

Przesłano do redakcji:
03.27.2025
Zaakceptowano po recenzji:
01.07.2025
Opublikowano:
17.09.2025

Wkład autorów:
A Projekt badania
B Gromadzenie danych
C Analiza danych
D Interpretacja danych
E Przygotowanie pracy
F Przegląd literatury
G Gromadzenie funduszy

Skuteczność zdalnego treningu odczytywania mowy z ust u dorosłych

Effectiveness of remote lipreading training in adults

Martyna Bryłka^{1A-F} , Joanna Beck^{1ABDE} , Hanna Cygan^{1AD-F} 

Instytut Fizjologii i Patologii Słuchu, Światowe Centrum Słuchu, Naukowe Centrum Obrazowania Biomedycznego, Warszawa/Kajetany

Streszczenie

Wprowadzenie: Odczytywanie mowy z ust jest istotnym elementem wspierającym percepcję mowy, szczególnie w warunkach zakłóconego sygnału słuchowego lub u osób z ubytkiem słuchu. Pomimo udowodnionego znaczenia odczytywania mowy z ust tworzenie efektywnych treningów nadal pozostaje przedmiotem badań. Celem niniejszego projektu było sprawdzenie efektywności zdalnego komputerowego programu treningowego w poprawie umiejętności odczytywania mowy z ust u dorosłych.

Materiał i metody: W badaniu wzięło udział 64 dorosłych uczestników ze słuchem w normie (31 mężczyźn, 33 kobiety), którzy zrealizowali miesięczny program treningowy obejmujący sześć sesji online oraz jedną stacjonarną sesję wprowadzającą. Przed rozpoczęciem treningu i po jego zakończeniu przeprowadzono test oceniający rozumienie słów, zdań i wypowiedzi prezentowanych w formie nagrań wideo pozbawionych dźwięku. Zastosowano analizę wariancji z powtarzanymi pomiarami, uwzględniając poziom złożoności materiału językowego oraz płeć uczestników.

Wyniki: Trening znacząco poprawił ogólną poprawność odczytywania mowy z ust ($F(1,62) = 48,18, p < .001, \eta^2_p = 0,437$). Największy wzrost zaobserwowano w rozumieniu zdań (średni wzrost o 15,7 punktu procentowego), mniejszy w przypadku wypowiedzi (11,5 p.p.) i słów (7,6 p.p.). Kobiety osiągały ogólnie lepsze wyniki niż mężczyźni ($F(1,62) = 7,08, p = .01$), ale nie stwierdzono istotnych interakcji efektów treningu z płcią.

Wnioski: Zdalny trening odczytywania mowy z ust jest skuteczną metodą poprawy percepcji wizualnej mowy u dorosłych. Uzyskane wyniki wskazują na istotną rolę zarówno procesów *bottom-up*, jak i *top-down*, szczególnie przy przetwarzaniu materiału bardziej złożonego. Wyniki mają potencjał aplikacyjny w terapii osób z ubytkiem słuchu oraz w sytuacjach komunikacyjnych o ograniczonej dostępności dźwięku.

Słowa kluczowe: odczytywanie mowy z ust • percepcja mowy • trening zdalny • wizualna percepcja, niedosłuch

Abstract

Introduction: Lipreading is a crucial skill that supports speech perception, particularly in noisy environments or for individuals with hearing loss. Despite its relevance, the effectiveness of lipreading training remains under investigation. This study aimed to assess the impact of a remote, computer-based training program on adult lipreading abilities.

Material and methods: Sixty-four adult participants (31 men, 33 women) with normal hearing completed a month-long training consisting of six online sessions and one introductory session. Lipreading skills were assessed before and after the training using a test comprising isolated words, sentences, and short narratives presented as silent video recordings. A repeated measures analysis of variance was conducted, taking into account the linguistic complexity of the material and participants' gender.

Results: Training significantly improved overall lipreading accuracy ($F(1,62) = 48.18, p < .001, \eta^2_p = 0.437$). The greatest improvement was observed in sentence comprehension (+15.7 percentage points), followed by narratives (+11.5 p.p.) and single words (+7.6 p.p.). Women performed better overall than men ($F(1,62) = 7.08, p = .01$), with no significant interactions between time and gender.

Conclusions: Remote lipreading training is an effective tool for enhancing visual speech perception in adults. The observed gains highlight the combined influence of bottom-up and top-down mechanisms, particularly when processing complex linguistic material. These findings have practical implications for auditory rehabilitation and communication in sound-limited contexts.

Keywords: lipreading • speech perception • remote training • visual perception • hearing loss

Autor korespondencyjny: Martyna Bryłka, Naukowe Centrum Obrazowania Biomedycznego, Światowe Centrum Słuchu, Instytut Fizjologii i Patologii Słuchu, ul. Mochanckiego 10, 02-042 Warszawa; email: m.brylka@ifps.org.pl

Wprowadzenie

W niniejszym artykule przyjmuje się rozróżnienie zaproponowane przez Kurcz [1], inspirowane koncepcją Chomskiego [2], zgodnie z którym język to system znaków i reguł ich użycia, natomiast mowa to konkretny akt posługiwania się tym systemem przez człowieka w sytuacji komunikacyjnej. Zgodnie z ujęciem Chomskiego język jako system formalny składa się z trzech głównych komponentów: składnika syntaktycznego, semantycznego oraz fonologicznego. Struktury te mają swoje odpowiedniki w systemie poznawczym człowieka w postaci analizatorów językowych, które wspólnie tworzą kompetencję językową – wewnętrzną reprezentację wiedzy o języku. Analizator fonologiczny odpowiada za przetwarzanie dźwięków mowy, analizator leksykalny (tzw. słownik umysłowy) zarządza wiedzą o słowach i ich znaczeniach, natomiast analizator syntaktyczny umożliwia rozpoznawanie i przetwarzanie struktur gramatycznych zdań. Percepcja mowy odnosi się w tym ujęciu do procesów rozpoznawania i interpretowania sygnałów mowy odbieranych drogą słuchową lub wzrokową (np. ruchy artykulacyjne). Obejmuje ona zarówno identyfikację dźwięków mowy (fonemów), jak i wyższe poziomy przetwarzania, prowadzące do rozumienia słów i zdań [3]. Przyswajanie języka mówionego to proces rozwoju kompetencji językowej w naturalnym środowisku, który obejmuje zarówno rozwój rozumienia, jak i produkcji mowy z uwzględnieniem modalności słuchowej i wzrokowej [4,5].

Język mówiony jest przede wszystkim przekazywany za pomocą sygnałów dźwiękowych, jednak widoczne wskazówki towarzyszące artykulacji odgrywają istotną rolę w kształtowaniu percepcji mowy. Umiejętność pozyskiwania informacji werbalnych z ruchów artykulacyjnych aparatu mowy wykorzystujemy na poziomie nieświadomym przez całe życie. Już między drugim a trzecim miesiącem życia niemowlęta zaczynają angażować się w interakcje twarzą w twarz z opiekunem [6,7]. Kontakt wzrokowy nie jest przy tym jedynie odruchem, lecz stanowi celową formę komunikacji społecznej, opartą na wzajemnym dopasowywaniu sygnałów – wzroku, mimiki i wokalizacji [8]. W kolejnych miesiącach życia obserwuje się stopniowe przesuwanie uwagi wzrokowej niemowląt z okolic oczu na usta osoby mówiącej. Intensywność tego procesu wzrasta znacząco w trakcie pierwszego roku życia dziecka, przy czym zmiana ta jest zauważalna już w pierwszym półroczu i utrzymuje się stabilnie przez kolejne miesiące [9]. Zwiększona koncentracja niemowląt na ustach rozmówcy odgrywa istotną rolę w dalszym przyswajaniu języka oraz w kształtowaniu się percepcji mowy. Chociaż percepcja mowy opiera się przede wszystkim na modalności słuchowej, modalność wzrokowa stanowi istotne wsparcie w rozwoju języka mówionego, zwłaszcza w pierwszym roku życia dziecka [10,11]. Styczek [10] w swojej pracy konstatuje, że największe znaczenie w procesie przyswajania mowy ma słuch, ale rozwój mowy wymaga również wielokrotnego kontaktu dziecka z wizualnym obrazem artykulacji oraz z samym desygnatem znaczeniowym danego wyrazu. Brak możliwości obserwacji ułożenia i ruchów narządów artykulacyjnych może być źródłem problemów w opanowaniu właściwej realizacji głosek [12]. Jak pokazują badania nad dziećmi z prawidłowym słuchem, lecz niewidomymi od urodzenia, rozwój językowy tej grupy

przebiega wolniej niż u widzących rówieśników [13], co może wynikać z trudności w przyswajaniu fonemów trudnych do odróżnienia wyłącznie słuchowo, lecz łatwych do zidentyfikowania na podstawie informacji wzrokowych, takich jak /m/ i /n/ [14].

Integracja sygnałów wizualnych i słuchowych jest zatem naturalnym mechanizmem wspierającym percepcję i przyswajanie języka mówionego, w tym jego aspektów fonologicznych [15]. Artykulatory – przede wszystkim wargi, zęby i język – przekazują istotne informacje dotyczące miejsca artykulacji spółgłosek i samogłosek. Choć ruchy ust wspierają percepcję mowy, zakres możliwych do uchwycenia informacji wizualnych jest ograniczony – szacuje się, że jednoznacznie widocznych jest zaledwie ok. 30% fonemów języka polskiego [16]. Co więcej, wiele głosek dzieli wspólne cechy wizualne i tworzy tzw. kinem (w literaturze międzynarodowej używa się sformułowania *visem*), czyli klasę fonemów nieodróżnialnych wzrokowo mimo ich akustycznej odmienności. Klasycznym przykładem są grupy: /p/, /b/ i /m/ czy /v/ i /f/ [16, 17]. Zjawisko to znacząco ogranicza jednoznaczność sygnału wizualnego i sprawia, że samo patrzenie na usta rozmówcy nie pozwala na pełną rekonstrukcję przekazu językowego [18]. W praktyce percepcja wizualna mowy opiera się więc nie tylko na samych danych wzrokowych, lecz także na aktywnym wykorzystywaniu kontekstów: językowego, pragmatycznego i sytuacyjnego [16]. Słuchacz (lub obserwator) posługuje się wiedzą o świecie, przewidywaniami semantycznymi oraz strukturą wypowiedzi, aby uzupełniać brakujące lub niejednoznaczne informacje fonetyczne [3,19]. Skuteczność tego procesu wzrasta, gdy wypowiedź ma przewidywalną strukturę, a temat rozmowy jest znany [18].

