunner T., Kramer S.E. (2017), Effects of hearing impairment and hearing aid amplification on listening effort: A systematic review, *Ear and Hearing*, 38(3), s. 267–281. doi: 10.1097/AUD.0000000000000396.

Olds C., Pollonini L., Abaya H., Larky J., Loy M., Bortfeld H., Beauchamp M.S., Oghalai J.S. (2016), Cortical Activation Patterns Correlate with Speech Understanding After Cochlear Implantation, *Ear Hear*, 37(3), s. 160–172. doi: 10.1097/AUD.0000000000000258.

O'Shaughnessy D. (2000), *Speech communications. Human and machine*, IEEE Press, Nowy Jork, s. 35–107.doi: 10.1109/9780470546475.ch3.

Ozimek E., Kutzner D., Sęk A., Wicher, A. (2009), Polish sentence tests for measuring the intelligibility of speech in interfering noise, *Int J Audiol*, 48(7), s. 433–443. doi: 10.1080/14992020902725521.

Ozimek E, Warzybok A, Kutzner D. (2010), Polish sentence matrix test for speech intelligibility measurement in noise, *Int J Audiol*, 49(6), s. 444–454. doi: 10.3109/14992021003681030.

Ozimek E. (2009), Development and evaluation of Polish digit triplet test for auditory screening, Speech Communication, 2009, 51(4), s. 307–316. doi: 10.1016/j.specom.2008.09.007.

Ozimek E., Kutzner D., Libiszewski P. (2012), Speech intelligibility tested by the Pediatric Matrix Sentence test in 3–6 year old children, *Speech Communication*, 54(10), s. 1121–1131. doi: 10.1016/j.specom.2012.06.001.

Quignard P. (2016), *The Hatred of Music*, Yale University Press, New Haven, s. 71. doi: 10.12987/9780300220940.

Palmiero A.J., Symons D., Morgan J.W., Shaffer R.E. (2016), Speech intelligibility assessment of protective facemasks and air-purifying respirators, *Journal of Occupational and Environmental Hygiene*, 13(12), s. 960–968. doi: 10.1080/15459624.2016.1200723

Parducci A., Perrett L.F. (1971), Category rating scales: Effects of relative spacing and frequency of stimulus values, *Journal of Experimental Psychology*, 89(2), s. 427–452. doi: 10.1037/h0031258.

Pastusiak A., Niemiec D., Kociński J., Warzybok A. (2019), The Benefit of Binaural Hearing Among Listeners with Sensorineural Hearing Loss, *Archives of Acoustics*, 44(4), s. 709–717. doi: 10.24425/aoa.2019.129726.

Pastusiak A., Hafke-Dys H., Kociński J., Szarzyński K., Janeczek K. (2024), Advancing Auscultation Education: Signals Visualization as a Novel Tool for Enhancing Pathological Respiratory Sounds Detection, *Polish Journal of Medical Physics and Engineering*, 30(1),s. 1–10. doi: doi.org/10.2478/pjmpe-2024-0001.

Paul B.T., Bruce I.C., Roberts L.E. (2017), Evidence that hidden hearing loss underlies amplitude modulation encoding deficits in individuals with and without tinnitus, *Hear Res*, 344, s. 170–182.

Peelle J.E. (2018), Listening Effort: How the cognitive consequences of acoustic challenge are reflected in brain and behavior, *Ear and Hearing*, 39, s. 204–214. doi: 10.1097/ aud.0000000000000494.

Peters R.W., Moore B.C., Baer T. (1998), Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people, *J Acoust Soc Am*, 103(1), s. 577–587. doi: 10.1121/1.421128.

Pichora-Fuller M.K, Kramer S.E., Eckert M.A., Edwards B., Hornsby B.W., Humes L.E., Lemke U., Lunner T., Matthen M., Mackersie C.L., Naylor G., Phillips N.A., Richter M., Rudner M., Sommers M.S., Tremblay K.L., Wingfield A. (2016), Hearing Impairment and Cognitive Energy: The Framework for Understanding Effortful Listening (FUEL), *Ear Hear*, 37(1), 5S–27S. doi: 10.1097/AUD.0000000000000312.

