

WŁODZISŁAW DUCH  orcid.org/0000-0001-7882-4729

Laboratorium Neurokognitywne, Interdyscyplinarne Centrum Nowoczesnych Technologii
Katedra Informatyki Stosowanej
Wydział Fizyki, Astronomii i Informatyki Stosowanej
Uniwersytet Mikołaja Kopernika w Toruniu
Neurocognitive Laboratory, Center for Modern Interdisciplinary Technologies
Department of Informatics
Institute of Technical Sciences
Faculty of Physics, Astronomy and Informatics, Nicolaus Copernicus University in Toruń
e-mail: wduch@umk.pl

Projekty MetaLab i ManyBabies. Metaanalizy w psychologii rozwojowej¹

MetaLab and the ManyBabies Projects. Meta-Analysis in Developmental Psychology

Abstract. Psychological research has a serious problem with replicability. Several collaborative replication initiatives in cognitive and social psychology try to address such issues. Situation in developmental research is exceptionally challenging due to the difficulties in performing experiments, limitation of objective measurement methods, and small samples. The ManyBabies project is a multi-laboratory large-scale effort that gives a chance to provide more robust, replicable results in infancy research. After a brief discussion of the problems in psychological research the ManyBabies project is introduced and the MetaLab tools for statistical analysis of collaborative data at the meta-level are presented, using two examples in early language acquisition that are of particular interest to our lab: native and non-native vowel discrimination.

Keywords: developmental psychology, early language acquisition, meta-analysis, infancy research, vowel discrimination

Słowa kluczowe: psychologia rozwojowa, wczesne przyswajanie języka, metaanaliza, badania nad niemowlętami, rozróżnianie samogłosek

WPROWADZENIE

Badania procesów rozwojowych dzieci są technicznie trudne, trwają bardzo długo, a rezultaty często nie są jednoznaczne i trudno jest o ich replikację. Standardową metodą jest zebranie dostatecznie dużej grupy badanych, by wykazać istnienie statystycznie istotnych efektów. Indywidualne różnice, niekontrolowane zmien-

ne, drobne zmiany warunków eksperymentów, trudności zebrania reprezentatywnej, w miarę homogenicznej i odpowiednio licznej grupy badanych i grupy kontrolnej – to tylko niektóre problemy utrudniające badania rozwojowe. Dyskusja nad powtarzalnością wyników badań psychologicznych toczyła się od dawna. Od 2010 roku stała się bardzo intensywna, a jej rezultatem jest znaczna poprawa procedur ba-

dawczych w ostatnich latach (Nelson, Simmons, Simonsohn, 2018).

Przed podobnymi trudnościami stoją liczne dyscypliny badawcze zarówno w obrębie neuro nauk (Button i in., 2013), jak i medycyny, dlatego w ostatnich latach pojawiła się idea medycyny spersonalizowanej, zwanej również stratyfikowaną lub precyzyjną, która jest oparta na zindywidualizowanym fenotypie pacjenta, precyzyjnej diagnostyce. Potrzebna jest do tego wielopoziomowa fenomika, czyli szczegółowy opis organizmu i zachowania na wszystkich dostępnych poziomach opisu, od genetyki i epigenetyki przez proteomikę, metabolomikę, budowę komórek, tkanek, narządów, całą anatomię i fizjologię, mechanizmy regulacyjne układu immunologicznego aż do całych układów regulujących homeostazę, zachowania behawioralne, reakcje emocjonalne i interakcje społeczne. W efekcie próbuje się zidentyfikować homogeniczne grupy pacjentów, dla których można lepiej dobrać specyficzną terapię. Wymaga to jednak zarówno bardzo dobrze rozwiniętych metod diagnostycznych, jak i opisu wielu przypadków, gdyż z powodu dużej wariacji indywidualnej im więcej szczegółów, tym trudniej wyodrębnić podobne do siebie przypadki. Potrzebujemy modeli, które podzielią całą populację na stosunkowo niewiele odróżnialnych od siebie rodzajów fenotypów, opisujących stosunkowo często spotykane przypadki. Niektóre cechy fenotypów umożliwiają wstępną charakteryzację, ale nie mając dostatecznie dużo informacji, nie wiemy, czy w całej przestrzeni cech definiujących fenotypy można znaleźć wyraźne skupienia podobnych do siebie przypadków, pozwalające na odwołanie się do prototypów opisujących daną osobę, a więc umożliwiające generalizację indywidualnych wyników.

W dalszej perspektywie psychologia, podobnie jak medycyna i neuropsychiatria, będzie zapewne coraz silniej opierać się na fenomice i opisie indywidualnego rozwoju organizmu i jego funkcji. Będzie to możliwe dzięki indywidualnym modelom pracy mózgu, takim jak The Virtual Brain (Stefanovski i in., 2016), które pozwalają na integrację wyników neuroobrazowania, pomiarów elektrofizjologicznych i rezultatów badań behawioralnych (Schirner

i in., 2017). Wielki krok w tym kierunku stanowią badania rozwoju konektomów prowadzone w ramach The Developing Human Connectome Project (Makropoulos i in., 2018). Konektom, czyli kompletna mapa sieci połączeń neuronalnych, pozwala zrozumieć, które obszary mózgu współpracują z sobą, realizując różne funkcje. Dokładne określenie połączeń pomiędzy neuronami w przypadku złożonych mózgów nie jest możliwe, ale badania połączeń między większymi grupami neuronów, na przykład kolumnami korowymi, również niosą wiele informacji. Ustalenie struktury połączeń można wykonać bezinwazyjnie metodami traktografii, czyli obrazowania strukturalnego za pomocą rezonansu magnetycznego (MRI). Na sieciach neuronowych o tej samej strukturze można zaobserwować różne zmieniające się w czasie aktywacje, świadczące o funkcjonalnej współpracy niektórych obszarów. Służy do tego funkcjonalny rezonans magnetyczny (fMRI), mierzący poziom tlenu we krwi dopływającej do danego obszaru mózgu (sygnał BOLD, Blood-Oxygen Level Dependent). W ten sposób ustalono, które obszary mózgu biorą udział w procesach percepcji, uwagi spontanicznej i intencjonalnej, pamięci roboczej, istotności emocjonalnej, sieci czołowo-ciemieniowych, związanych z wykonaniem zadań i aktywnych w czasie spoczynku.

Rezultaty badań rozwoju konektomu łączą z sobą badania za pomocą fMRI, a także badania genetyczne, obserwacje kliniczne i behawioralne. Wymagało to rozwoju nowych metod związanych z trudnościami skanowania płodów, obrazowania struktur anatomicznych i połączeń pomiędzy nimi oraz funkcjonalnych połączeń widocznych w czasie spoczynku (*resting-state fMRI*). Pod koniec 2018 roku wykonano już prawie 800 skanów, a rok później udostępniono dane ponad 500 noworodków w wieku od 24. do 45. tygodnia od poczęcia. Wszystkie badania wykonywano w czasie snu. Badane są zarówno dzieci zdrowe, jak i zagrożone różnymi chorobami spowodowanymi zmianami genetycznymi czy czynnikami środowiskowymi, takimi jak autyzm.

Podsumowując, badania rozwoju konektomów pokazują, że komunikacja pomiędzy półkulami mózgowymi zaczyna się w 24. tygodniu

od poczęcia. Najpierw rozwijają się połączenia w pierwotnych obszarach zmysłowych i ruchowych, a później drugo- i trzeciorzędowe, integrujące informacje. Ogólny schemat połączeń jest już ustalony w momencie narodzin, ale przedwczesny poród nie pozwala na prawidłowe wykształcenie długich połączeń wzgórzo-korowych. Szkodliwe substancje mają wpływ na receptory neurotransmiterów i rozwój połączeń pomiędzy podkorowymi jądrami i obszarami czołowymi, wpływając na kontrolę emocji. Większość chorób neurologicznych i psychicznych zaczyna się od zaburzeń formowania się połączeń mózgu już w okresie prenatalnym. Indywidualne zdolności zależą od tego, jak rozwinięte są konekty w wczesnym okresie. Indywidualne różnice już w momencie narodzin są duże, a to przekłada się na różnice w wynikach badań, które opierają się głównie na wieku dziecka lub ogólnej ocenie wieku rozwojowego, nie kontrolując innych czynników. Stopniowa internalizacja istotnych bodźców pojawiających się w otoczeniu dziecka możliwa jest przede wszystkim dzięki formowaniu się konektom. Takie ujęcie łączy zarówno natywistyczne podejście, w naukach kognitywnych wywodzące się od Noama Chomsky'ego, jak i konstruktywistyczne podejście empiryczne do nabywania języka. Spór natywistów i empirystów był oderwany od rzeczywistych mechanizmów odpowiedzialnych za nabywanie języka. Właściwymi pytaniami są: Jak rozwijają się struktury i sieci mózgu od początku formowania się układu nerwowego do osiągnięcia względnej stabilności w wieku dojrzałym? Jaki wpływ ma na nie środowisko i do jakiego stopnia mogą się pod jego wpływem zmienić? Jak to wpływa na nasze możliwości komunikacji?

Zrozumienie indywidualnych ścieżek rozwojowych to kwestia bardziej odległej przyszłości. Inna możliwość stworzenia dobrej teorii procesów rozwojowych to metaanaliza wszystkich eksperymentów badających określone mechanizmy poznawcze. Wymaga to zebrania w repozytorium danych opublikowanych rezultatów z wielu eksperymentów wykonanych w różnych laboratoriach i użycia odpowiednich narzędzi statystycznych pozwalających na ich analizę. Jest to celem koordynowanego przez Uniwer-

sytet Stanforda projektu MetaLab (Frank i in., 2017; Bergmann i in., 2018)², zbierającego dane na temat różnych aspektów nabywania języka. To pionierskie przedsięwzięcie naukowe ma na celu integrację badań laboratoriów zajmujących się psychologią rozwojową na świecie. Interakcyjne narzędzia dostępne na stronie tego projektu pozwalają na przeprowadzenie analiz statystycznych, pomagając w planowaniu eksperymentów.