Mimo tych ograniczeń wizualna informacja artykulacyjna pozostaje niezwykle cennym źródłem danych wspierających rozumienie mowy. Integracja informacji wzrokowych z ruchu warg z bodźcami słuchowymi umożliwia osobom z prawidłowym słuchem dokładniejszą percepcję mowy, nawet gdy sygnał akustyczny pozostaje nienaruszony [20]. Badania pokazują, że zarówno dorośli, jak i dzieci wykorzystują informacje wizualne w sposób zautomatyzowany i często nieuświadomiony, co zwiększa ich precyzję percepcji mowy [21,22]. Gijbels i wsp. [23], badając poziom rozumienia mowy przez dzieci w wieku 4–15 lat w warunkach hałasu, wykazali, że w momencie prezentacji filmu wideo z lektorem, którego mowa była zsynchronizowana z dźwiękiem, dzieci osiągały lepsze wyniki w rozumieniu przekazu. Co istotne, nawet najmłodsze dzieci czerpały wymierne korzyści z obecności wskazówek wizualnych, a efekt ten pozostawał stabilny na w dalszych etapach rozwoju.

Znaczenie percepcji wizualnej w komunikacji jest szczególnie istotne dla osób z ubytkiem słuchu. Szacuje się, że prawie jedna na pięć osób na świecie doświadcza niedosłuchu, a około 5% wymaga specjalistycznej rehabilitacji [24]. Zdecydowana większość z nich to osoby starsze, u których utrata słuchu prowadzi do ograniczenia relacji społecznych, zmniejszenia aktywności związanej z zainteresowaniami oraz pogłębiającego się poczucia izolacji [25,26]. W tej grupie wiekowej obserwuje się jednak, że możliwość wspomagania percepcji słuchowej

obserwacją ruchu warg znacząco poprawia efektywność komunikacji, co z kolei wspiera ich aktywność społeczną i samodzielność [27,28]. Dzieje się tak dlatego, że nawet w przypadku znacznych ubytków słuchu minimalne informacje dźwiękowe, jak np. sam sygnał rozpoczęcia mowy, mogą wspierać analizator wzrokowy i poprawiać rozumienie przekazu mówionego [16]. Warto podkreślić, że wiele osób starszych rozwija umiejętność odczytywania mowy z ust spontanicznie, nie zawsze zdając sobie z tego sprawę, co wskazuje na jej intuicyjny i adaptacyjny charakter [28].

Odczytywanie mowy z ust to umiejętność, która może odgrywać podstawową rolę nie tylko w przypadku osób z ubytkiem słuchu, lecz także w trudnych warunkach akustycznych, w których samo słyszenie – nawet przy użyciu aparatów słuchowych – nie jest wystarczające [29,30]. Trening odczytywania mowy z ust może przyczynić się do lepszego rozumienia wypowiedzi, gdy sygnały dźwiękowe są zniekształcone, niewyraźne lub całkowicie niedostępne [31]. Co istotne, łączenie informacji wzrokowych i słuchowych pozwala maksymalizować efektywność rozpoznawania mowy, i to niezależnie od wieku czy rodzaju niedosłuchu [30]. Jak pokazują badania Ewing [32] oraz van Udena [33], osoby z głębokim ubytkiem słuchu osiągały istotnie wyższe wyniki w rozumieniu przekazu, gdy mogły korzystać zarówno ze wzroku, jak i ze szczątkowych informacji słuchowych. Przykładowo dzieci z ubytkiem słuchu na poziomie 100–109 dB rozumiały przekaz wzrokowo w 81%, ale przy wzrokowo-słuchowym odbiorze – już w 94%. Podobnie u dzieci z jeszcze większym ubytkiem (brak reakcji słuchowej) – skuteczność rozumienia mowy wzrastała z 53% do 69% [32,33]. Wyniki te dowodzą, że nawet minimalne wsparcie słuchowe może istotnie poprawić efektywność komunikacji opartej na czytaniu z ust.

Wprowadzenie treningu odczytywania mowy z ust może w efekcie znacząco poprawić rozumienie mowy u osób niedosłyszących [18,34], a także zwiększyć efektywność rozpoznawania wypowiedzi w hałasie u osób ze słuchem w normie [35,36]. Badania pokazują, że skuteczność tej treningu odczytywania mowy z ust zależy od wielu czynników: od precyzji artykulacji mówcy i wyrazistości mimiki twarzy, przez indywidualne predyspozycje osoby uczestniczącej w treningu, aż po warunki środowiskowe i jakość technologii wspomagającej komunikację [31]. Chociaż starsze osoby zazwyczaj mają większe trudności z odczytywaniem mowy z ust niż osoby młodsze [37,38], to jednak dostęp do bodźców wzrokowych może istotnie zmniejszyć różnice w rozumieniu mowy między grupami wiekowymi [37,39]. Poprawa zdolności rozumienia mowy przekłada się bezpośrednio na lepszą integrację społeczną, większe możliwości edukacyjne oraz ogólną jakość życia. Dla osób głuchych i niedosłyszących odczytywanie mowy z ust stanowi jedno z podstawowych narzędzi wspierających odbiór języka mówionego [40].

W badaniach przeprowadzonych przez Domagałę-Zyśk [41], obejmujących 124 gimnazjalistów z kilkunastu ośrodków specjalnych w Polsce, aż 24% uczniów z ubytkiem słuchu zadeklarowało, że posługuje się wyłącznie mową oraz odczytywaniem z ust, nie używając języka migowego. Kolejne 34% uczniów korzystało z mowy, odczytywania z ust oraz języka migowego lub systemu językowo-migowego jednocześnie, co oznacza, że niemal

60% badanych uczniów wykorzystywało odczytywanie mowy z ust jako jeden z głównych sposobów komunikacji [41]. Umiejętność ta umożliwia bardziej efektywne przetwarzanie informacji z otoczenia i zwiększa poziom niezależności [40]. W praktyce klinicznej i logopedycznej trening odczytywania mowy z ust stanowi ważny element rehabilitacji, szczególnie u osób z trudnościami w komunikacji. Dzięki rozwijaniu zdolności interpretowania wizualnych aspektów mowy uczestnicy takich programów zyskują większą pewność siebie i skuteczność w interakcjach społecznych.

Odpowiednio zaprojektowany trening odczytywania mowy z ust może zatem stanowić skuteczne narzędzie wspierające percepcję wizualną mowy, szczególnie w sytuacjach, w których sygnał dźwiękowy jest niedostępny lub ograniczony. W przeprowadzonym badaniu podłużnym sprawdzono skuteczność opracowanego przez nas treningu odczytywania mowy z ust, któremu poddano grupę zdrowych dorosłych. Założono, że trening wpłynie na poprawę umiejętności identyfikacji pojedynczych słów oraz rozumienia przekazu zawartego w niemych wypowiedziach.

Materiał i metody

Grupa badana

W badaniu wzięło udział 64 uczestników w wieku od 19 do 39 lat (średni wiek = 26,9; $SD = 4,4$). Ze względu na niejednoznaczne wyniki wcześniejszych badań dotyczących różnic w umiejętności odczytywania mowy z ust pomiędzy kobietami a mężczyznami, zadbano o równoliczną reprezentację obu płci (31 mężczyźni i 33 kobiety). Analizy wykazały, że grupy nie różniły się istotnie pod względem wieku. Wszyscy uczestnicy badania byli praworęczni, posiadali co najmniej średnie wykształcenie oraz prawidłowy słuch. Kryteriami włączenia były: ogólny dobry stan zdrowia, brak zgłaszanych zaburzeń neurologicznych i psychicznych oraz brak przyjmowania leków wpływających na funkcjonowanie ośrodkowego układu nerwowego. Badana grupa obejmowała osoby uczestniczące w projekcie z wykorzystaniem protokołu *neurofeedback* funkcjonalnego rezonansu magnetycznego (ang. *functional magnetic resonance imaging*, fMRI), realizowanym w Światowym Centrum Słuchu IFPS w Kajetanach w okresie od lipca 2023 do marca 2024.

Badanie zostało przeprowadzone w ramach grantu NCN SYMFONIA 4 (nr 2016/20/W/NZ4/00354) i uzyskało pozytywną opinię Komisji ds. Etyki Badań Naukowych Uniwersytetu Mikołaja Kopernika (numer wniosku: 22/2023/FT), zgodnie z wytycznymi Deklaracji helsińskiej. Wszyscy uczestnicy wyrazili pisemną, świadomą zgodę na udział w badaniu.

Procedura

W ramach badania uczestnicy realizowali zdalny komputerowy program treningowy, poprzedzony jedną stacjonarną sesją wprowadzającą przeprowadzoną przez profesjonalną trenerkę. Trening polegał na ćwiczeniu umiejętności rozumienia wypowiedzi o różnym stopniu złożoności, które były pozbawione dźwięku. Składał się z sześciu sesji o stopniowo rosnącym poziomie trudności. Każda sesja

wymagała od badanych identyfikacji słów i znaczenia wypowiedzi na podstawie ruchu warg lektora prezentowanego na ekranie. Założono, że systematyczna ekspozycja na zadania o rosnącej trudności wpłynie na poprawę zdolności uczestników do efektywnego odczytywania mowy z ust. Wszystkie nagrania wideo – zarówno te wykorzystane w testach, jak i w trakcie treningu i ekspozycji pasywnej – zawierały wypowiedzi tej samej lektorki, co zapewniało pełną spójność artykulacyjną i eliminowało zmienność źródła mowy. Procedury treningowe, a także testy mierzące poziom przed- i potreningowy zostały przygotowane w programie PsychoPy [42], natomiast treningi online realizowano za pośrednictwem platformy Pavlovia.org, umożliwiającej zdalne prowadzenie eksperymentów psychologicznych zbudowanych w PsychoPy.