Pickles J.O. (1988), *An introduction to the physiology of hearing*, Academic Press, San Diego.

Picou E.M., Gordon J., Ricketts T.A. (2016), The Effects of Noise and Reverberation

on Listening Effort in Adults With Normal Hearing, Ear Hear, 37(1), s. 1–13. doi: 10.1097/AUD.0000000000000222.

Piłka E. (2015), Testy słowne dostępne i wykorzystywane w Polsce w audiometrii mowy – rys historyczny, *Now Audiofonol*, 4(4), s. 67–74. doi: 10.17431/895188.

Pinker S. (1994), *The Language Instinct: How the Mind Creates Language*, William Morrow and Company, Nowy Jork.

Plomp R., Mimpen A.M. (1979), Improving the reliability of testing the speech reception threshold for sentences, *Audiology*, 18(1), s. 43–52. doi: 10.3109/00206097909072618.

Pollack I. (1975), Auditory informational masking, *J Acoust Soc Am,* 57(1), S5. https://doi.org/10.1121/1.1995329.

Pruszewicz A., Surmanowicz-Demenko G., Jastrzębska M. (2011), Polskie testy do badania audiometrią mowy, [w:] *Wybrane zagadnienia z audiometrii mowy*, red. A. Obrębowski, Wydawnictwo Naukowe UM, Poznań.

Puglisi G.E., Warzybok A., Hochmuth S., Visentin C., Astolfi A., Prodi N., Kollmeier B. (2015), An Italian matrix sentence test for the evaluation of speech intelligibility in noise, *Int J Audiol*, 2015, 54(2), s. 44–50. doi: 10.3109/14992027.2015.1061709.

Puglisi G.E., di Berardino F., Montuschi C., Sellami F., Albera A., Zanetti D., Albera R., Astolfi A., Kollmeier B., Warzybok A. (2021), Evaluation of Italian Simplified Matrix Test for Speech-Recognition Measurements in Noise, *Audiol Res*, 11(1), s. 73–88. doi: 10.3390/audiolres11010009.

Rabbitt P. (1968), Channel-capacity, intelligibility and immediate memory, *Q J Exp Psychol*, 20, s. 241–248. doi: 10.1080/14640746808400158.

Rahne T., Wagner T.M., Kopsch A.C., Plontke S.K., Wagner L. (2023), Influence of Age on Speech Recognition in Noise and Hearing Effort in Listeners with Age-Related Hearing Loss, *J Clin Med*, 12(19), 6133. doi: 10.3390/jcm12196133.

Remez R.E., Rubin P.E., Pisani D.B., Carrell T.D. (1981), Speech perception without traditional speech cues, *Science*, 212, s, 947–950. doi: 10.1126/science.7233191.

Rhebergen K.S., Versfeld N.J., Dreschler W.A. (2005), Release from informational masking by time reversal of native and non-native interfering speech, *J Acoust Soc*

Am, 118(3), s. 1274–1277. doi: 10.1121/1.2000751.

Ross L.A., Saint-Amour D., Leavitt V.M., Javitt D.C., Foxe J.J. (2007), Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments, *Cerebral Cortex*, 17, s. 1147–1153. doi: 10.1093/cercor/bhl024.

Rosemann S., Thiel C.M. (2020), Neuroanatomical changes associated with age-related hearing loss and listening effort, *Brain Struct Funct*, 225(9), 2689–2700. doi: 10.1007/s00429-020-02148-w.

Rothman S. (1984), Synaptic release of excitatory amino acid neurotransmitter mediates anoxic neuronal death, *J Neurosci*, 4(7), s. 1884–1891. doi: 10.1523/JNEUROSCI.04-07-01884.1984.

Rudner M., Lunner T., Behrens T., Thorén E.S., Rönnberg J. (2012), Working memory capacity may influence perceived effort during aided speech recognition in noise, *J Am Acad Audiol*, 23, s. 577–589. doi: 10.3766/jaaa.23.7.7

Šabic E., Henning D., Myüz H., Morrow A., Hout M.C., MacDonald J.A. (2020), Examining the Role of Eye Movements During Conversational Listening in Noise, *Front. Psychol*, 11, 200. doi: 10.3389/fpsyg.2020.00200.