Od czasu klasycznej książki Erica Lenneberga *Biological Foundations of Language* (1967), w której pojawiła się hipoteza okresów krytycznych w rozwoju umiejętności językowych, nasza wiedza o szczegółach tego procesu i możliwych trajektoriach rozwojowych znacznie wzrosła. Jednak w rezultacie licznych trudności z prowadzeniem i interpretacją eksperymentów z małymi dziećmi wiarygodność hipotez i teorii opartych na badaniach rozwojowych jest często niewielka. Pojawiły się problemy z powtarzalnością wyników badań, w których jest wiele pułapek wynikających z niewielkiej liczebności badanych grup i mocy stosowanych testów. Trajektorie rozwoju indywidualnych zdolności budowane na podstawie interpretacji wyników eksperymentów nie zawsze dają się z sobą uzgodnić. Badania dzieci próbują zwykle odseparować rozwój specyficznych zdolności w oderwaniu od rozwoju innych, które są z nimi nierozłącznie związane.

Stopniowe zmniejszanie się maksymalnej wrażliwości na oddziaływanie środowiska na tworzenie wewnętrznych reprezentacji ma wpływ na nabywanie języka w okresie niemowlęcym i uczenie się innych języków w późniejszym okresie. Chociaż dorośli często radzą sobie dobrze z nauką języków obcych, to w miarę pogarszania się sprawności ogólnych mechanizmów poznawczych ich zdolności językowe również ulegają zmniejszeniu (Saxton, 2010).

PROBLEMY PSYCHOLOGII – NIE TYLKO ROZWOJOWEJ

Każdy pojedynczy eksperyment może być obciążony wieloma błędami związanymi z wyborem grupy, specyficznym sposobem prowadzenia

obserwacji, rezultaty nie muszą być prawdziwe w każdej sytuacji, mogą znacznie odbiegać od wcześniej opublikowanych wyników, zależeć od nieświadomych czynników. Wpływ różnych czynników na wariancję wyników i trudności z replikacją rezultatów można badać za pomocą symulacji komputerowych (Stanley, Spence, 2014). Im słabsza jest statystyczna moc testów, tym większa wariancja wyników i trudności z ich replikacją (Button i in., 2013). Większość wyników badań dotyczących ludzi – w medycynie, neuronaukach czy psychologii – zależy od licznych czynników niemożliwych do ścisłej standaryzacji. Jeśli efekt jest wyraźnie widoczny tylko dla części badanej grupy, może ulec rozmyciu w wyniku uśredniania. Trudno jest opublikować wyniki, dla których nie ma wyraźnego statystycznego wsparcia. Nieuprawnione wnioski bywają więc wynikiem manipulacji (nie zawsze świadomie prowadzonych) mających na celu zwiększenie statystycznej istotności wyników (*p-hacking*, Simmons i in., 2011).

Jak można się zabezpieczyć przed takimi problemami? W 2016 roku w celu poprawy wiarygodności badań psychologicznych powstało The Society for the Improvement of Psychological Science (SIPS)³. Towarzystwo organizuje coroczne konferencje i wydaje czasopismo *Collabra: Psychology*, zwracając szczególną uwagę na aspekty metodologiczne. Druga ważna inicjatywa to powołanie sieci laboratoriów Psychological Science Accelerator (PSA)⁴, w której na początku 2019 roku było już ponad 350 laboratoriów z 45 krajów. Organizacja prowadzi rekrutację laboratoriów do różnych projektów. W niektórych projektach uczestniczy ponad 100 laboratoriów, gdzie przebadano w sumie ponad 2000 osób. Nadchodzi więc era projektów psychologicznych na dużą skalę, gdzie – podobnie jak w przypadku projektów fizyki czy astronomii – liczba autorów publikacji może być bardzo duża. Publikacja grupy 96 autorów *The Psychological Science Accelerator: Advancing Psychology Through a Distributed Collaborative Network* (Moshontz i in., 2018) popiera ideę rozproszonych sieci laboratoriów współpracujących na zasadach crowdsourcingu, czyli otwartych konkursów na wykonanie częściowych badań. Publikacje

w ramach projektu ManyLabs, zmierzające do replikacji badań psychologicznych i zrozumienia przyczyn rozbieżności wyników eksperymentów, mają obecnie nawet 190 autorów (Klein i in., 2014; Ebersole i in., 2016; Klein i in., 2018).

W psychologii rozwojowej zaproponowano (Frank i in., 2017) następującą procedurę, mającą na celu zwiększenie zaufania do publikowanych wyników:

1. Należy ustalić przed rozpoczęciem badań, ile danych zostanie zebranych. Pozwoli to uniknąć problemu zbyt wczesnego lub późnego zakończenia badań, gdy będzie to wygodne z punktu widzenia statystyki.
2. Mała liczebność obserwacji wymaga określenia, czy są to wyniki z jednego eksperymentu, czy z kilku powtarzanych aż do osiągnięcia założonego rezultatu.
3. Należy na początku określić wszystkie zmienne oceniane w badaniu, gdyż mając wiele zmiennych, zawsze jakieś można uznać za istotne.
4. Trzeba określić wszystkie warunki eksperymentu, w tym również zmiany, które były wynikiem niepowodzeń.
5. Jeśli pewnych obserwacji nie uwzględniono, trzeba podać statystykę z ich uwzględnieniem lub dokładnie wytłumaczyć, dlaczego trzeba je odrzucić.
6. Jeśli analiza zawiera zmienne niezależne, autorzy powinni przeprowadzić analizę statystyczną również bez takich zmiennych.
7. Recenzenci powinni wymagać, by autorzy uzasadnili niezależność swoich wyników od arbitralnych decyzji dotyczących analizy danych.

Są to problemy szczególnie istotne w badaniach niemowląt. Brak standaryzacji, niewielka liczebność badanej grupy, trudności w rekrutacji, znaczny ubytek liczby dzieci w czasie trwania badania, trudności z utrzymaniem uwagi, wytłumaczeniem procedury oraz liczne artefakty wynikające ze sposobu pomiaru (kierunku spojrzenia, obrotu głowy, pulsu czy elektroencefalografii [EEG]) to tylko niektóre przyczyny

powodujące, że badania niemowląt są znacznie trudniejsze niż badania dorosłych.

PROJEKT MANYBABIES

Badania procesu nabywania języka przez niemowlęta obejmują wiele aspektów, od rozpoznawania dźwięków mowy do rozumienia sensu słów. Te procesy tradycyjnie badano w niezależnych eksperymentach, niezależnie od siebie, jednak jest coraz więcej wyników pokazujących, że rozwój wielu umiejętności poznawczych potrzebnych do nauki języka zachodzi jednocześnie, potrzebne są więc badania systemowe i próba syntezy otrzymanych wyników (Lewis i in. 2016; Bergmann, Tsuji, Cristia, 2017). Metaanaliza wyników otrzymanych w wielu laboratoriach pozwala na wykrywanie różnych problemów i współzależności pomiędzy badanymi efektami. Projekt ManyBabies umożliwia takie analizy (Frank i in., 2019), gromadząc wyniki eksperymentów i dostarczając narzędzi do ich analiz. Ma też na celu koordynację sposobu zbierania i opisu danych pomiędzy laboratoriami na świecie, zrozumienie przyczyn różnic wyników eksperymentów, przeprowadzanych w podobny sposób, ale dających odmienne rezultaty, i oparcie się na rezultatach projektu ManyLabs (Klein i in., 2014; Ebersole i in., 2016; Klein i in., 2018).

W tym celu w 2017 roku rozpoczęto w wielu ośrodkach serię badań opartych na przedstawionych powyżej procedurach, zmierzających do replikacji najważniejszych procesów związanych z nabywaniem języka. Całkowita standaryzacja pomiędzy wieloma laboratoriami nie jest możliwa, dlatego położono nacisk na szczegółową dokumentację czynników, które mogą wpływać na wyniki przy użyciu kilku powszechnie stosowanych sposobów oceny preferencji niemowląt, takich jak habituacja, obrót głowy czy obserwacja ruchu oczu (*eye-tracking*). Zmniejszeniu wariancji wyników służą wspólnie opracowane protokoły dla każdej z metod badawczych, dokumentacja statusu społeczno-ekonomicznego i innych informacji wpływających na rozwój umiejętności językowych. Analiza ma pokazać, jak dodatkowe

czynniki wpływają na wielkość obserwowanego efektu, jaką część wariancji wyników można przypisać różnicom pomiędzy warunkami eksperymentu w różnych laboratoriach. Wiele czynników związanych ze sposobem rekrutacji czy samej analizy statystycznej lepiej jest ustalić przed rozpoczęciem eksperymentów, by uniknąć efektu obserwatora, wpływu na błędne pozytywne wyniki. Pozwoli to na ustalenie „standardowej procedury operacyjnej”, wstępną rejestrację protokołu eksperymentu. Same dane pozostaną otwarte dla dowolnych analiz, a dostęp do danych z wielu laboratoriów umożliwi ich metaanalizy. Takie podejście pomaga wykrywać ukryte tendencje w opublikowanych wynikach badań, zwiększyć lub zmniejszyć wiarygodność hipotez, ocenić powtarzalność wyników, uniknąć selekcji wyników „publikowanych” (zgodnych z oczekiwaniami), lepiej ocenić metody badawcze, wpływ wieku i innych parametrów, a dzięki temu lepiej opracować teorie procesów rozwojowych. Ułatwia to również planowanie przyszłych eksperymentów. Globalna współpraca wielu laboratoriów staje się niezbędna dla dalszego postępu w wielu dyscyplinach badawczych.