Komputerowy program treningowy

W ramach badania uczestnicy realizowali zdalny komputerowy trening odczytywania mowy z ust, obejmujący sześć 15-minutowych sesji treningowych, poprzedzonych jedną sesją stacjonarną z udziałem profesjonalnej trenerki. Celem treningu było rozwijanie zdolności rozpoznawania słów, zdań i sensu wypowiedzi wyłącznie na podstawie obserwacji ruchów artykulacyjnych ust, bez dostępu do informacji dźwiękowej.

W trakcie sesji wprowadzającej trenerka prezentowała uczestnikom przykładowe nagrania, wskazując istotne cechy ruchów artykulacyjnych twarzy i ich znaczenie dla różnicowania dźwięków mowy w języku polskim.

Program treningowy online został zaprojektowany w sposób progresywny – wraz z kolejnymi sesjami wzrastała złożoność językowa materiału oraz poziom trudności zadań. Struktura każdej sesji była zorganizowana według stałego schematu i obejmowała trzy etapy: 1) prezentację materiału z dźwiękiem, 2) prezentację z napisami oraz 3) zadania bez żadnego wsparcia – wyłącznie z obrazem. Uczestnicy ćwiczyli rozpoznawanie słów, zdań i wypowiedzi, wybierając poprawne odpowiedzi z listy lub wpisując je samodzielnie. Od pierwszej do drugiej sesji trening obejmował proste listy słów należące do czterech kategorii tematycznych (*jedzenie, moda, sport, pogoda*). W trzeciej i czwartej sesji uczestnicy pracowali ze zdaniami dłuższymi i bardziej złożonymi. W piątej i szóstej sesji wykorzystywano już krótkie wypowiedzi narracyjne, składające się z dwóch lub trzech zdań.

Każda sesja składała się z kilku bloków ćwiczeń obejmujących:

- rozpoznawanie słów lub fragmentów wypowiedzi (poprzez wybór poprawnej odpowiedzi z listy lub wpisanie jej samodzielnie na klawiaturze komputera);
- informację zwrotną – po każdej odpowiedzi uczestnik otrzymywał informację o poprawności;
- naukę na błędach – w wariantach bez wsparcia po udzieleniu odpowiedzi uczestnik mógł ponownie obejrzeć nagranie, tym razem z dźwiękiem lub napisami, aby porównać swoją odpowiedź i wyciągnąć wnioski (**rycina 1**).

Aby zwiększyć motywację uczestników do systematycznej pracy, wprowadzono system punktacji oparty na liczbie poprawnych odpowiedzi. W każdej sesji uczestnicy mogli

zdobywać punkty zarówno za wybór poprawnych wariantów odpowiedzi, jak i za samodzielne wpisywanie odpowiedzi oraz za poprawne wskazanie jej tematyki. Na podstawie liczby zdobytych punktów przyznawano symboliczne nagrody finansowe (od 1 zł do 10 zł za sesję treningową w zależności od poziomu wykonania treningu), co dodatkowo zwiększało zaangażowanie w trening. Trening, z wyjątkiem sesji wprowadzającej, był realizowany w całości w warunkach domowych, w dogodnym dla uczestników czasie. Taka forma umożliwiała systematyczne ćwiczenie percepcji wizualnej mowy w naturalnym rytmie, z zachowaniem standaryzacji materiałów i procedur.

Oprócz dwóch treningowych sesji online realizowanych każdego tygodnia, uczestnicy byli również eksponowani na materiał językowy dotyczący odczytywania mowy z ust w ramach odrębnej procedury w postaci pasywnej prezentacji materiału wideo bez dźwięku, będącej częścią protokołu *neurofeedback* fMRI. Każda z takich ekspozycji trwała około 25 minut i obejmowała zróżnicowane nagrania przedstawiające osoby wypowiadające słowa, zdania i dłuższe wypowiedzi. Uczestnicy nie udzielali w tym czasie żadnych odpowiedzi ani nie otrzymywali informacji zwrotnej. Choć ekspozycja stanowiła element protokołu *neurofeedback*, należy zaznaczyć, że pełniła również funkcję poznawczą – jej celem było utrwalanie wzorców artykulacyjnych oraz wspieranie percepcji wizualnej mowy w sposób naturalny i niezwiązany z oceną wykonania.

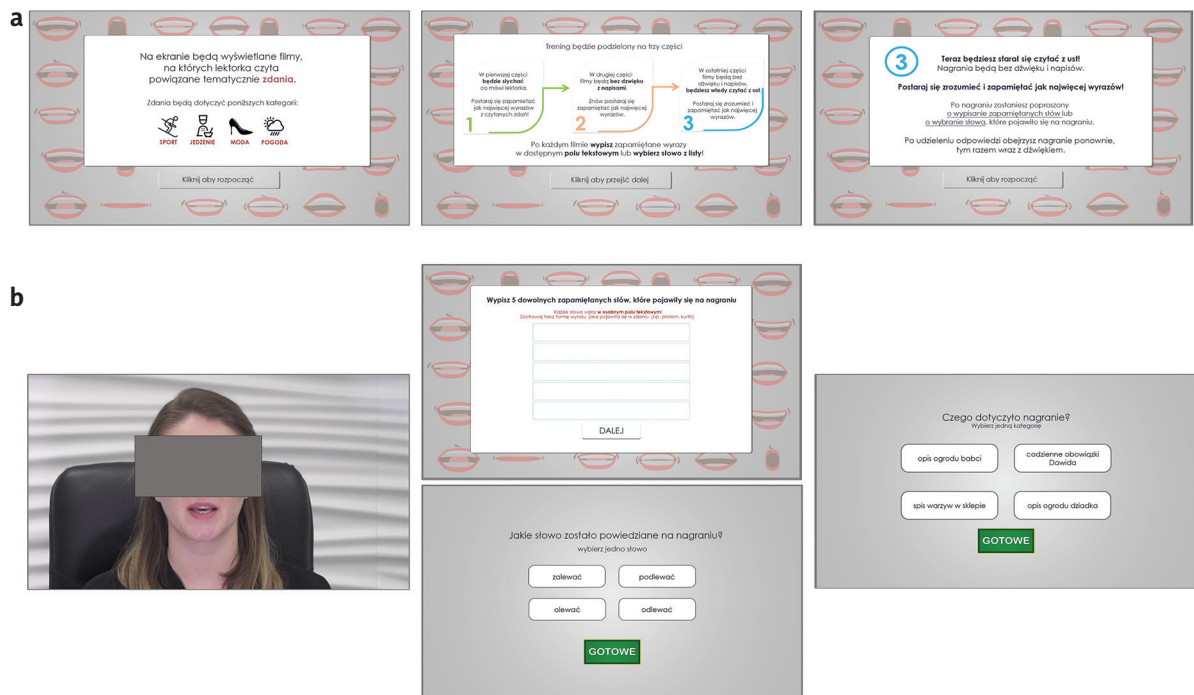
Ocena umiejętności odczytywania mowy z ust

Test oceniający umiejętność odczytywania mowy z ust, opracowany przez psychologów i logopedów na potrzeby badania, przeprowadzony był dwukrotnie – przed rozpoczęciem treningu i po jego zakończeniu. Każda wersja testu zawierała inny zestaw materiałów językowych, jednak struktura zadań pozostała identyczna, co pozwoliło na rzetelne porównanie wyników. Wszystkie bodźce prezentowano w formie nagrań wideo bez dźwięku, w których lektorka artykułowała treści językowe w sposób naturalny, wyraźny i jednoznaczny artykulacyjnie (**rycina 2**).

Test składał się z trzech typów zadań: 1) rozpoznawania pojedynczych wyrazów, 2) rozumienia zdań oraz 3) rozumienia dłuższych wypowiedzi.

W części pierwszej uczestnicy oglądali krótkie filmy, w których lektorka wypowiadała pojedyncze wyrazy (**rycina 2b**). Po każdej prezentacji należało wskazać poprawną odpowiedź spośród czterech możliwości. Każdy zestaw odpowiedzi zawierał jeden wyraz zgodny z wypowiedzianym materiałem oraz trzy dystraktory (podobne, ale błędne odpowiedzi), dobrane na podstawie podobieństwa fonologicznego, artykulacyjnego lub rytmicznego do wyrazu docelowego. Poprawne odpowiedzi należały do jednej z czterech kategorii semantycznych: *jedzenie, moda, sport* lub *pogoda*. Takie zestawienie umożliwiało precyzyjną ocenę umiejętności odczytywania mowy z ruchu warg i eliminowało ryzyko przypadkowego trafienia.

W drugiej części testu uczestnicy oglądali nagrania przedstawiające całe zdania wypowiadane bez dźwięku (**rycina 2c**). Po każdej prezentacji należało wybrać dwa wyrazy spośród sześciu zaproponowanych, które



Opis: **Panel a** – ekrany instruktażowe wprowadzające uczestników w strukturę treningu. Uczestnikom przedstawiono tematykę ćwiczeń (cztery kategorie: *jedzenie, moda, sport, pogoda*) oraz etapy treningu. Każda sesja składała się z trzech części: 1) prezentacji nagrania z dźwiękiem, 2) prezentacji z napisami, 3) prezentacji bez dźwięku i napisów (odczytywanie mowy z ust). Instrukcje zawierały także informacje o konieczności zapamiętywania słów i rozumienia wypowiedzi na podstawie obserwacji artykulacji. **Panel b** – przykładowe etapy zadaniowe w trakcie sesji treningowej. Po obejrzeniu każdego nagrania – niezależnie od etapu sesji – uczestnicy wykonywali zawsze ten sam zestaw zadań: wpisywali pięć zapamiętanych słów lub wybierali jedno słowo, które pojawiło się w nagraniu spośród czterech opcji oraz wskazywali tematykę wypowiedzi, wybierając jedną z czterech kategorii. W ostatniej części każdej sesji (nagrania bez dźwięku i napisów) uczestnicy po udzieleniu odpowiedzi otrzymywali informację zwrotną w postaci ponownego odtworzenia nagrania z dźwiękiem, co miało wspierać proces uczenia się na błędach i rozwijanie percepcji wizualnej mowy.