Sabine W.C. (1922), *Collected papers on acoustics*, Harvard University Press, Cambridge (MA), s.279.

Sakai H., Setoguchi H., Ando Y. (1998), Individual subjective preference of listeners for vocal music sources in relation to the subsequent reverberation time of sound fields, The *J Acoust Soc Am*, 103(5), 2826. doi: 10.1006/jsvi.1999.269.

Sarampalis A., Kalluri S., Edwards B., Hafter E. (2009), Objective measures of listening effort: effects of background noise and noise reduction, *J Speech Lang Hear Res*, 52(5), s. 1230–1240. doi: 10.1044/1092-4388(2009/08-0111).

Saunders G.H., Jackson I.R., Visram A.S. (2021), Impacts of face coverings on communication: an indirect impact of COVID-19, *Int J Audiol*, 60, s. 495–506. doi: 10.1080/14992027.2020.1851401

Schneider BA, Daneman M, Murphy DR. (2005) Speech comprehension difficulties in older adults: cognitive slowing or age-related changes in hearing? *Psychol Aging*, 20(2), s. 261–271. doi: 10.1037/0882-7974.20.2.261.

Schneider B.A., Li L., Daneman M. (2007), How competing speech interferes with speech comprehension in everyday listening situations, *J Am Acad Audiol,* 18(7), s.

559–572. doi: 10.3766/jaaa.18.7.4.

Schmidt-Nielsen A. (1992), *Intelligibility and Acceptability testing for Speech Technology*, http://www.dtic.mil/dtic/tr/fulltext/u2/a252015.pdf. [dostęp: 06.11.2024]

van Schoonhoven J., Schulte M., Boymans M., Wagener K.C., Dreschler W.A., Kollmeier B. (2016), Selecting Appropriate Tests to Assess the Benefits of Bilateral Amplification With Hearing Aids, *Trends Hear*, 20, 2331216516658239. doi: 10.1177/2331216516658239.

Schreiner C.E., Urbas J.V. (1986), Representation of amplitude modulation in the auditory cortex of the cat. I. The anterior auditory feld (AAF), *Hear Res*, 21(3), s. 227–241. doi: 10.1016/0378-5955(88)90146-3

Seeman S., Sims R. (2015), Comparison of Psychophysiological and Dual-Task Measures of Listening Effort, *J Speech Lang Hear Res*, 58(6), s. 1781–1792. doi: 10.1044/2015_JSLHR-H-14-0180.

Sellick P.M., Patuzzi R., Johnstone B.M. (1982), Measurement of basilar membrane motion in the guinea pig using the Mössbauer technique, *J Acoust Soc Am*, 72(1), s. 131–141. doi: 10.1121/1.387996.

Sergeyenko Y., Lall K., Liberman M.C., Kujawa S.G. (2013), Age-related cochlear synaptopathy: an early-onset contributor to auditory functional decline, *J Neurosci*, 33(34), s. 13686-13694. doi: 10.1523/JNEUROSCI.1783-13.2013.

Shams L., Kamitani Y., Shimojo S. (2000), What you see is what you hear, *Nature*, 408, 788. doi: 10.1038/35048669.

Shannon R.V., Zeng F.G., Kamath V., Wygonski J., Ekelid M. (1995), Speech recognition with primarily temporal cues, *Science*, 270(5234), s. 303–304. doi: 10.1126/science.270.5234.303.

Schiller I.S., Aspöck L., Schlittmeier S.J. (2023), The impact of a speaker's voice quality on auditory perception and cognition: a behavioral and subjective approach, *Front Psychol*,14, 1243–1249. doi: 10.3389/fpsyg.2023.1243249.

Shinn-Cunningham B.G., Kopco N., Martin T.J. (2005), Localizing nearby sound sources in a classroom: binaural room impulse responses, *J Acoust Soc Am*, 117(5), s. 3100–3115. doi: 10.1121/1.1872572.