Na początek (w drodze głosowania biorących udział w projekcie laboratoriów) ustalono, by powtórzyć badania jednego z najlepiej opisanych efektów, jakim jest preferencja dla mowy skierowanej do dzieci (IDS, *infant-directed speech*)⁵. Ponieważ jest to efekt powtarzalny i silny, można będzie próbować oddzielić zmienność indywidualną i część zależną od laboratorium, w którym prowadzono badania. Z metaanaliz statystycznych wiadomo, że występuje tu efekt obserwatora – rezultaty badań o małej precyzji są silniejsze niż badań o większej precyzji. Większość badań prowadzono w Ameryce Północnej, niewiele wiemy o innych krajach i językach. Zmiany preferencji IDS z wiekiem nie zostały jeszcze dobrze opisane (Newman, Hussain, 2006). Mechanizmy wpływu takich preferencji na nabywanie języka są nadal dyskutowane. Może to być związane z wpływem mowy skierowanej na mechanizmy uwagowe, przyspieszającym rozwój zdolności do rozróżnień elementów mowy na poziomie fonologii, rozpoznawania wyrazów czy łączenia

słów z kontekstem (Kaplan i in., 1995), różnych specyficznych cech języka (Kuhl i in., 2010), rozwoju specjalizacji w kierunku języka natywnego⁶ w pierwszym roku życia (Werker, Tees, 1984; Kuhl i in., 2006; Kuhl 2007). Być może różnice wymowy brytyjskiej, w której jest mniej prozodycznych modyfikacji niż w wymowie amerykańskiej, mogą spowodować szybszy zanik preferencji IDS. Niewiele wiemy o takich efektach dla innych języków, stąd potrzeba włączenia laboratoriów z wielu krajów.

Jest tu więc kilka istotnych pytań badawczych, na które trwające od 2017 roku meta-badania, nawet w przypadku tak dobrze opisanego efektu, mogą dać nowe odpowiedzi.

METALAB

W ramach projektu MetaLab dokonano meta-analiz wielu efektów opisanych w literaturze przedmiotu i zebranych w bazach danych. Nabywanie zdolności językowych jest złożonym procesem, wymagającym rozwoju wielu umiejętności, od rozwoju słuchu fonematycznego do rozumienia sensu abstrakcyjnych słów. Umiejętności te rozwijają się w sposób z sobą skorelowany, gdyż odpowiedzialne za nie zmiany mózgu są z sobą powiązane. W sumie MetaLab na początku 2020 roku zawierała metaanalizy w 22 obszarach związanych z rozwojem poznawczym, wykorzystując 477 publikacji, w których zbadano 28 270 dzieci, uwzględniając 1804 wielkości efektów. Baza projektu MetaLab związana z nabywaniem języka przez niemowlęta i dzieci w wieku do 5. roku życia zawierała informacje na temat 12 badanych zjawisk opisanych w 174 publikacjach, porównania wielkości 569 efektów zmierzonych dla 9395 badanych. Badane zjawiska dotyczyły:

1. Rozróżniania samogłosek natywnych.
2. Rozróżniania samogłosek obcych.
3. Kategoryzacji sylab, manipulowania cechami akustycznymi.
4. Symboliki dźwięków – relacji pomiędzy formą a sensem wyrazów.
5. Nauki zależności fonotaktycznych.
6. Segmentacji słów.
7. Nauki pierwszych słów z wykorzystaniem badania kierunku spojrzenia.

8. Preferencji mowy matczynej.
9. Użyteczności etykiet w uczeniu się kategorii.
10. Reakcji wzrokowych na znane słowa.
11. Wskazywania i budowy słownika.
12. Wskazywania i budowy słownika, badań podłużnych.

Ze względu na charakter naszych własnych badań (projekt NCN „NeuroPerKog: rozwój słuchu fonematycznego i pamięci roboczej u niemowląt i dzieci”) poniżej przedstawię przykłady analiz dotyczących tylko dwóch pierwszych typów badań, pokazując, co można osiągnąć, wykorzystując narzędzia pozwalające na metaanalizy.

ISTOTNOŚĆ EFEKTU I MOC TESTU W METALAB

W wielu pracach istotność efektu określa się za pomocą p wartości, czyli prawdopodobieństwa, że przy określonych założeniach dotyczących statystyki obserwowanych rozkładów mierzonych wartości różnice pomiędzy średnimi wartościami dla grupy badanej i kontrolnej będą większe lub równe obserwowanym. Hipoteza głosząca brak różnic jest odrzucana, jeśli p jest mniejsze niż założony poziom istotności. Ponieważ to prawdopodobieństwo zależy od liczby prób, wielkość efektu ocenia się często za pomocą miary d Cohena, nazwanej estymatorem największej wiarygodności, czyli różnicy pomiędzy średnimi wartościami \bar{x}_2 dla pomiarów o liczebności n_1, n_2 podzielonymi przez odchylenie standardowe w obu niezależnych próbach.

$$(\bar{x}_1 - \bar{x}_2) / \sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}}$$

gdzie s_1, s_2 to odchylenia standardowe dla każdej z prób. Wartość $d = 1$ oznacza więc różnicę średnich o jedno odchylenie standardowe, zakładając jednakową liczebność i takie samo odchylenie standardowe dla obu grup. Używany jest też współczynnik g Hedgesa, który uwzględnia w nieco inny sposób odchylenie standardowe: zamiast n_1 i n_2 w liczniku powyższego wzoru pojawia się $(n_1 - 1)$ oraz $(n_2 - 1)$.

Inną miarą wielkości efektu jest współczynnik korelacji r Pearsona, który jest też niezależny od liczebności próby. Stosowany jest też logarytm ilorazu szans (*log odds ratio*), równy logarytmowi naturalnemu stosunku szans (prawdopodobieństw) tego, że w grupie badanej występuje oczekiwany efekt $p(E)$ do braku efektu $p(B)$, czyli szansy $p_B(E)/p_B(-E)$, do analogicznej szansy w grupie kontrolnej tego, że wystąpił w niej efekt $p_K(E)/p_K(-E)$.

$$L = \log \left(\frac{p_B(E)/p_K(-E)}{p_B(-E)/p_K(E)} \right)$$

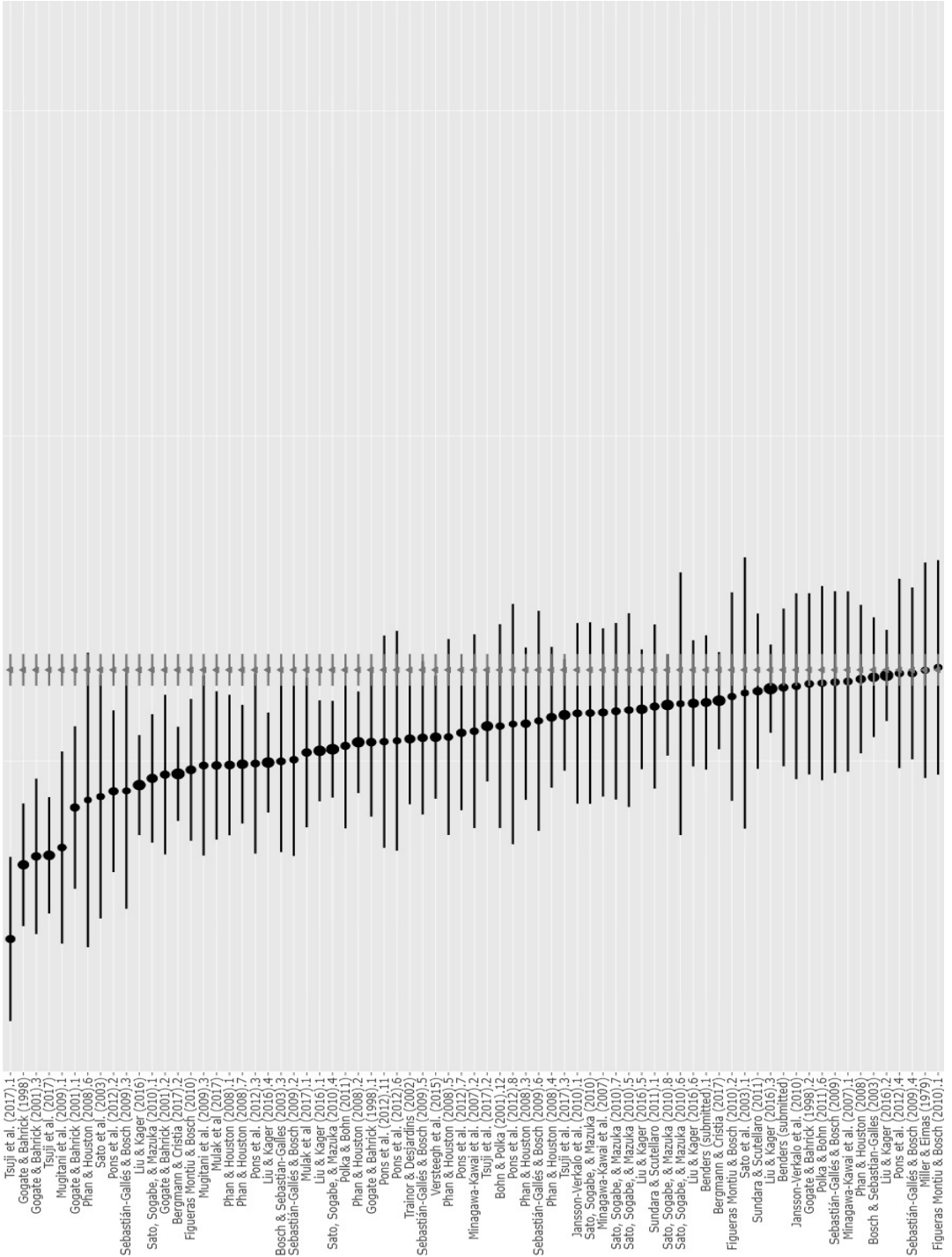
Należy też określić statystyczną moc testu, definiowaną jako jeden minus prawdopodobieństwo uznania braku efektu, podczas gdy efekt występuje, a więc dopełnienie prawdopodobieństwa popełnienia błędu drugiego rodzaju, $1-\beta$. Zalecane wartości mocy testu w badaniach behawioralnych to .8. MetaLab oferuje narzędzia pozwalające określić wymaganą liczebność grup potrzebną do osiągnięcia takiej mocy testu w zależności od wielkości efektu. Analizy oparte są na wzorach z książki *Statistical Methods for Meta-Analysis* (Hedges, Olkin, 2014). Serwer dokonujący analiz tych efektów działa na podstawie skryptów napisanych w języku *R* do obliczeń statystycznych.