Rycina 1. Procedura treningowa odczytywania mowy z ust

Figure 1. Lipreading training procedure

faktycznie wystąpiły w danym zdaniu. Dystraktory zostały tak dobrane, by były artykulacyjnie lub fonologicznie zbliżone do wyrazów poprawnych, np. miały podobny układ ust, liczbę sylab lub rytm. Pozwoliło to na ocenę umiejętności całościowego rozpoznawania wypowiedzi w kontekście oraz zdolności różnicowania podobnych ruchów artykulacyjnych.

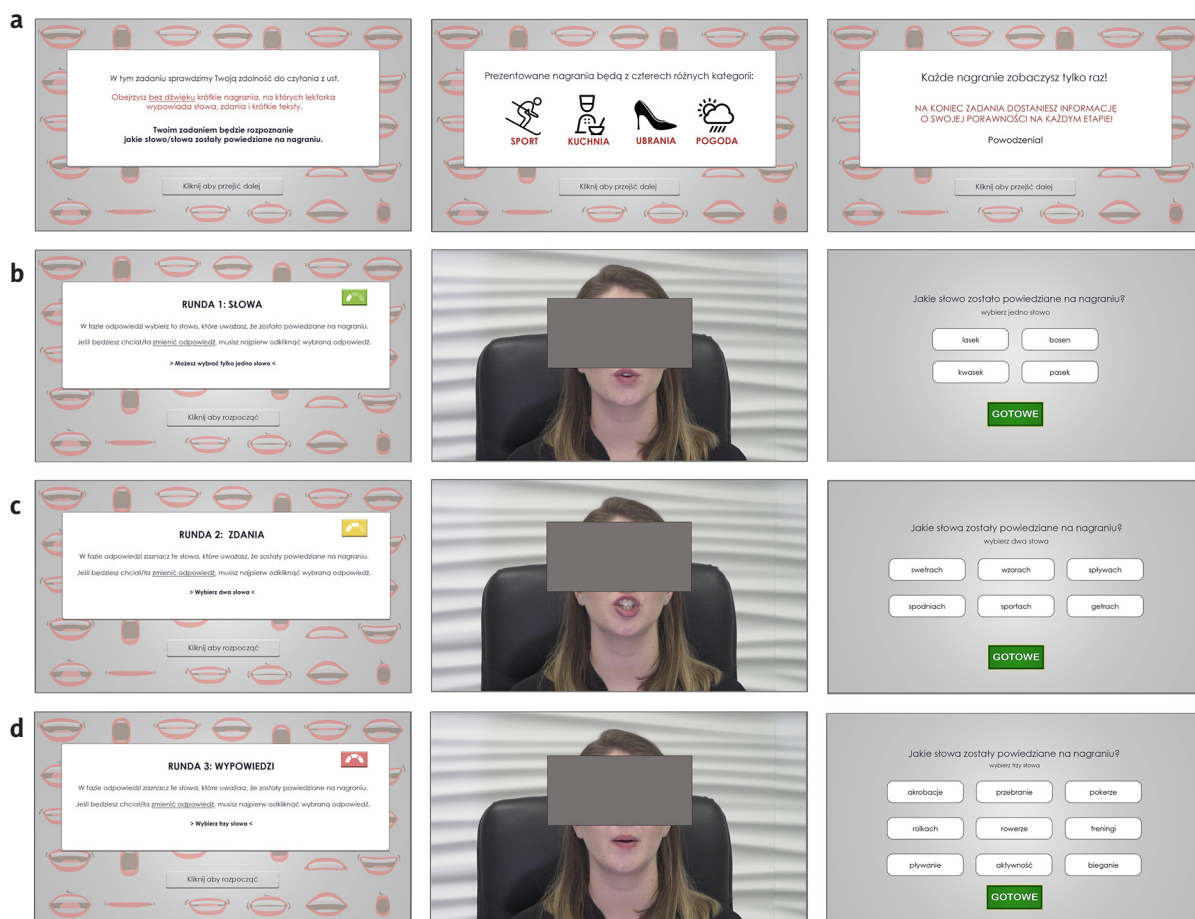
W trzeciej części testu uczestnicy oglądali dłuższe, kilkunastosekundowe wypowiedzi lektorki, również pozbawione dźwięku (**rycina 2d**). Po każdej prezentacji badani wybierali trzy wyrazy z listy dziesięciu propozycji, które rzeczywiście pojawiły się w nagraniu. Dystraktory – podobnie jak w poprzednich częściach – były starannie dobrane pod względem wizualnym i fonologicznym, a jednocześnie nie występowały w wypowiedzi. Wypowiedzi obejmowały różnorodne treści językowe należące do tych samych czterech kategorii semantycznych (*jedzenie, moda, sport, pogoda*), co zapewniało spójność tematyczną i umożliwiało analizę wyników w odniesieniu do rodzaju materiału.

Tak skonstruowany test umożliwił ocenę zarówno podstawowych umiejętności rozpoznawania pojedynczych słów, jak i bardziej złożonych zdolności związanych z rozumieniem kontekstu i większych jednostek językowych na podstawie wyłącznie informacji wzrokowych.

Dzięki zastosowaniu różnych typów zadań możliwe było uzyskanie wieloaspektowego obrazu kompetencji w zakresie odczytywania mowy z ust. Procentowa poprawność udzielanych odpowiedzi została obliczona na podstawie wyników uzyskanych przez uczestników, a następnie wykorzystana w analizach statystycznych.

Analiza statystyczna

W celu dokonania oceny efektów treningu analizie poddano poprawność odpowiedzi w teście mierzącym zdolność odczytywania mowy z ruchu warg przed treningiem i po jego zakończeniu, uwzględniając trzy poziomy złożoności językowej: słowa, zdania, wypowiedzi. Zastosowano analizę wariancji z powtarzanymi pomiarami (ang. *repeated measures analysis of variance*, rmANOVA) w układzie 2×3 (pomiar $\times$ poziom złożoności materiału językowego). Dodatkowo w analizie uwzględniono czynnik grupowy w postaci płci uczestników, ponieważ istnieją dowody na istotne różnice międzypłciowe w zakresie dokładności odczytywania mowy z ruchu warg, przy czym kobiety zazwyczaj osiągają lepsze wyniki niż mężczyźni [43]. Siłę efektu oceniano za pomocą wskaźnika częściowego eta-kwadrat (η^2_p). Testy *post-hoc* przeprowadzono z zastosowaniem korekty Tukeya. Wszystkie analizy statystyczne wykonano w programie Jamovi 2.3 [44], a wizualizacja



Opis: **Panel a** – ekrany instruktażowe wprowadzające uczestników do testu. Uczestnicy oglądali nagrania bez dźwięku, w których lektorka wypowiadała słowa, zdania lub dłuższe wypowiedzi. Zadaniem było rozpoznawanie wypowiadanych słów. Materiał pochodził z czterech kategorii: *jedzenie, moda, sport, pogoda*. **Panel b** – rozpoznawanie pojedynczych słów. Po obejrzeniu nagrania uczestnicy wybierali jedno słowo z czterech możliwych. **Panel c** – rozumienie zdań. Po obejrzeniu nagrania uczestnicy wybierali dwa słowa spośród sześciu, które pojawiły się w zdaniu. **Panel d** – rozumienie dłuższych wypowiedzi. Po obejrzeniu nagrania uczestnicy wybierali trzy słowa z dziewięciu, które rzeczywiście wystąpiły w wypowiedzi.

Rycina 2. Procedura oceny umiejętności odczytywania mowy z ust przed i po treningu

Figure 2. Lipreading assessment procedure before and after training

wyników została przygotowana z wykorzystaniem języka programowania R [45].

Wyniki

Analiza wariancji wyników testów przeprowadzonych przed realizacją komputerowego treningu odczytywania mowy z ust i po jego zakończeniu wykazała istotny główny efekt czynnika *pomiar* ($F(1,62) = 48,18$, $p < .001$, $\eta^2_p = 0,437$), z dużą siłą efektu. Wynik ten wskazuje, że trening odczytywania mowy z ust istotnie poprawił wyniki uczestników w zakresie percepcji wizualnej mowy, średnio o 11,6 punktu procentowego (pp.), niezależnie od poziomu złożoności materiału językowego.