Shinn-Cunningham B.G., Best V. (2008), Selective attention in normal and impaired

hearing. *Trends Amplif*, 12(4), s. 283–299. doi: 10.1177/1084713808325306.

Sininger Y. (2002), Identification of auditory neuropathy in infants and children, *Semin Hear*, 23(3), s. 193–200. doi: 10.1055/s-2002-34456.

Skrodzka E. (2014), Audiometria mowy, [w:] *Protetyka Słuchu*, red. E. Hojan, Wydawnictwo Naukowe UAM, Poznań, s. 275–308.

Smits C., Kapteyn T.S., Houtgast T. (2004), Development and validation of an automatic speech-in-noise screening test by telephone, *Int J Audiol,* 43(1), s. 15-28. doi: 10.1080/14992020400050004.

Springer F. (2019), *Springer szuka ciszy*, https://www.vogue.pl/a/springer-szuka-ciszy-halas [dostęp: 26.10.2024 r.]

Solhjoo S., Haigney M.C., McBee E. (2019), Heart Rate and Heart Rate Variability Correlate with Clinical Reasoning Performance and Self-Reported Measures of Cognitive Load, *Sci Rep*, 9, 14668. doi: 10.1038/s41598-019-50280-3.

Sommers M.S., Phelps D. (2016), Listening Effort in Younger and Older Adults: A Comparison of Auditory-Only and Auditory-Visual Presentations, *Ear Hear*, 37(1), 62S–8S. doi: 10.1097/AUD.0000000000000322.

Speaks C., Karmen J.L., Benitez L. (1967), Effect of a competing message on synthetic sentence identification, *J Speech Hear Res*, 10(2), s. 390–395. doi: 10.1044/jshr.1002.390.

Starr A., McPherson D., Patterson J., Don M., Luxford W., Shannon R., Sininger Y., Tonakawa L., Waring M. (1991), Absence of both auditory evoked potentials and auditory percepts dependent on timing cues, *Brain*, 114(3), s. 1157–1180. doi: 10.1093/brain/114.3.1157.

Steeneken H.J.M. (1992), Quality Evaluation of Speech Processing Systems, [w:] Digital Speech Processing, red A.N. Ince, T*he Kluwer International Series in Engineering and Computer Science,* vol 155. Springer, Boston, MA, 155. doi: 10.1007/978-1-4757-2148-5_6.

Steinbrink J., Villringer A., Kempf F., Haux D., Boden S., Obrig H. (2006), Illuminating the BOLD signal: Combined fMRI-fNIRS studies, *Magnetic Resonance Imaging*, 24(4), s. 495–505. doi: 10.1016/j.mri.2005.12.034.

Stevens K.N. (2000), Acoustic Phonetics, [w:] *Current Studies in Linguistics*, MIT Press, Cambridge, MA.

Stevens S.S. (1957), On the psychophysical law, *Psychological Review*, 64(3), s. 153–181. doi:10.1037/h0046162.

Stickney G.S., Assmann P.F. (2001), Acoustic and linguistic factors in the perception of bandpass-filtered speech, *J Acoust Soc Am*, 109(3), s. 1157–1165. doi: 10.1121/1.1340643.

Stollman M.H., Kapteyn T.S. (1994), Effect of time scale modification of speech on the speech recognition threshold in noise for elderly listeners, *Audiology*, 33(5), s. 280–290. doi: 10.3109/00206099409071888.

Strand J.F., Brown V.A., Merchant M.B., Brown H.E., Smith J. (2018), Measuring Listening Effort: Convergent Validity, Sensitivity, and Links With Cognitive and Personality Measures, *J Speech Lang Hear Res*, 61(6), s. 1463–1486. doi: 10.1044/2018_JSLHR-H-17-0257.

Strouse A.L., Hall J.W., Burger M.C. (1995), Central auditory processing in Alzheimer's disease, *Ear Hear*, 16, s. 230–238. doi: 10.1097/00003446-199504000-00010.

Sumby W.H., Pollack I. (1954), Visual contribution to speech intelligibility in noise, *J Acoust Soc Am*, 26, s. 212–215. doi: 10.1121/1.1907309.