ROZRÓŻNIANIE SAMOGŁOSEK NATYWNYCH

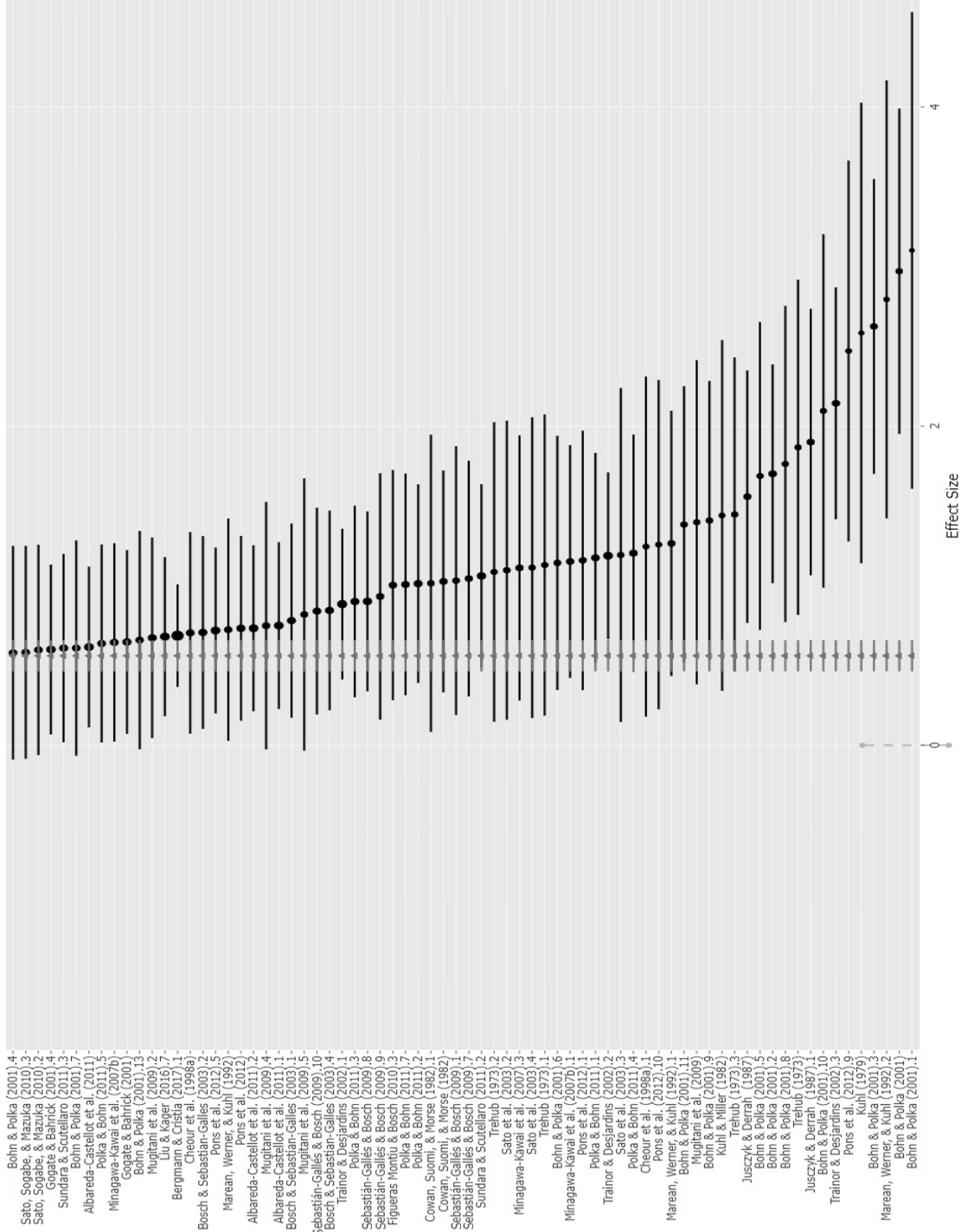
Dla ilustracji możliwości MetaLabu zostanie tu przedstawiony problem rozróżniania samogłosek natywnych i obcych przez niemowlęta. Jest to bardzo ważne zagadnienie ze względu na trudności uczenia się języków obcych o odmiennej strukturze fonetycznej. Pionierskie badania Patrycji Kuhl z University of Washington (Kuhl, 2010; 2011) i wielu innych badaczy pokazały, że już w okresie niemowlęcym pojawiają się krytyczne okresy, w których następuje szybka specjalizacja funkcji związanych z percepcją słuchową. Rozumienie mowy opiera się na kwantyzacji, czyli dzieleniu ciągłego sygnału akustycznego przekształcanego w sekwencję fonemów. Pozwala to na rozumienie mowy

niezależnie od wysokości głosu, intonacji czy jej wyrazistości. Do 8. miesiąca życia nie widać preferencji do analizy fonemów języka ojczystego, ale już w 10.–12. miesiącu życia niemowlę przestaje rozpoznawać kontrasty fonetyczne, których brakuje w języku ojczystym. W rezultacie Japończycy nie słyszą różnicy pomiędzy *ra/la*, Hiszpanie *vi/bi*, Polacy ważnego w języku angielskim rozróżnienia *s/th*, ludzie niestykający się z językami tonalnymi (dialekty chińskiego, języki południowo-wschodniej Azji, liczne języki afrykańskie i południowo-amerykańskie) nie dostrzegają istotnych informacji leksykalnych i gramatycznych ukrytych w tonach. Biernie oglądanie programów telewizyjnych czy słuchanie taśm nie wpływają na postrzeganie istotnych różnic fonetycznych w językach obcych. Panuje przekonanie (znane pod nazwą Social Gating Hypothesis – Kuhl, 2007), że do nauki fonetyki danego języka konieczne są kontakty z opiekunami mówiącymi tym językiem.

Badania sprawdzające zdolności niemowląt do rozróżniania dźwięków mogą być prowadzone w różny sposób, ale odnoszą się do tych samych zdolności, można je więc poddać metaanalizie. Zwykle stosuje się procedurę określaną jako Preferencje Kierunku Spojrzenia (Headturn Preference Paradigm, HPP), lub procedurę Uwarunkowanego Obrótu Głowy (Conditioned Head-Turning, CHT) opisaną w pracy Janet Werker i współautorów (Werker, Polka, Pegg, 1997) i używaną między innymi przez Patrycję Kuhl i współpracowników (Kuhl i in., 2006). Dzieci siedzą na kolanach swoich opiekunów, a bodźce akustyczne (np. sylaby *la* albo *ra*) podawane są przez słuchawki lub głośniki ustawione po obu stronach głowy. Po okresie familiaryzacji, w którym bodziec audiowizualny pojawia się jednostronnie tak długo, aż dziecko przestaje się nim interesować, następuje zmiana bodźca i zmiana strony, po której pojawia się bodziec wizualny. Uwarunkowanie na zmianę bodźca pozwala zbadać, czy dziecko rozpoznaje różnice fonemów, które są istotne dla rozumienia danego języka. W czasie testu, zależnie od tego, czy jest to ten sam bodziec, czy nowy (np. sylaba lub słowo ją zawierające), wybierany w sposób losowy, notowany



Rysunek 1. Rozkłady wartości efektu zmierzonego w różnych publikacjach uporządkowane wg rosnących wartości efektu



jest kierunek spojrzenia, który powinien się zmienić przy pojawieniu się nowego bodźca. Obecnie głównie wykorzystane są okulometry automatycznie zapisujące kierunek spojrzenia, zsynchronizowane z wideorejestracją zachowania dziecka. Dawniej notowano kierunek obrotu głowy. Stosuje się też procedurę wzrokowej habituacji (*visual habituation, central fixation*), w której bodziec audiowizualny pojawia się centralnie, i mierzy się czas habituacji, a w fazie testowej ocenia się reakcje na ten sam bodziec lub różne bodźce.

Mierzone są też bezpośrednio reakcje mózgu na prezentowane bodźce. Elektroencefalografia jest tu najczęściej stosowaną metodą, chociaż technicznie nie jest łatwa, wymaga założenia dziecku czepka z wieloma elektrodami, oraz żmudnej procedury czyszczenia sygnału EEG z licznych artefaktów wywołanych ruchami głowy i skurczami mięśni. Analizuje się potencjały wywołane (*Event Related Potentials, ERP*), czyli wielokrotnie uśrednione zapisy EEG w okienku czasowym bliskim pojawienia się bodźca, oraz „falę niezgodności” (*Mismatch Negativity Response*), pojawiającą się u niemowląt około 200–300 milisekund po bodźcu, automatyczną reakcję na pojawienie się nowego bodźca.

Stanowisko do badania reakcji niemowląt pozwala na filmowanie sesji, wykorzystuje szeroki ekran, na którym pojawiają się postacie z kreskówek, słuchawki, przez które podawane są słowa różniące się różnymi fonemami oraz pomiar kierunku spojrzenia za pomocą kulometru. By ocenić reakcje mózgu, a nie tylko reakcje behawioralne, używaliśmy też 64-elektrodowe EEG i zaawansowane metody analizy sygnału (Aghabeig i in., 2019). W ten sposób można testować zdolności do rozróżniania wyrazów zawierających trudne do odróżnienia fonemy zarówno w języku ojczystym, jak i w językach o odmiennej strukturze fonetycznej.