Zaobserwowano również istotny główny efekt *poziomu złożoności materiału językowego* ($F(2,124) = 11,79$, $p < .001$, $\eta^2_p = 0,16$), co sugeruje, że poziom złożoności materiału miał istotny wpływ na wyniki uczestników, niezależnie od momentu pomiaru (przed czy po treningu). Analiza *post-hoc* wykazała, że słowa i zdania nie różniły się istotnie pod

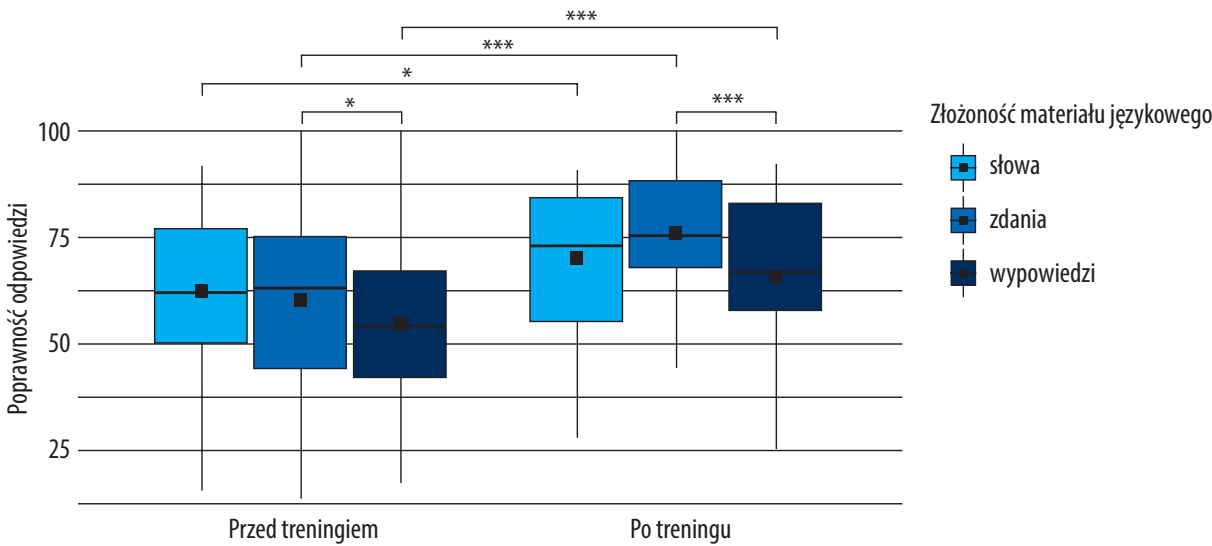
względem poziomu trudności ($t(63) = -1,10$, $p = 0,515$), a średnia różnica procentowej poprawności między słowami a zdaniami wynosiła $-2,04$ pp. Wypowiedzi wielozdaniowe, zgodnie z oczekiwaniami, okazały się istotnie trudniejsze od słów ($t(63) = 3,10$, $p = 0,008$, średnia różnica = 6,12 pp.) oraz od zdań ($t(63) = 6,25$, $p < .001$, średnia różnica = 8,16 pp.).

Co istotne, analiza wykazała również słabą, ale istotną interakcję między *pomiarem a poziomem złożoności materiału językowego* ($F(2,124) = 3,45$, $p = 0,035$, $\eta^2_p = 0,053$). Analiza porównań *post-hoc* dla tej interakcji wykazała, że przed rozpoczęciem treningu poziom poprawności rozumienia pojedynczych słów nie różnił się istotnie od poziomu poprawności rozumienia zdań ($t(62) = 0,73$, $p = 0,977$, średnia różnica = 1,97) ani od rozumienia wypowiedzi ($t(62) = 2,74$, $p = 0,081$, średnia różnica = 8,01). Istotna różnica pojawiła się jednak między zdaniami a wypowiedziami ($t(62) = 3,47$, $p = 0,012$, średnia różnica = 6,04). Po zakończeniu treningu nie odnotowano istotnej różnicy między poziomem poprawności dla słów i zdań ($t(62) = -2,79$, $p = 0,072$,

Tabela 1. Statystyki opisowe poprawności [%] w trzech typach zadań przed i po treningu, z podziałem na płeć
Table 1. Descriptive statistics of accuracy [%] in three task types before and after training, by gender

		Przed treningiem			Po treningu		
		słowa	zdania	wypowiedzi	słowa	zdania	wypowiedzi
Średnia	mężczyźni	58,5	56	52,7	63,5	71,6	60,5
	kobiety	65,6	64,2	55,5	75,9	80	70,7
	wszystkie osoby badane	62,2	60,3	54,1	69,9	75,9	65,7
Mediana	mężczyźni	62	56	58	64	75	58
	kobiety	69	63	50	82	81	75
	wszystkie osoby badane	62	63	54	73	75	67
Odchylenie standardowe	mężczyźni	19,2	15,5	18,7	17,2	14,8	14,6
	kobiety	19,1	21,1	21	16,7	15,8	15,8
	wszystkie osoby badane	19,3	18,9	19,8	17,9	15,8	15,9

Opis: Tabela przedstawia statystyki opisowe (średnią, medianę i odchylenie standardowe) poprawności odpowiedzi wyrażonej w punktach procentowych (0–100) w trzech typach zadań językowych: rozpoznawaniu słów, rozumieniu zdań oraz rozumieniu dłuższych wypowiedzi. Wyniki zaprezentowano osobno dla pomiaru przed i po treningu odczytywania mowy z ruchu warg, z uwzględnieniem podziału na płeć uczestników. Dane ilustrują zmiany w poziomie poprawności w zależności od rodzaju zadania i etapu pomiaru.



Opis: Poprawność odpowiedzi przed i po treningu odczytywania mowy z ust w zależności od złożoności materiału językowego (słowa, zdania, wypowiedzi). Wykresy pudełkowe przedstawiają medianę (pozioma linia), średnią (czarny kwadrat) oraz zakres wyników. Istotne różnice zaznaczono gwiazdką: * $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$.

Rycina 3. Poprawność rozumienia mowy na podstawie ruchu warg przed i po treningu w zależności od złożoności materiału językowego
Figure 3. Accuracy of speech comprehension based on lipreading before and after training, depending on the linguistic complexity of the material

średnia różnica = 2,18) ani między słowami a wypowiedziami ($t(62) = 1,88$, $p = 0,426$, średnia różnica = -6,09). Istotna różnica wystąpiła ponownie między zdaniami a wypowiedziami, gdzie zdania były rozumiane znacznie lepiej niż dłuższe wypowiedzi ($t(62) = 5,14$, $p < .001$, średnia różnica = 10,22).

Porównania wyników przed i po treningu wykazały istotną poprawę na wszystkich poziomach złożoności materiału językowego. Poprawność w rozumieniu pojedynczych

słów znacząco wzrosła po treningu o średnio 7,6 pp. ($t(62) = 3,26$, $p = 0,021$). Rozumienie zdań poprawiło się najbardziej ($t(62) = 6,74$, $p < .001$), osiągając najwyższy wzrost poprawności, średnio o 15,67 pp. Wypowiedzi również wykazały istotny wzrost ($t(62) = 4,36$, $p < .001$), poprawiając się średnio o 11,49 pp. po treningu (tabela 1).

W analizie uwzględniającej płeć uczestników jako czynnik grupowy, czynnik płci okazał się istotny wyłącznie na poziomie efektu głównego ($F(1,62) = 7,08$, $p = 0,01$, $\eta^2_p = 0,102$),

co wskazuje, że niezależnie od momentu pomiaru (przed czy po treningu) oraz poziomu złożoności materiału językowego, kobiety osiągały lepsze wyniki niż mężczyźni, średnio o 8,2 pp. Siła tego efektu jest umiarkowana, co sugeruje, że płeć stanowi istotny czynnik wpływający na zdolność wizualnej percepcji mowy. Natomiast żaden z efektów interakcyjnych nie okazał się istotny statystycznie. Nie stwierdzono znaczącej interakcji *pomiar* $\times$ *pleć* ($F(1,62) = 1,66$, $p = 0,203$, $\eta^2_p = 0,026$), *poziom złożoności materiału językowego* $\times$ *pleć* ($F(2,124) = 0,45$, $p = 0,64$, $\eta^2_p = 0,007$) ani *pomiar* $\times$ *poziom złożoności materiału językowego* $\times$ *pleć* ($F(2,124) = 0,75$, $p = 0,473$, $\eta^2_p = 0,012$). Oznacza to, że poprawa wyników po treningu była podobna u kobiet i mężczyzn, wpływ poziomu trudności materiału na wyniki uczestników nie zależał od płci, a zmiany wyników po treningu w zależności od poziomu trudności były porównywalne w obu grupach (**rycina 3**).

Dyskusja

Wyniki przeprowadzonego badania potwierdziły poprawę umiejętności odczytywania mowy z ust uczestników opracowanego przez nas treningu. Wyniki wykazały, że ukierunkowany trening odczytywania mowy z ust u osób ze słuchem w normie, które wcześniej nie ćwiczyły tej umiejętności, może poprawić percepcję mowy wizualnej. Co ważne, wzrost umiejętności potwierdził się także w przypadku materiału, który nie był wcześniej trenowany. Jest to zgodne z doniesieniami wielu wcześniejszych badań, które wykazały umiarkowaną, lecz istotną poprawę umiejętności odczytywania mowy z ust w następstwie treningu – zarówno w przypadku osób słyszących (dorośli [46,47] i dzieci [48]), jak i populacji osób z ubytkami słuchu [49–51].

Co istotne, najnowsze badania podkreślają, że skuteczność treningu zależy od zastosowanych metod i warunków uczenia się. Tradycyjne programy treningowe, skupiające się np. na odgadywaniu całych zdań, często przynosiły ograniczone efekty [46]. Zgodnie z literaturą trudności związane z poprawą umiejętności odczytywania mowy z ust wynikają z kilku powodów, a szczególnie istotny jest fakt, że artykulacja wielu fonemów wygląda bardzo podobnie. Zjawisko to opisano szeroko w polskiej literaturze jako jeden z kluczowych czynników ograniczających dokładność percepcji wizualnej — przykładowo, głoski /p/, /b/ i /m/ są reprezentowane przez ten sam kinem artykulacyjny i są niemal nierozróżnialne bez kontekstu [16]. Niemniej treningi wykorzystujące zasadę tzw. stopniowego percepcyjnego uczenia się (ang. *perceptual learning*) przyniosły obiecujące rezultaty. Na przykład Bernstein i wsp. [46] wykazali, że zastosowanie specjalistycznej informacji zwrotnej podczas treningu (ukierunkowanej na błędy percepcji fonemów) prowadzi do znaczącej poprawy rozpoznawania mowy z ust, a efekt ten uogólnia się na nowe materiały testowe oraz poprawia rozumienie mowy audiowizualnej w szumie.

Jak wspomniano, zaobserwowane efekty poprawy umiejętności odczytywania mowy z ust można wyjaśnić w kategoriach uczenia się percepcyjnego [52], jak i reorganizacji procesów poznawczych zaangażowanych w percepcję mowy. Uczenie percepcyjne definiuje się jako długotrwałą zmianę w systemie percepcyjnym organizmu, poprawiającą jego zdolność reagowania na bodźce. W kontekście

przeprowadzonego badania oznacza to, że trening mógł usprawnić ekstrakcję cech wizualnych mowy na poziomie fonetycznym, jak również zoptymalizować sposób, w jaki uczestnicy wykorzystują wskazówki kontekstowe i swoją wiedzę językową podczas oglądania ruchu warg.