Świdziński, P., Wiskirska-Woźnica, b., Pruszewicz, A. 2011, Metodologia jakościowych i ilościowych badań słuchu mową, [w:] *Wybrane zagadnienia z audiometrii mowy*, red. A. Obrębowski,Wydawnictwo Uniwersytetu Medycznego, Poznań, s. 56–68.

Tabatabaeizadeh S.A. (2021), Airborne transmission of COVID-19 and the role of face mask to prevent it: a systematic review and meta-analysis, *Eur J Med Res*, 26(1), 1. doi: 10.1186/s40001-020-00475-6.

Takahashi G.A., Bacon S.P. (1992), Modulation detection, modulation masking, and speech understanding in noise in the elderly, *J Speech Hear Res*, 35(6), s. 1410–1421. doi: 10.1044/jshr.3506.1410.

Tchorz J. (2022), Measuring listening effort, [w:] *Techniques of Hearing*, Routledge, s. 47–54.

Tremblay P., Dick A.S. (2016), Broca and Wernicke are dead, or moving past the classic model of language neurobiology, *Brain Lang*, 162, s. 60–71. doi: 10.1016/j.bandl.2016.08.004.

Treutwein B., Strasburger H. (1999), Fitting the psychometric function, *Percept Psychophys*, 61(1), s. 87–106. doi: 10.3758/bf03211951.

Tun P.A. (1998), Fast noisy speech: Age differences in processing rapid speech with background noise, *Psychol Aging,* 13(3), s. 424–434. doi: 10.1037//0882-7974.13.3.424

Tun P.A., O'Kane G., Wingfield A. (2002), Distraction by competing speech in young and older adult listeners, *Psychol.Aging*, 17(3), s. 453–467. doi: 10.1037/0882-7974.17.3.453.

Tyler R.S., Tye-Murray N. (1986), Frequency Resolution Measured by Adaptively Varying the Notchwidth: Results from Normals and Hearing Impaired, [w:] *Auditory Frequency Selectivity*, red. B.C.J. Moore, R.D. Patterson, Springer, Boston, MA, 119, s. 323–330.

Vaughan N.E., Letowski T. (1997), Effects of age, speech rate, and type of test on temporal auditory processing, *J Speech Lang Hear Res*, 40(5), s. 1192–1200. doi: 10.1044/jslhr.4005.1192.

Verghese J., Lipton R.B., Katz M.J. (2003), Leisure activities and the risk of dementia in the elderly, *N Engl J Med*, 348, s. 2508–2516. doi: 10.1056/NEJMoa022252.

Versfeld N.J., Dreschler W.A. (2002), The relationship between the intelligibility of time-compressed speech and speech in noise in young and elderly listeners, *J Acoust Soc Am*, 111(1), s. 401–408. doi: 10.1121/1.1426376.

Vetulani J. (2017), *Mózg: fascynacje, problemy, tajemnice*, Homini, Kraków.

PAU (2007), *Po drogach uczonych, Tom I, Z członkami Polskiej Akademii Umiejętności rozmawia Andrzej M. Kobos*, Kraków, s. 491–524.

Voiers W.D. (1983), Evaluating processed speech using the diagnostic rhyme test, *Speech Technology,* s. 30–39.

Yost W.A. (2007), *Fundamentals of Hearing: An Introduction*, Emerald Group Publishing Limited.

Kollmeier B., Wesselkamp M. (1997), Development and evaluation of a German sentence test for objective and subjective speech intelligibility assessment, *J Acoust Soc Am*, 102(4), s. 2412–2421. doi: 10.1121/1.419624.

Wagener K., Josvassen J.L., Ardenkjaer R. (2003), Design, optimization and

evaluation of a Danish sentence test in noise, *Int J Audiol*, 42(1), s. 10–17. doi: 10.3109/14992020309056080.

Wagener K., Brand T. (2006), The role of silent intervals for sentence intelligibility in fluctuating noise in hearing-impaired listeners, *Int J Audiol*, 45(1), s. 26–33, doi: 10.1080/14992020500243851.