Metaanalizie poddano obecnie 33 publikacje, w których opisano 145 eksperymentów z udziałem 2491 dzieci (wstępne rezultaty zob. Tsuji, Cristia, 2014). W każdym przypadku w czasie prezentacji sekwencji sylab zbieramy informację o kierunku spojrzenia w tę samą lub

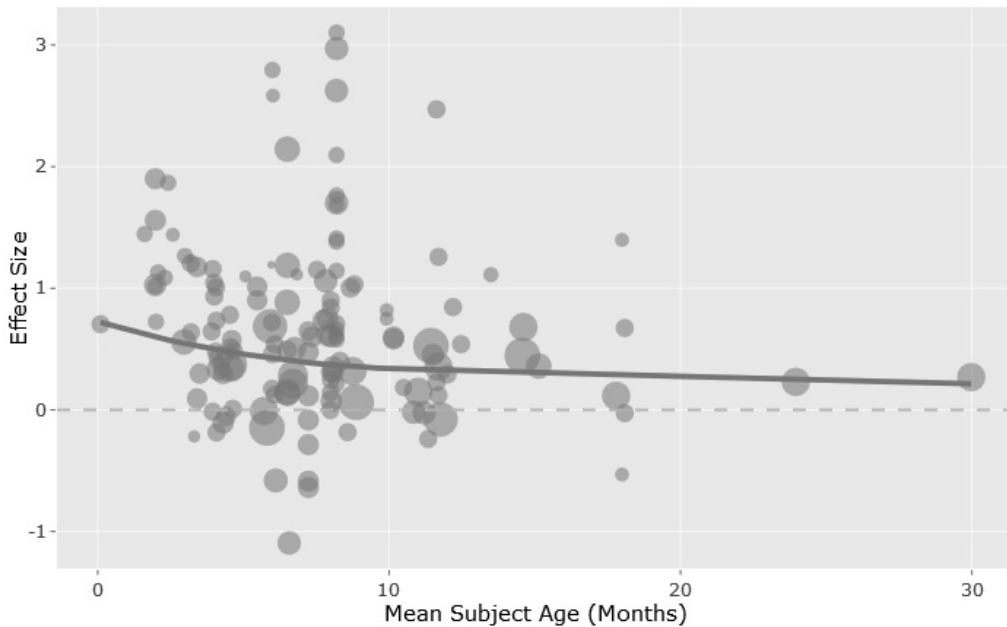
przeciętną stronę, po której pojawił się ciekawy bodziec. To pozwala zobaczyć, czy dziecko potrafi rozpoznać i skojarzyć specyficzny bodziec dźwiękowy z nagrodą. Na tej podstawie możemy obliczyć różne miary wielkości efektu, czyli reakcji na natywne i obce: d , g , r oraz logarytm ilorazu szans L .

Na rysunku 1 przedstawiono wyniki dla każdego ze 145 eksperymentów, w których otrzymano wielkości efektów wraz z przedziałami ufności. Wielkość efektu oceniana była na podstawie ułamka prawidłowych orientacji przy pojawieniu się nowego bodźca lub różnic amplitud w pomiarach ERP.

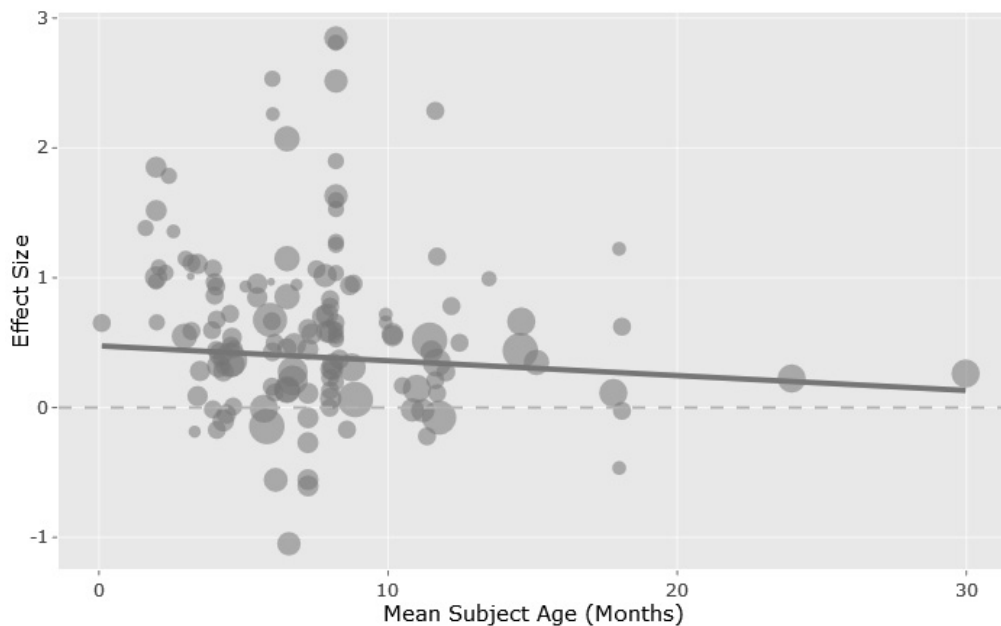
Wariancja metamodelu $\pm .08$ widoczna jest na poziomej linii na poziomie .58. Wszystkie wykresy wygenerowano za pomocą narzędzi MetaLab (<http://metalab.stanford.edu>).

Model metaanalityczny stworzony na podstawie tych wyników pokazuje ocenianą wartość efektu $.58 \pm .08$. Pozioma linia jest na poziomie .58. Jak widać, rozbieżność wyników jest duża, również wariancja każdego z badań jest znaczna (liczba badanych dzieci jest zwykle mała), ale wariancja metamodelu jest niewielka. Przyczyny takich rozbieżności wyjaśnia teoria PRIMIR (Processing Rich Information from Multidimensional Interactive Representations) opracowana przez Janet Werker i Suzanne Curtin (2005) i rozwijana w pracy Curtin i Werker (2018). Teoria ta opiera się na hipotezie różnych mechanizmów uczenia się analizy bogatej informacji sensorycznej, wrodzonych preferencji do zwracania uwagi na różne aspekty bodźców akustycznych i poszczególnych wpływów środowiskowych, co w sumie wpływa na powstanie różnych trajektorii nabywania kompetencji językowych. W różnych publikacjach badane są odmienne kontrasty fonetyczne, co może również mieć wpływ na wyniki. Dodatkową przyczyną może być uwzględnienie w metaanalizach obok rezultatów badań behawioralnych wyników badań EEG i NIRS (Near InfraRed Spectroscopy), które stanowiły około $\frac{1}{4}$ wszystkich analizowanych wyników.

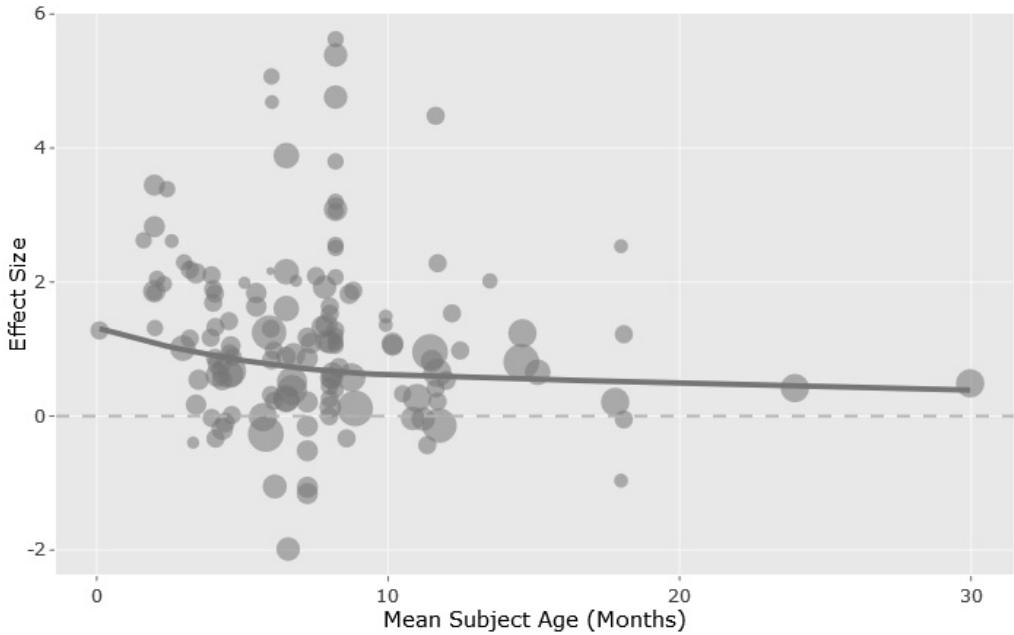
Na rysunkach 2–5 pokazano wykresy wielkości efektu w zależności od wieku, generowane przez serwer MetaLab na podstawie zgromadzonej bazy danych.