Jednym z kluczowych mechanizmów wydaje się zwiększenie uwagi na informacje subfonemiczne (subwizemiczne) dostępne w ruchach artykulatorów. W literaturze podkreśla się, że choć wiele fonemów należy do tych samych kinemów, wprawni czytający z ust potrafią dostrzec nawet drobne różnice artykulacyjne poniżej poziomu kinemu [53]. Bernstein i wsp. wykazali, że niektóre osoby (zarówno głuche, jak i słyszące) są w stanie rozróżniać elementy mowy na podstawie subtelnych szczegółów ruchu warg, co dowodzi istnienia percepcji subwizemicznej. Uzyskane wyniki sugerują, że trening mógł pomóc uczestnikom w bardziej precyzyjnym wychwytywaniu subtelnych wskazówek wizualnych, co przełożyło się na lepsze rozpoznawanie poszczególnych fonemów i słów. Innymi słowy, praktyka mogła udoskonalić wewnętrzne reprezentacje fonetyczne mowy wizualnej oraz zwiększyć wrażliwość percepcyjną na ruchy ust odpowiadające różnym dźwiękom mowy.

Warto również rozważyć wpływ procesów integracji informacji sensorycznych na uzyskane efekty. Mowa jest zjawiskiem multisensorycznym – informacje wzrokowe i słuchowe w typowych warunkach są łączone przez system poznawczy w spójną percepcję. Klasycznym przykładem potwierdzającym dwumodalny odbiór mowy jest efekt McGurka [54], w przypadku którego niezgodność bodźca wzrokowego i słuchowego (np. widzimy [ga], słyszymy [ba]) skutkuje powstaniem odmiennej percepcji ([da]), która jest połączeniem obu modalności. Zjawisko to pokazuje, że mózg analizuje ruch warg w kategoriach fonetycznych i łączy je z dźwiękami mowy w ramach zintegrowanej percepcji. Skuteczność treningów opartych na percepcji wzrokowej może być zatem wzmacniana przez takie procedury, które sprzyjają integracji sensorycznej i wspierają samoregulację percepcji. Jak wykazano w badaniach Bernstein i wsp. [55], zasadniczą rolę w procesie uczenia się oraz uogólniania nabytych umiejętności na sytuacje wymagające percepcji audiowizualnej odgrywa dostępność informacji zwrotnej – uczestnicy, którzy podczas treningu czytania z ust otrzymywali pełne informacje zwrotne, osiągalni istotnie lepsze wyniki także w rozumieniu mowy w szumie. Podobne założenie przyjęto w prezentowanym treningu: po każdej odpowiedzi udzielonej przez uczestnika w zadaniu bez dźwięku i bez napisów materiał był prezentowany ponownie – tym razem z dźwiękiem, co umożliwiało porównanie własnych hipotez percepcyjnych z poprawną realizacją fonetyczną. Tego rodzaju mechanizm, oparty na informacji zwrotnej, sprzyja nie tylko rozwojowi strategii predykcyjnych, lecz także integracji modalności oraz kalibracji percepcji w oparciu o różne typy wskazówek.

Trening odczytywania mowy z ust mógł zatem usprawnić wewnętrzne modele predykcyjne mowy – uczestnicy mogli nauczyć się lepiej przewidywać, jaki dźwięk odpowiada danemu ruchowi ust, a także skuteczniej wykorzystywać informacje wzrokowe do rozstrzygania niejednoznaczności słuchowych. Plutecka [56] wskazuje, że skuteczność

odczytywania mowy z ust zależy w dużej mierze od zdolności różnicowania optycznych właściwości głosek oraz doświadczenia w porozumiewaniu się, a także od znajomości kinemów artykulacyjnych. W literaturze opisano zjawisko przeniesienia się efektów treningu między modalnościami, co zaobserwowano również w niektórych badaniach: na przykład uczestnicy trenowani w czytaniu z ust poprawiali nie tylko rozpoznawanie mowy wizualnej, lecz także rozpoznawanie słuchowe w warunkach hałasu [55,57].

Po treningu odnotowano silniejszy wzrost umiejętności rozumienia zdań i dłuższych wypowiedzi niż pojedynczych słów. Wynik ten wskazuje, że materiał językowy, który początkowo sprawiał badanym więcej trudności, po treningu stał się bardziej zrozumiały, osiągając poziom nie różniący się istotnie od poziomu rozumienia pojedynczych słów. Wynik ten może wskazywać, że opracowany przez nas trening był szczególnie skuteczny w odniesieniu do bardziej złożonego materiału językowego, który w największym stopniu odzwierciedla naturalną komunikację językową. Różnicowany wpływ przeprowadzonego treningu na poprawę rozumienia mowy w oparciu tylko o kanał wzrokowy dla materiału o różnym stopniu złożoności wskazuje na możliwe mechanizmy oddziaływania treningu.

Jak zauważa Kaczmarek [58], słuchacz nie tylko odczytuje poszczególne głoski, ale dzięki przechowywanym w pamięci modelom fonemów oraz znajomości reguł językowych jest w stanie uzupełniać brakujące elementy wypowiedzi i właściwie rekonstruować jej znaczenie, szczególnie gdy kontekst wspiera proces dekodowania. Z kolei Styczek [10] podkreśla, że skuteczna percepcja mowy wymaga nie tylko analizy fonematycznej, lecz także integracji elementów semantycznych i składniowych w celu zrekonstruowania spójnej treści komunikatu. Takie podejście do przetwarzania mowy – zakładające aktywne formułowanie hipotez znaczeniowych na podstawie dostępnych wskazówek – znajduje potwierdzenie także w badaniach Krakowiak [59], która wskazuje, że skuteczne odczytywanie mowy z ust opiera się zarówno na kompetencjach wzrokowych, jak i językowych, umożliwiających dekodowanie bardziej złożonych struktur.

W rozumieniu zdań i dłuższych wypowiedzi pozbawionych dźwięku biorą zatem udział zarówno zdolności percepcyjno-sensoryczne, jak i kontekstowo-semantyczne. Pierwsze, o charakterze *bottom-up*, dzięki uczeniu się przez doświadczenie pozwalają na coraz sprawniejsze wychwytywanie drobnych elementów różnicujących fonemy w ruchach artykulacyjnych. Drugie, o charakterze *top-down*, wykorzystują wiedzę o kontekście wypowiedzi czy informacje składniowe, które mogą ułatwić rozpoznanie treści podczas odczytywania mowy z ust [60].

Rozumienie pojedynczych słów bazuje głównie na mechanizmach *bottom-up*, natomiast zdania i dłuższe wypowiedzi wymagają zaangażowania obu mechanizmów jednocześnie [46]. Podczas przeprowadzonego treningu ćwiczone było rozumienie zarówno pojedynczych słów, jak i zdań oraz wypowiedzi, co umożliwiło równoległe wzmacnianie obu typów mechanizmów poznawczych. Wyniki sugerują, że zaobserwowana poprawa rozumienia mowy z ruchów warg dotyczyła wszystkich poziomów złożoności materiału

językowego (słów, zdań oraz dłuższych wypowiedzi), co wskazuje na jednoczesne wzmocnienie zarówno mechanizmów *bottom-up*, jak i *top-down*. Istotna interakcja pomiędzy pomiarem a poziomem złożoności materiału wykazała jednak, że największą poprawę odnotowano w zakresie rozumienia zdań, które po treningu były rozumiane istotnie lepiej niż dłuższe wypowiedzi. Może to sugerować, że o ile trening skutecznie poprawił zdolność wykorzystywania wizualnych wskazówek percepcyjnych i informacji kontekstowych na wszystkich poziomach złożoności, o tyle złożoność dłuższych wypowiedzi mogła wciąż przekraczać dostępne zasoby poznawcze uczestników, ograniczając efektywność wykorzystania informacji kontekstowych w najbardziej wymagającym materiale. Zdolność do efektywnego wykorzystania obu typów mechanizmów zależy od wielu czynników, w tym kompetencji językowych oraz zasobów poznawczych [61].

Podsumowując, trening sprawił, że uczestnicy nauczyli się lepszego różnicowania artykulacji fonemów o podobnej wizualnej charakterystyce, co zmniejszyło wcześniejsze trudności wynikające z istnienia wizemów grupujących fonemy trudne do rozróżnienia wzrokowego [16,17,62]. Trening zwiększył wrażliwość na subtelne różnice artykulacyjne, takie jak tempo mowy, ułożenie warg czy mimowolne ruchy mięśni, i w rezultacie doprowadził do wzrostu skuteczności rozpoznawania pojedynczych słów, tworząc przestrzeń do szerszego wykorzystania informacji *top-down*, co wpłynęło na polepszenie rozumienia przekazów zawierających aspekt kontekstu.