Wan G., Corfas G. (2017), Transient auditory nerve demyelination as a new mechanism for hidden hearing loss, *Nat Commun*, 8, 14487. doi: 10.1038/ncomms14487.

Wang C., Wang Z., Xie B., Shi X., Yang P., Liu L., Qu T., Qin Q., Xing Y., Zhu W., Teipel S.J., Jia J., Zhao G., Li L., Tang Y. (2022), Binaural processing deficit and cognitive impairment in Alzheimer's disease, *Alzheimers Dement*, 18(6), s. 1085–1099. doi: 10.1002/alz.12464.

Wardenga N., Batsoulis C., Wagener K.C., Brand T., Lenarz T., Maier H. (2015), Do you hear the noise? The German matrix sentence test with a fixed noise level in subjects with normal hearing and hearing impairment, *Int J Audiol*, 54(2), s. 71–79. doi: 10.3109/14992027.2015.1079929.

Warren R.M. (1970), Perceptual restoration of missing speech sounds, *Science*, 167(3917), s. 392–393. doi: 10.1126/science.167.3917.392.

Warzybok A., Brand T., Wagener K.C., Kollmeier B. (2015), How much does language proficiency by non-native listeners influence speech audiometric tests in noise?, *Int J Audiol*, 54(2), s. 88–99. doi: 10.3109/14992027.2015.1063715.

Warzybok A., Zhilinskaya E., Goykhburg M., Tavartkiladze G., Kollmeier B., Boboshko M. (2020), Clinical validation of the Russian Matrix test – effect of hearing loss, age, and noise level, *Int J Audiol,* 59(12), s. 930–940. doi: 10.1080/14992027.2020.1806368.

Watson C.S., Kelly W.J., Wroton H.W. (1976), Factors in the discrimination of tonal patterns. II. Selective attention and learning under various levels of stimulus uncertainty, *J Acoust Soc Am*, 60(5), s. 1176–1186. doi: 10.1121/1.381220

Wicher A., Sęk A., Konieczny J. (2005), *Akustyczno-fonetyczne cechy mowy polskiej*, Instytut Akustyki UAM, Poznań.

Willberg T., Sivonen V., Linder P., Dietz A. (2021), Comparing the Speech Perception of Cochlear Implant Users with Three Different Finnish Speech Intelligibility Tests in Noise, *J Clin Med*,10(16), 3666. doi: 10.3390/jcm10163666.

Wojnowski W., Obrębowski A., Pruszewicz A., Demenko G., Wiskirska-Woźnica B. (2006), Testy utrudnione w audiometrii mowy dla języka polskiego. Część I: Przygotowanie materiału testowego, *Otolaryngol Pol*, 60(3), s. 385–390.

Wood N.L., Cowan N. (1995), The cocktail party phenomenon revisited: attention and memory in the classic selective listening procedure of Cherry, *J Exp Psychol Gen*, 124(3), s. 243–262.

Woods T.M., Recanzone G.H. (2004), Visually Induced Plasticity of Auditory Spatial Perception in Macaques, *Current Biology*, 14(17), s. 1559–1564. doi: 10.1016/j.cub.2004.08.059).

Zekveld A.A., Kramer S.E., Festen J.M. (2010), Pupil response as an indication of effortful listening: the influence of sentence intelligibility, *Ear Hear*, 31(4), s. 480–490. doi: 10.1097/AUD.0b013e3181d4f251.

Zekveld A.A., Kramer S.E., Festen J.M. (2011), Cognitive load during speech perception in noise: the influence of age, hearing loss, and cognition on the pupil response, *Ear Hear*, 32(4), s. 498–510. doi: 10.1097/AUD.0b013e31820512bb.

Zokaei N., Husain M. (2019), Working Memory in Alzheimer's Disease and Parkinson's Disease, *Curr Top Behav Neurosci*, 41, s. 325–344. doi: 10.1007/7854_2019_103.

Zokoll M.A., Wagener K.C., Kollmeier B. (2017), Diagnosing and Screening in a Minority Language: A Validation Study, *Am J Audiol*, 26(3S), s. 369–372. doi: 10.1044/2017_AJA-16-0138.