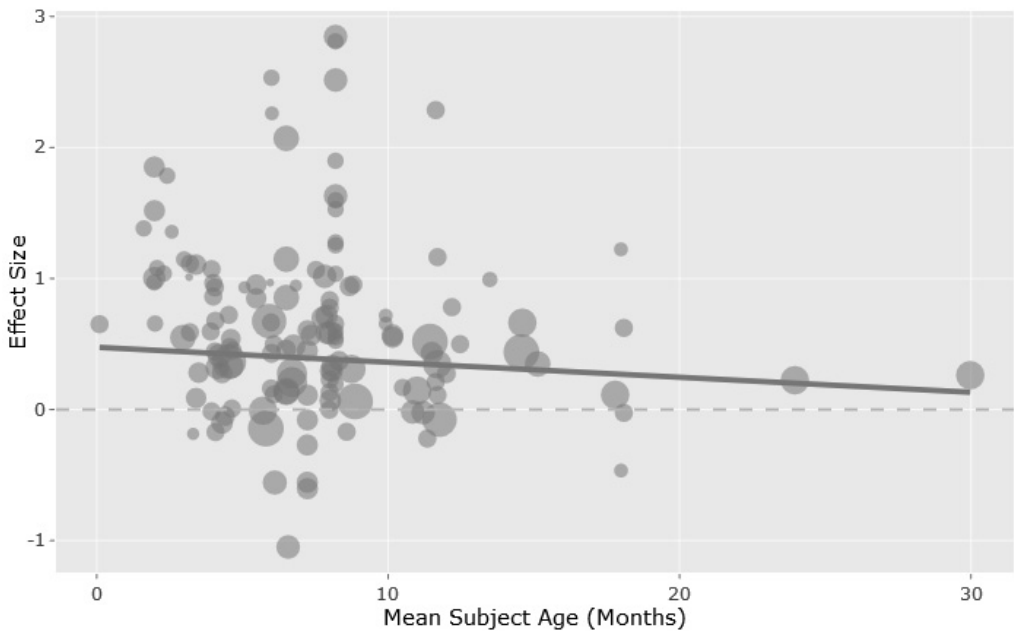
Rysunek 2. Wykres rozrzutu wartości d Cohena w zależności od wieku. Krzywa pokazuje lokalną liniową regresję (*loess*). Jednakże rozrzut wielkości efektu jest bardzo duży



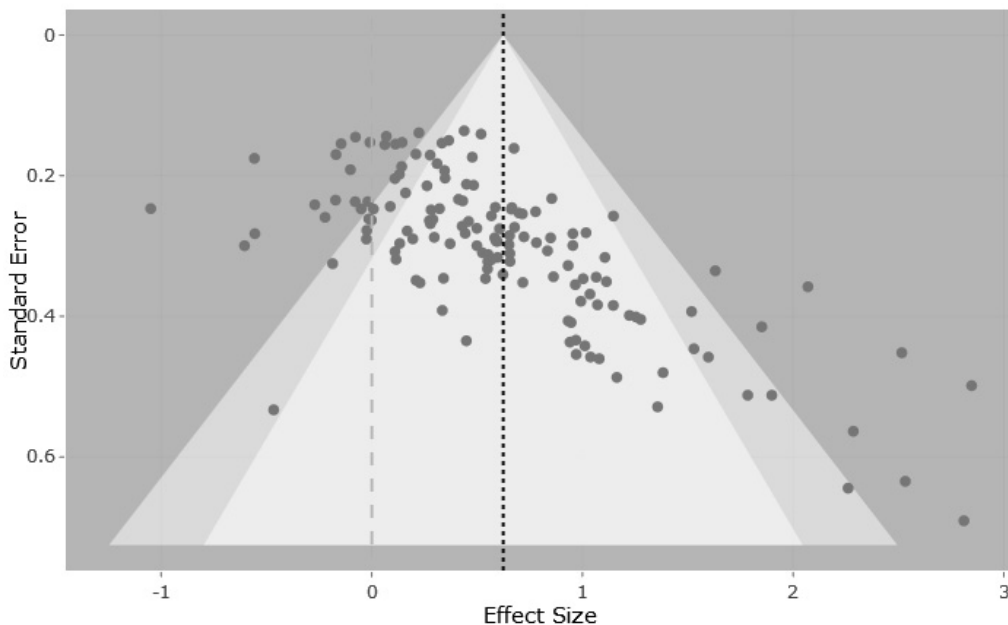
Rysunek 3. Wykres rozrzutu wartości g Hedgesa w zależności od wieku jest bardzo podobny do d Cohena. Prosta pokazuje dopasowanie ważonej regresji liniowej



Rysunek 4. Wykres rozrzutu wartości r Pearsona w zależności od wieku. Prosta pokazuje dopasowanie ważonej regresji liniowej



Rysunek 5. Wykres rozrzutu wartości logarytm ilorazu szans L w zależności od wieku. Krzywa pokazuje lokalną liniową regresję (*loess*)



Rysunek 6. Wykres lejkowy zależności standardowego błędu od wielkości efektów. Obszar dla $p < .01$ jest jasny, $p < .05$ lekko zaciemniony, a $p > .05$ ciemniejszy. Badania wyraźnie odbiegające od średniego efektu obarczone są dużym błędem standardowym i są mniej wiarygodne

W przypadku rozróżniania samogłosek natywnych wszystkie cztery miary dają podobne wykresy rozrzutu. Rozmiar kółek jest odwrotnie proporcjonalny do standardowego błędziego do wielkości efektu. Metaanaliza pokazuje powolne zmniejszanie się efektu z wiekiem, ale asymptotycznie efekt utrzymuje się do 30. miesiąca (taki był średni wiek najstarszej badanej grupy). Modele metaanalityczne usuwają wyniki powyżej i poniżej trzech odchyłeń standardowych od średniej wszystkich wyników. Ponieważ wyniki dla różnych miar są, jak widać, bardzo podobne, wystarczy podawać tylko wyniki dla miary d Cohena. Najwięcej badań dotyczyło niemowląt do 10. miesiąca życia, i tu obserwujemy bardzo duży rozrzut wielkości obserwowanych efektów.

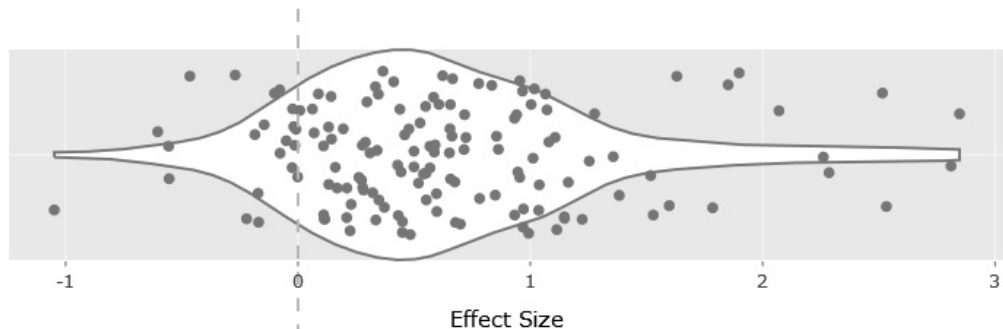
Wykresy „lejkowe” (rysunek 6) ilustrują ukryte tendencje (*bias*), pokazując standardowe błędy w zależności od wielkości efektu.

Badania o największej dokładności powinny być blisko średniej wielkości efektu. Jeśli w analizowanych badaniach nie ma ukrytych

tendencji, to wyniki powinny być rozłożone symetrycznie względem średniej. Widoczna asymetria pokazuje, że im większy błąd, tym większe rozmiary efektu, co podkreśla tendencyjność publikacji. MetaLab pozwala również rozróżnić wyniki w zależności od metody badania. Badania oparte na elektroencefalografii wskazują na zmniejszanie się efektu z wiekiem, badania z użyciem okulometrii przewidują stabilnie średni efekt niezależnie od wieku w granicach 5–30 miesięcy, a badania behawioralne uwidaczniają wyraźny efekt na poziomie $d = 1$, ale były ograniczone do 8. miesiąca życia.

Innym sposobem ilustracji wyników wielu publikacji jest ocena liczby publikacji (lub prawdopodobieństwa wartości efektu określonego na podstawie liczby publikacji taki efekt raportujących) za pomocą „wykresów skrzypcowych” (*violin plots*), zawierających podobne informacje jak histogramy.

Jak widać na rysunku 7, publikacji podających wartości znacznie odbiegające od średniej jest stosunkowo niewiele.

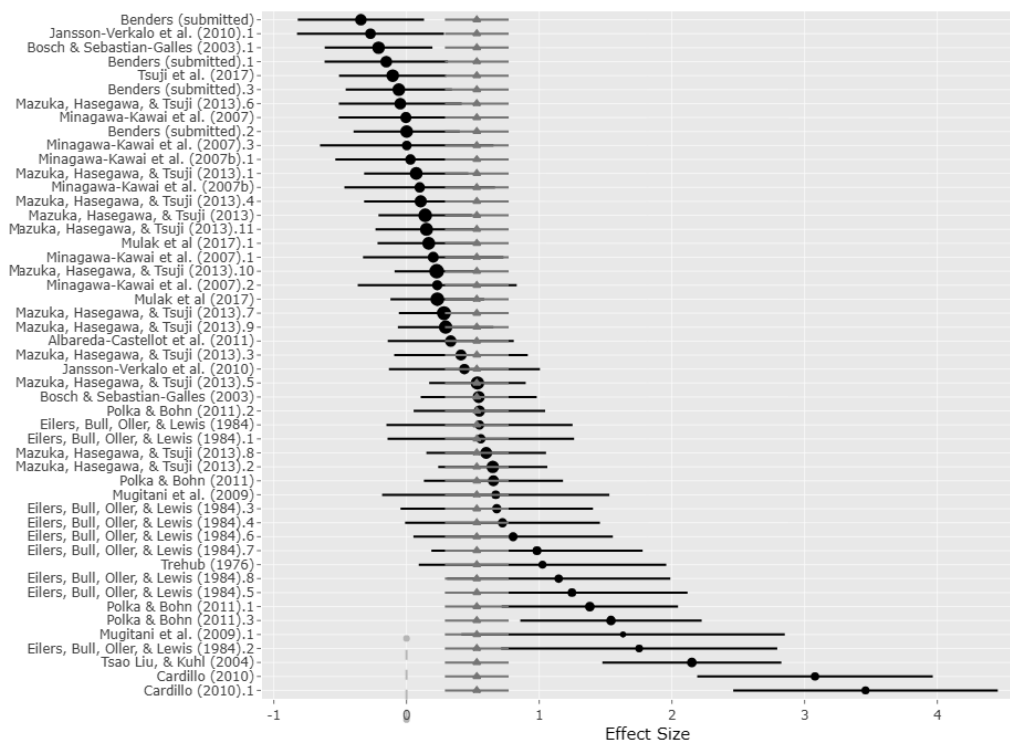


Rysunek 7. Wykres skrzypcowy pokazujący, jak często raportowane są wielkości efektu w różnych pracach. Najgrubszy odcinek wskazuje, jakie wartości efektów występują najczęściej w publikacjach