Wyniki wykazały także różnice międzypłciowe w poziomach wykonania testów odczytywania mowy z ust przed treningiem i po jego ukończeniu. Sugerują one jednak, że chociaż kobiety osiągały ogólnie wyższe wyniki niż mężczyźni, to wpływ treningu oraz poziomu trudności materiału był podobny w obu grupach. Brak istotnych interakcji wskazuje, że zdolności poznawcze, na które ukierunkowany był trening, usprawniły się w zbliżonym stopniu niezależnie od płci uczestników, a zatem trening oddziałuje w podobny sposób zarówno w odniesieniu do kobiet, jak i mężczyzn. Z drugiej strony, odnotowane różnice międzypłciowe pozostawały stabilne zarówno przed treningiem, jak i po jego ukończeniu, co wskazuje na wyższy poziom zdolności odczytywania mowy z ust u kobiet niezależny od treningu. Dodatkowo przewaga kobiet była stała we wszystkich typach bodźców – zarówno w przypadku rozpoznawania pojedynczych słów, zdań, jak i wypowiedzi. Może to świadczyć o zaangażowaniu dodatkowych zdolności poznawczych niezwiązanych ze zdolnościami, na które ukierunkowany był trening, np. wiążących się z rozpoznaniem niewerbalnej komunikacji i społecznym poznaniem. Różnice międzypłciowe w mechanizmach rozpoznawania mowy wizualnej sugerowane były także we wcześniejszych badaniach behawioralnych [43,63] i neuroobrazowych [64]. Spójnie z rezultatem prezentowanego eksperymentu, wcześniejsze prace wykazały także, że kobiety przeciętnie osiągają lepsze wyniki w czytaniu mowy z ust niż mężczyźni [37,43,65,66]. Z perspektywy analizy statystycznej istotne jest, że zarówno kobiety, jak i mężczyźni wykazali porównywalny wzrost kompetencji w zakresie odczytywania mowy z ust, a brak efektów interakcyjnych między płcią a pomiarem wskazuje na równorzędną skuteczność zastosowanego treningu

w obu grupach. Rezultaty te wspierają założenie o uniwersalnym charakterze opracowanego programu.

Uzyskane wyniki mają istotne implikacje praktyczne, zwłaszcza dla programów rehabilitacji słuchu i edukacji osób z ubytkiem słuchu. Trening odczytywania mowy z ust może stać się cennym uzupełnieniem tradycyjnej terapii logopedycznej i treningów słuchowych dla osób niedosłyszących. W przypadku dorosłych z nabytą utratą słuchu, którzy korzystają z aparatów słuchowych lub implantów ślimakowych, nabycie umiejętności efektywnego wykorzystywania informacji wizualnych może znacząco poprawić rozumienie mowy w codziennych sytuacjach. Zgodnie z polskimi koncepcjami surdopedagogicznymi odczytywanie mowy z ust należy traktować jako jeden z kluczowych elementów strategii kompensacyjnych, wspierających zarówno edukację, jak i integrację społeczną osób z niedosłuchem [16,67]. Już w klasycznych ujęciach Góralówny [68] podkreślano, że rozwój umiejętności wizualnej percepcji mowy powinien być integralnym komponentem rehabilitacji słuchu, a ćwiczenia w zakresie czytania z ust powinny być wdrażane od najwcześniejszych etapów terapii dziecka z ubytkiem słuchu. Uzyskane rezultaty dostarczają empirycznego wsparcia dla włączenia tego typu treningu do programów rehabilitacyjnych. Pacjenci, u których wdrożono by sprzężoną terapię słuchową i wzrokową, mieliby szansę na większe wykorzystanie resztek słuchu w połączeniu ze wzrokiem, co w efekcie zwiększyłoby efektywność ich strategii kompensacyjnych w komunikacji.

Wnioski

Przeprowadzone badanie wykazało, że miesięczny zdalny trening odczytywania mowy z ust prowadzi do istotnej

poprawy percepcji wizualnej mowy u dorosłych. Efekty treningu były widoczne na wszystkich poziomach złożoności materiału językowego – od pojedynczych słów, przez zdania, aż po dłuższe wypowiedzi – co sugeruje jego skuteczność zarówno w usprawnianiu oddolnych mechanizmów percepcyjnych (*bottom-up*), jak i odgórnych mechanizmów kontekstowo-semantycznych (*top-down*). Szczególnie duża poprawa została zaobserwowana w rozumieniu zdań, co może świadczyć o wysokiej przydatności treningu w sytuacjach komunikacyjnych zbliżonych do codziennego użycia języka. Ponadto, choć kobiety osiągały ogólnie wyższe wyniki niż mężczyźni, poprawa po treningu była podobna w obu grupach, co wskazuje na uniwersalność zastosowanego programu niezależnie od płci. Wyniki te podkreślają potencjał praktyczny treningu odczytywania mowy z ust – zarówno w kontekście wspierania osób z ubytkiem słuchu, jak i jako narzędzia poprawiającego komunikację w trudnych warunkach akustycznych. Opracowany zdalny program treningowy może również stanowić podstawę do wdrożenia aplikacji terapeutycznej wspomagającej rehabilitację słuchu i mowy. Tego typu narzędzie mogłoby znaleźć zastosowanie zarówno w terapii indywidualnej, jak i w ramach wsparcia oferowanego przez placówki edukacyjne czy kliniczne. Włączenie takiej aplikacji do programów logopedycznych lub audiologicznych zwiększyłoby dostępność skutecznych form wsparcia, zwłaszcza dla osób z ograniczonym dostępem do specjalistów.

Finansowanie

Niniejsze badania i artykuł nie otrzymały żadnej dotacji od agencji działających w sektorze publicznym, komercyjnym lub non-profit.

Piśmiennictwo

- Kurcz I. Język i psychologia. Warszawa: Państwowe Wydawnictwo Naukowe; 1992.
- Chomsky N. Aspects of the Theory of Syntax. Cambridge, MA: MIT Press; 1965.
- Mattys SL, Davis MH, Bradlow AR, Scott SK. Speech recognition in adverse conditions: a review. *Lang Cogn Process*, 2012; 27(7–8): 953–78; <https://doi.org/10.1080/01690965.2012.705006>.
- Kuhl PK. Early language acquisition: cracking the speech code. *Nat Rev Neurosci*, 2004; 5(11): 831–43; <https://doi.org/10.1038/nrn1533>
- Tomasello M. Constructing a Language: A Usage-Based Theory of Language Acquisition. Cambridge (MA): Harvard University Press; 2003.
- Lavelli M, Fogel A. Developmental changes in mother-infant face-to-face communication: birth to 3 months. *Dev Psychol*, 2002; 38(2): 288–95; <https://doi.org/10.1037/0012-1649.38.2.288>.
- Murray L, De Pascalis L, Bozicevic L, Hawkins L, Sclafani V, Ferrari PF. The functional architecture of mother-infant communication, and the development of infant social expressiveness in the first two months. *Sci Rep*, 2016; 6: 39019; <https://doi.org/10.1038/srep39019>.
- Colonnesi C, Zijlstra BJ, van der Zande A, Bögels SM. Coordination of gaze, facial expressions and vocalizations of early infant communication with mother and father. *Infant Behav Dev*, 2012; 35(3): 523–32; <https://doi.org/10.1016/j.infbeh.2012.02.004>.
- Alviar C, Sahoo M, Edwards LA, Jones W, Klin A, Lense M. Infant-directed song potentiates infants' selective attention to adults' mouths over the first year of life. *Dev Sci*, 2023; 26(5): e13359; <https://doi.org/10.1111/desc.13359>.
- Styczek I. Logopedia. Warszawa: PWN; 1979.
- Bergeson TR, Houston DM, Miyamoto RT. Effects of congenital hearing loss and cochlear implantation on audiovisual speech perception in infants and children. *Restor Neurol Neurosci*, 2010; 28(2): 157–65; <https://doi.org/10.3233/RNN-2010-0522>.
- Kaczorowska-Bray K, Milewski S. Specyfika rozwoju umiejętności komunikacyjnych dzieci z uszkodzeniem narządu wzroku. *Logopedia*, 2022; 51(2): 77–95.
- Pruszewicz A. Fizjologiczny rozwój mowy. W: Foniatria kliniczna. Pruszewicz A (red.). Warszawa: PZWL; 1992, s. 222–5.
- Rosenblum LD. Primacy of multimodal speech perception. W: The handbook of speech perception. Pisoni DB, Remez RE (red.). Oxford: Blackwell Publishing; 2005, s. 51–78.
- Matyjek M, Kita S, Torralba Cuello M, Soto Faraco S. Multisensory integration of speech and gestures in a naturalistic paradigm. *Hum Brain Mapp*, 2024; 45(11): e26797; <https://doi.org/10.1002/hbm.26797>.
- Szczepankowski B. Fonetyka wizualna. *Logopedia*, 2022; 51(2): 65–75.

17. Auer ET Jr, Bernstein LE, Singh M. Comparing cortical activity during the perception of two forms of biological motion for language communication. International Conference on Auditory-Visual Speech Processing – AVSP 2001.
18. Bernstein LE, Demorest ME, Tucker PE. Speech perception without hearing. *Percept Psychophys*, 2000; 62: 233–52; <https://doi.org/10.3758/BF03205546>.
19. Kurcz I. Język a reprezentacja świata w umyśle. Warszawa: PWN; 1987.
20. Summerfield Q. Lipreading and audio-visual speech perception. *Philos Trans R Soc Lond B Biol Sci*, 1992; 335(1273): 71–8; <https://doi.org/10.1098/rstb.1992.0009>.
21. Chen Y, Hazan V. Developmental factors and the non-native speaker effect in auditory-visual speech perception. *J Acoust Soc Am*, 2009; 126(2): 858–65; <https://doi.org/10.1121/1.3158823>.
22. Cienkowski KM, Carney AE. Auditory-visual speech perception and aging. *Ear Hear*, 2002; 23(5): 439–49; <https://doi.org/10.1097/00003446-200210000-00006>.
23. Gijbels L, Yeatman J, Lalonde K, Doering P, Lee A. Audiovisual speech perception benefits are stable from preschool through adolescence. *Multisens Res*, 2024; 37(4–5): 317–40; <https://doi.org/10.1163/22134808-bja10128>.
24. World Health Organization. World report on hearing. Geneva: WHO; 2021.
25. Amieva H, Ouvrard C, Meillon C, Rullier L, Dartigues JF. Death, depression, disability, and dementia associated with self-reported hearing problems: a 25-year study. *J Gerontol A Biol Sci Med Sci*, 2018; 73(10): 1383–9; <https://doi.org/10.1093/gerona/glx250>.
26. Lawrence BJ, Jayakody DMP, Bennett RJ, Eikelboom RH, Gasson N, Friedland PL. Hearing loss and depression in older adults: a systematic review and meta-analysis. *Gerontologist*, 2020; 60(3): e137–54; <https://doi.org/10.1093/geront/gnz009>.
27. Mikkola TM, Portegijs E, Rantakokko M, Gagné J-P, Rantanen T, Gagne J-P, Rantanen T, Viljanen A. Association of self-reported hearing difficulty to objective and perceived participation outside the home in older community-dwelling adults. *J Aging Health*, 2015; 27(1): 103–22; <https://doi.org/10.1177/0898264314538662>.
28. Bennion A, Forshaw MJ. Insights from the experiences of older people with hearing impairment in the United Kingdom: recommendations for nurse-led rehabilitation. *Int J Older People Nurs*, 2013; 8(4): 270–8; <https://doi.org/10.1111/j.1748-3743.2012.00318.x>.
29. Schmitt R, Meyer M, Giroud N. Improvements in naturalistic speech-in-noise comprehension in middle-aged and older adults after 3 weeks of computer-based speechreading training. *NPJ Aging*, 2023; 9: 1–8; <https://doi.org/10.1038/s41539-023-00179-6>.
30. Bernstein JGW, Auer ET, Tucker PE. Visual speech perception: a tutorial on letter-to-sound speechreading. *J Speech Lang Hear Res*, 2020; 63(7): 2341–62; https://doi.org/10.1044/2020_JSLHR-19-00133.
31. Santana RS, Menezes EC, Ralin V, Givigi S, Batorowicz B. Lipreading as a communication strategy to enhance speech recognition in individuals with hearing impairment: a scoping review. *CoDAS*, 2023; 35: e20220273; <https://doi.org/10.1590/2317-1782/20232022273>.
32. Ewing I. O odczytywaniu mowy z ust i aparatach słuchowych. Góralówna M (oprac.), Szczepankowski B (tłum.). Warszawa: Polski Związek Głuchych; 1974.
33. van Uden A. A world of language for deaf children. Amsterdam: Swets & Zeitlinger; 1977.
34. Campbell R. Speechreading: advances in understanding its cortical bases and implications for deafness and speech rehabilitation. *Scand Audiol Suppl*, 1998; 49: 80–6; <https://doi.org/10.1080/010503998420694>.
35. MacLeod A, Summerfield Q. Quantifying the contribution of vision to speech perception in noise. *Br J Audiol*, 1987; 21(2): 131–41; <https://doi.org/10.3109/03005368709077786>.
36. Ross LA, Saint-Amour D, Leavitt VM, Javitt DC, Foxe JJ. Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. *Cereb Cortex*, 2007; 17(5): 1147–53; <https://doi.org/10.1093/cercor/bhl024>.
37. Tye-Murray N, Sommers MS, Spehar B. Audiovisual integration and lipreading abilities of older adults with normal and impaired hearing. *Ear Hear*, 2007; 28(5): 656–68; <https://doi.org/10.1097/AUD.0b013e31812f7185>.
38. Feld JE, Sommers MS. Lipreading, processing speed, and working memory in younger and older adults. *J Speech Lang Hear Res*, 2009; 52(6): 1555–65; [https://doi.org/10.1044/1092-4388\(2009/08-0137\)](https://doi.org/10.1044/1092-4388(2009/08-0137)).
39. Spehar BP, Tye-Murray N, Sommers M, Barcroft J. Audiovisual speech perception and lipreading abilities of older adults with normal and impaired hearing. *Ear Hear*, 2008; 29(3): 398–408; <https://doi.org/10.1097/AUD.0b013e31816453ac>.
40. Kovačević T, Isaković L. The role and significance of lipreading. *Teme*, 2023; XLVII(4): 923–38.
41. Domagała-Zyśk E. Kompetencje młodzieży gimnazjalnej z uszkodzonymi słuchem w zakresie pisania w języku angielskim. W: Teoria i praktyka oddziaływań profilaktyczno-wspierających rozwój osób z niepełnosprawnościami. Baran J, Gunia G (red.), 2013; t. 2, vol. 1, s. 155–75.
42. Peirce J, Gray JR, Simpson S, MacAskill M, Höchenberger R, Sogo H, i wsp. PsychoPy2: experiments in behavior made easy. *Behav Res Methods*, 2019; 51(1): 195–203; <https://doi.org/10.3758/s13428-018-01193-y>.
43. Bernstein LE. Response errors in females' and males' sentence lipreading necessitate structurally different models for predicting lipreading accuracy. *Lang Learn*, 2018; 68: 127–58; <https://doi.org/10.1111/lang.12281>.
44. The jamovi project. Jamovi, version 2.3 [software], 2022.
45. R Core Team. R: a language and environment for statistical computing, version 4.1 [software], 2021.
46. Bernstein LE, Jordan N, Auer ET, Eberhardt SP. Lipreading: a review of its continuing importance for speech recognition with an acquired hearing loss and possibilities for effective training. *Am J Audiol*, 2022; 31(2): 453–69; https://doi.org/10.1044/2021_AJA-21-00112.
47. Blumsack JT, Bower CR, Ross ME. Comparison of speechreading training regimens. *Percept Mot Skills*, 2007; 105(3 Pt 1): 988–96; <https://doi.org/10.2466/pms.105.3.988-996>.
48. Buchanan-Worster E, Hulme C, Dennan R, MacSweeney M. Speechreading in hearing children can be improved by training. *Dev Sci*, 2021; 24(6): e13124; <https://doi.org/10.1111/desc.13124>.
49. Bernstein LE, Auer ET Jr, Tucker PE. Enhanced speechreading in deaf adults: can short-term training/practice close the gap for hearing adults? *J Speech Lang Hear Res*, 2001; 44(1): 5–18; [https://doi.org/10.1044/1092-4388\(2001/001\)](https://doi.org/10.1044/1092-4388(2001/001)).
50. Bothe HH. Training of speechreading for severely hearing-impaired persons by human and computer. 7th IEEE International Conference on Advanced Learning Technologies ICALT 2007, s. 913–4; <https://doi.org/10.1109/ICALT.2007.269>.
51. Lonka E. Speechreading instruction for hard-of-hearing adults: effects of training face-to-face and with a video programme. *Scand Audiol*, 1995; 24(3): 193–8.

52. Goldstone RL. Perceptual learning. *Annu Rev Psychol*, 1998; 49: 585–612; <https://doi.org/10.1146/annurev.psych.49.1.585>.
53. Bernstein LE. Visual speech perception. W: *Audiovisual speech processing*. Vatikiotis-Bateson E, Bailly G, Perrier P (red.). Cambridge: Cambridge University Press; 2012, s. 21–39; <https://doi.org/10.1017/CBO9780511843891.004>.
54. McGurk H, MacDonald J. Hearing lips and seeing voices. *Nature*, 1976; 264(5588): 746–8; <https://doi.org/10.1038/264746a0>.
55. Bernstein LE, Auer ET, Eberhardt SP. During lipreading training with sentence stimuli, feedback controls learning and generalization to audiovisual speech in noise. *Am J Audiol*, 2022; 31(1): 57–77; https://doi.org/10.1044/2021_AJA-21-00034.
56. Plutecka K. Metody kształtowania mowy u dzieci niesłyszących. *Przegląd stanowisk. Logopedia Silesiana*, 2016; 5: 59–72; https://journals.us.edu.pl/index.php/LOGOPEDIA_SILESIANA/article/view/7267.
57. Rosenblum LD, Miller RM, Sanchez K. Lip-read me now, hear me better later: cross-modal transfer of talker-familiarity effects. *Psychol Sci*, 2007; 18(5): 392–6; <https://doi.org/10.1111/j.1467-9280.2007.01911.x>.
58. Kaczmarek BLJ. Słuch fonematyczny a percepcja wypowiedzi słownych. *Kosmos*, 1998; 47(3): 271–6.
59. Krakowiak K, Kołodziejczyk R, Borowicz A, Domagała-Zyśk E. *Student niesłyszący we wspólnocie akademickiej*. Lublin: Wydawnictwo KUL; 2011.
60. Spehar B, Goebel S, Tye-Murray N. Effects of context type on lipreading and listening performance and implications for sentence processing. *J Speech Lang Hear Res*, 2015; 58(3): 1093–102; https://doi.org/10.1044/2015_JSLHR-H-14-0360.
61. Kupisiewicz M. Sprostować wyzwaniom – uczeń z wadą słuchu w edukacji inkluzyjnej. *Stud Teor Wych*, 2020; 11(4): 87–102.
62. Fisher CG. Confusions among visually perceived consonants. *J Speech Hear Res*, 1968; 11(4): 796–804; <https://doi.org/10.1044/jshr.1104.796>.
63. Irwin JR, Whalen DH, Fowler CA. A sex difference in visual influence on heard speech. *Percept Psychophys*, 2006; 68: 582–92; <https://doi.org/10.3758/BF03208760>.
64. Ruytjens L, Albers F, Van Dijk P, Wit H, Willemsen A. Neural responses to silent lipreading in normal hearing male and female subjects. *Eur J Neurosci*, 2006; 24(6): 1835–44; <https://doi.org/10.1111/j.1460-9568.2006.05072.x>.
65. Johnson FM, Hicks LH, Goldberg TE, Myslobodsky MS. Sex differences in lipreading. *Bull Psychon Soc*, 1988; 26(2): 106–8; <https://doi.org/10.3758/bf03334875>.
66. Auer ET Jr, Bernstein LE. Enhanced visual speech perception in individuals with early-onset hearing impairment. *J Speech Lang Hear Res*, 2007; 50(5): 1157–65.
67. Krakowiak B. *Fonogesty jako narzędzie formowania języka dzieci z uszkodzonym słuchem*. Lublin: Wydawnictwo UMCS; 1995, s. 67.
68. Góralówna M, Hołyńska B. *Rehabilitacja małych dzieci z wadą słuchu*. Warszawa: PZWL; 1984.

Mgr Martyna Bryłka, email: m.brylka@ifps.org.pl •  0000-0001-6817-5446

Dr n. ścisł. i przyrod. Joanna Beck, email: joannaludwikabeck@gmail.com •  0000-0001-7453-9242

Dr n. biol. Hanna Cygan, email: h.cygan@ifps.org.pl •  0000-0002-4518-8540