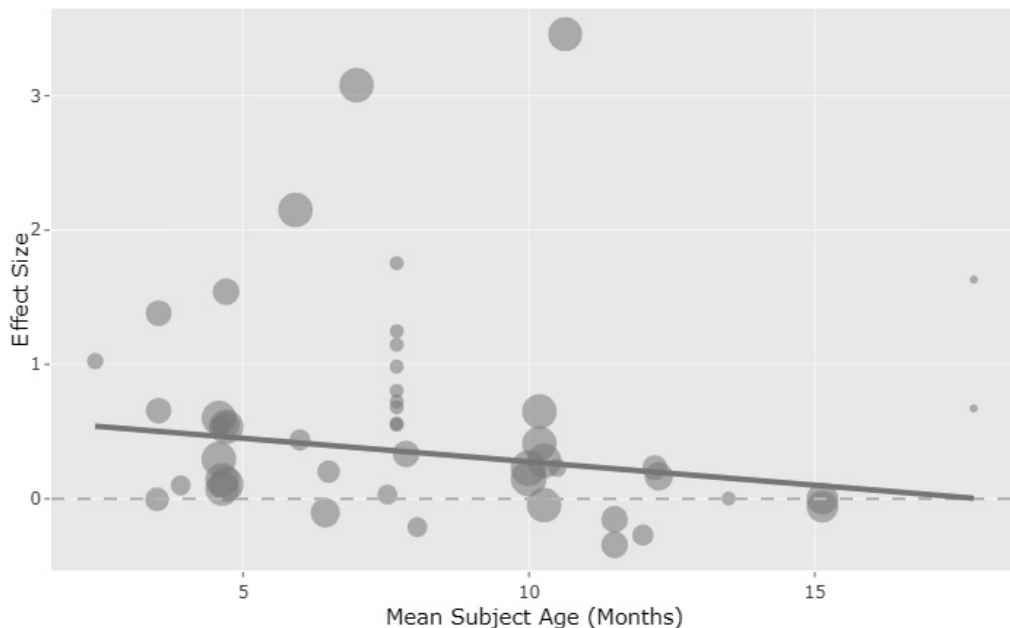
ROZRÓŻNIANIE SAMOGŁOSEK NIENATYWNYCH

W tym przypadku mamy znacznie mniej wyników. Analizie poddano 15 publikacji (od 1984 do 2017 roku), w których opisano 49 ekspery-

mentów z udziałem 573 dzieci. Średnia wielkość efektu jest na poziomie $.64 \pm .23$, a więc nieco większa niż w przypadku samogłosek natywnych, ale za to z dużo większym odchyleniem standardowym. Siedem eksperymentów dało ujemny wynik.



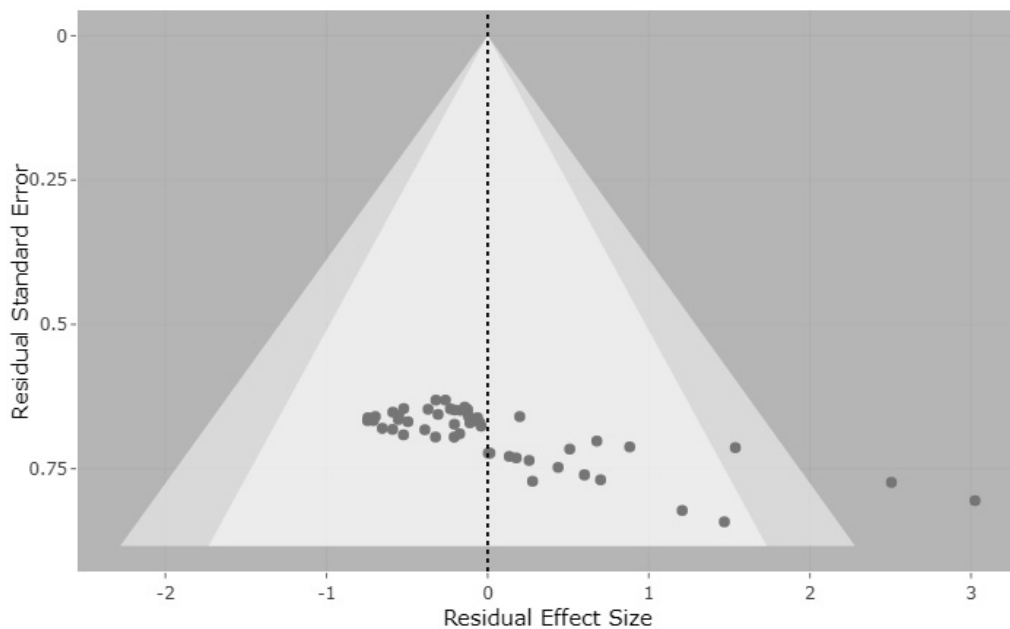
Rysunek 8. Wyniki dla 49 eksperymentów wraz z przedziałami ufności



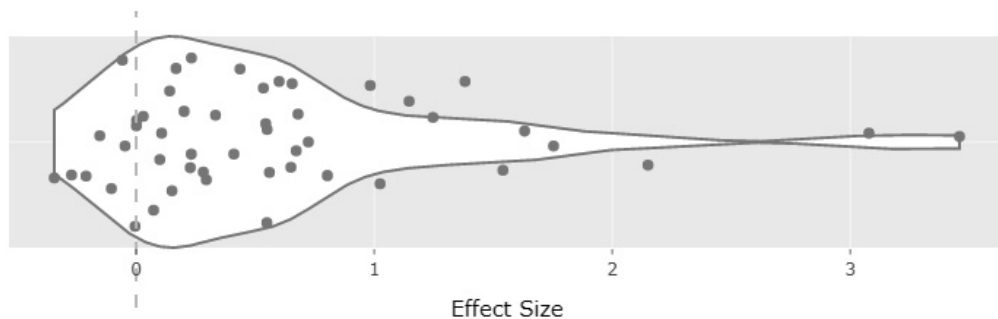
Rysunek 9. Wielkość efektu w zależności od wieku dla rozróżniania samogłosek nienatywnych

Na rysunku 9 pokazano wykres wielkości efektu rozróżniania samogłosek nienatywnych w zależności od wieku, generowane przez ser-

wer MetaLab na podstawie zgromadzonej bazy danych z 49 eksperymentów. Efekty powyżej trzech dotyczyły rozróżniania *u-y* w języku



Rysunek 10. Wykres lejkowy zależności standardowego błędu od wielkości efektów



Rysunek 11. Wykres skrzypcowy pokazuje, jak często raportowane są wielkości efektu w różnych pracach. Najgłębszy odcinek pokazuje, jakie wartości efektów występują najczęściej w publikacjach

angielskim i pochodzą z pracy doktorskiej G. Cardillo. Ponieważ wyraźnie odstają od pozostałych, nie mają wpływu na prostą otrzymaną z liniowej regresji, wskazującą na zmniejszanie się efektu, czyli zdolności do rozróżnień samogłosek nieobecnych w języku ojczystym wraz z wiekiem. Wykres lejkowy na rysunku 10 pokazuje te dwa odstające badania poza obszarem dla $p < .05$. Podobnie z wykresu skrzypcowego na rysunku 11 wynika, że tylko 3 eksperymenty dały wyniki powyżej $d = 2$.

Obszar dla $p < .01$ jest jasny, $p < .05$ lekko zacieniony, a $p > .05$ ciemniejszy.

DYSKUSJA

W omawianych przykładach prac związanych z nabywaniem języka w okresie niemowlęctwa, a w szczególności rozróżniania samogłosek natywnych i nienatywnych, uwidaczniają się liczne trudności tego typu badań. Choć w ostatnich dziesięcioleciach uzyskano liczne rezultaty przy użyciu nowoczesnych narzędzi, technik i metod badawczych, włączając w to badania behawioralne, analizę EEG i sygnałów NIRS, są one często kontrowersyjne, a efekty słabe i trudne do replikacji. Cel, jakim jest zrozumienie rozwoju procesów poznawczych w powiązaniu ze zmianami w mózgu niemowląt, jest nadal odległy. Wyniki wielu opublikowanych badań zebranych przez konsorcjum ManyBabies, koordynowane przez Uniwersytet Stanforda, pozwalają na dokonanie metaanaliz dla znacznie większej grupy badanych. Statystyczne metody

metaanaliz są już dobrze rozwinięte (Hedges, Olkin, 2014). Narzędzia do takich analiz udostępniono w ramach projektu MetaLab.

W indywidualnych badaniach zgromadzonych w repozytorium MetaLab, związanych z rozróżnianiem samogłosek, średni rozmiar grupy wynosił zaledwie 18 osób. Moc pojedynczych testów nie jest więc duża. Rozrzut wyników eksperymentów jest za to bardzo szeroki. Metaanalizy obejmują znacznie większe grupy. W przypadku pierwszego z omawianych tu efektów opisanych w tej pracy zjawisk było to prawie 2500, a drugiego prawie 600. Średnia wielkość efektu wynosiła $d = .58$ i $.64$. Na wykresach lejkowych widać korelację pomiędzy wielkością błędu a rozmiarem efektu, obrazującą tendencyjność niektórych opublikowanych wyników. Na wykresach skrzypcowych można zobaczyć, że większość publikacji otrzymała efekty o wielkości zbliżonej do średniej, a więc efekty są replikowalne. Dostępność takich narzędzi i danych odpowiednio przygotowanych do analiz umożliwia zastosowanie wyrafinowanych metod statystycznych (np. Tsuji, Cristia, 2014), które nie zostały tu omówione.

Metaanalizy potwierdziły, że w rozwoju percepcji widoczne są efekty związane z zanikiem zdolności do rozróżniania nienatywnych samogłosek wraz z wiekiem, ale wiek nie jest jedynym czynnikiem wpływającym na wyniki tych badań. Trudno jest rozpatrywać takie efekty w izolacji od ogólnych procesów rozwojowych. Dla nabywania języka ojczystego ważny jest rozwój behawioralny w zakresie różnicowania dźwięków mowy, naturalnych dźwięków dochodzących ze środowiska. Duża

neuroplastyczność młodego mózgu pozwala na formowanie się obwodów neuronalnych umożliwiających percepcję dźwięków mowy w postaci quasi-dyskretnych fonemów (Meltzoff i in., 2009). Przegląd literatury w przypadku treningu słuchu pokazał, że nie ma prostej reguły generalizacji percepcji bodźców akustycznych (Wright, Zhang, 2009). Podstawą realizacji mechanizmów nabywania języka oraz rozwijania kompetencji społecznych jest osiągnięcie odpowiedniego poziomu neuronalnej bazy **realizującej procesy uwagowe** i pamięci. W tym przypadku, zwłaszcza we wczesnym okresie rozwoju, chodzi głównie o uwagę wzrokową oraz pamięć roboczą (Amso, Scerif, 2015). Zwracanie uwagi na bodźce podawane w czasie eksperymentu jest jednym z problemów prowadzenia badań nad nabywaniem języka. Brak reakcji może być spowodowany rozproszeniem lub odwróceniem uwagi od zadania. Jest to duże wyzwanie i może być wyjaśnieniem rozbieżności wyników różnych publikacji.

Metaanalizy nie rozwiązują oczywiście wszystkich problemów dotyczących badań nad rozwojem percepcji niemowląt. Zbyt wiele czynników, oprócz wieku, może wpłynąć na wyniki tych badań. Średnie efekty dla dużej populacji badanych nie zawsze są bliskie wynikom dla mniejszych grup, zwłaszcza przy stosowaniu różnych metod badawczych. W przypadku percepcji fonemów natywnych i obcych były to badania behawioralne oparte na wariantach procedury Preferencji Kierunku Spojrzenia (HPP), Uwarunkowanego Obrótu Głowy (CHT), okulometrii, analizie potencjałów wywołanych ERP mierzonych za pomocą różnych systemów EEG,

oraz nielicznych badaniach z użyciem NIRS. Metaanalizy badań opartych na tak różnych procedurach są nadal możliwe, chociaż stosowanie różnych procedur zwiększa wariację uśrednionych wyników. Głębsze zrozumienie procesów rozwoju umiejętności językowych wymaga odwołania się do takich teorii, jak na przykład PRIMIR (Werker, Curtin, 2005), zakładających, że mózgi dzieci mogą nauczyć się wykorzystywać bogatą informację zawartą w sygnałach mowy na różny sposób. Zależy to od wrodzonych preferencji, reakcji na wzorce rytmiczne, percepcji wysokości dźwięków, stymulacji ze strony środowiska i struktury języka ojczystego. Jest wiele innych modeli dotyczących ściśle percepcji mowy (omówionych w pracy Werker i Curtin, 2005). PRIMIR oferuje szerszą perspektywę mechanizmów uczenia się mózgow, opartych na statystyce, oraz wynikające stąd rozwoju zdolności poznawczych.

Takie zindywidualizowane ujęcie nie da się łatwo pogodzić z metaanalizami badań różnego typu. Mechanizmy uwagowe oraz pamięciowe w rozwijającym się mózgu będą określały mechanizmy uczenia dla nabywania języka ojczystego oraz rozwijania kompetencji społecznych. Różnicowanie będzie dotyczyło rozwojowego kształtowania się uwagi i pamięci oraz sposobów jego badania i wyjaśniania (Amso, Scerif, 2015; D'Esposito, Postle, 2015; Constantinidis, Klingberg, 2016; Reynolds, Romano, 2016). Badania powinny więc skupiać się na sieciach uwagowych w kontekście uczenia się (Posner, Rothbart, 2007), mających również znaczenie dla wyjaśniania procesów rozwoju przez całe życie (Graysona, Fairc, 2017).

PRZYPISY

¹ Badania sfinansowane w ramach projektu „NeuroPerKog: rozwój słuchu fonematycznego i pamięci roboczej u niemowląt i dzieci” ze środków Narodowego Centrum Nauki przyznanych na podstawie umowy nr UMO-2013/08/W/HS6/00333.

² <http://metalab.stanford.edu> (dostęp: 16.08.2019).

³ <https://improvingpsych.org/> (dostęp: 16.08.2019).

⁴ <https://psysciacc.org/> (dostęp: 16.08.2019).

⁵ <https://osf.io/gf7vh/wiki/home/> (dostęp: 16.08.2019).

⁶ Język natywny to język występujący w naturalnych warunkach życia. Jeśli na przykład dziecko ma nianię, która mówi do niego w języku innym niż rodzice, którzy przez większość dnia przebywają w pracy, to językiem natywnym dla tego dziecka będzie język jego niani (przyp. red.).

BIBLIOGRAFIA

- Aghabeg M., Bałaj B., Dreszer J., Lewandowska M., Milner R., Pawlaczyk N., Piotrowski T., Szmytke N., Duch W. (2019), *Perception of Non-Native Phoneme Contrasts in 8–13 Months Infants: Tensor-Based Analysis of EEG Signals*. 27th European Signal Processing Conference, EUSIPCO 2019, Corna, Spain, 2–6 Sept. 2019, 2251–2256.
- Amso D., Scerif G. (2015), The Attentive Brain: Insights from Developmental Cognitive Neuroscience. *Nature Reviews. Neuroscience*, 16(10), 606–619.
- Bergmann C., Tsuji S., Cristia A. (2017), Top-Down versus Bottom-up Theories of Phonological Acquisition: A Big Data Approach. *Proc. Interspeech*, 2103–2107, doi: 10.21437/Interspeech.2017-1443.
- Bergmann C., Tsuji S., Piccinini P.E., Lewis M.L., Braginsky M., Frank M.C., Cristia A. (2018), Promoting Replicability in Developmental Research through Meta-analyses: Insights from Language Acquisition Research. *Child Development*, 89(6), 1996–2009.
- Button K.S., Ioannidis J.P.A., Mokrysz C., Nosek B.A., Flint J., Robinson E.S.J., Munafò M.R. (2013), Power Failure: Why Small Sample Size Undermines the Reliability of Neuroscience. *Nature Reviews Neuroscience*, 14(5), 365–376.
- Byers-Heinlein K., Bergmann C., Davies C., Frank M.C., Hamlin K., Kline M., Kominsky J.F., Kosie J.E., Lew-Williams C., Liu L., Mastroberardino M., Singh L., Waddell C.P.G., Zettersten M., Soderstrom M. (2019), Building a Collaborative Psychological Science: Lessons Learned from ManyBabies 1. *PsyArXiv*, <https://doi.org/10.31234/osf.io/dmhk2>.
- Constantinidis C., Klingberg T. (2016), The Neuroscience of Working Memory Capacity and Training. *Nature Review Neuroscience*, 26 May, 438–449.
- Curtin S., Werker J.F. (2018), PRIMIR on Tone. *Frontiers in Psychology*, 9, 1007, doi:10.3389/fpsyg.2018.01007.
- D’Esposito M., Postle B.R. (2015), The Cognitive Neuroscience of Working Memory. *Annual Review of Psychology*, 66, 115–142.
- Ebersole C.R., Atherton O.E., Belanger A.L., Skulborstad H.M. [...] (64 autorów) (2016), ManyLabs 3: Evaluating Participant Pool Quality Across the Academic Semester Via Replication. *Journal of Experimental Social Psychology*, 67, 68–82.
- Frank M.C., Bergelson E., Bergmann C., Cristia A., Floccia C., Gervain J. [...] Yurovsky D. (2017), A Collaborative Approach to Infant Research: Promoting Reproducibility, Best Practices, and Theory-Building. *Infancy*, 22(4), 421–435. <https://doi.org/10.1111/infa.12182>.
- Goswami U. (2015), *Children’s Cognitive Development and Learning*. York: Cambridge Primary Review Trust.
- Grayson D.S., Fairc D.A. (2017), Development of Large-Scale Functional Networks from Birth to Adulthood: A Guide to the Neuroimaging Literatur. *NeuroImage*, January.
- Hedges L., Olkin I. (2014), *Statistical Methods for Meta-Analysis*. Cambridge, MA: Academic Press.
- Kaplan P.S., Goldstein M.H., Huckey E.R., Owren M.J., Cooper R.P. (1995), Dishabituation of visual attention by infant- versus adult-directed speech: Effects of frequency modulation and spectral composition. *Infant Behavior and Development*, 18(2), 209–223. [https://doi.org/10.1016/0163-6383\(95\)90050-0](https://doi.org/10.1016/0163-6383(95)90050-0).
- Keunen K., Counsell S.J., Benders M.J.N.L. (2017), The Emergence of Functional Architecture during Early Brain Development. *NeuroImage*, 160, 2–14.
- Klein R.A., Ratliff K.A., Vianello M. [...] Nosek B.A. (51 autorów) (2014), Investigating Variation in Replicability: A “Many Labs” Replication Project. *Social Psychology*, 45(3), 142–152.
- Klein R.A., Vianello M., Hasselman F., Adams B.G. [...] Nosek B.A. (190 autorów) (2018), Many Labs 2: Investigating Variation in Replicability Across Samples and Settings. *Advances in Methods and Practices in Psychological Science*, 1(4), 443–490.
- Kuhl P.K. (2007), Is Speech Learning ‘Gated’ By The Social Brain? *Developmental Science*, 10(1), 110–120.
- Kuhl P.K. (2010), Brain Mechanisms in Early Language Acquisition. *Neuron*, 67(5), 713–727.
- Kuhl P.K. (2011), Early Language Learning and Literacy: Neuroscience Implications for Education. *Mind, Brain, and Education*, 5, 128–142.
- Kuhl P.K., Stevens E., Hayashi A., Deguchi T., Kiritani S., Iverson P. (2006), Infants Show a Facilitation Effect for Native Language Phonetic Perception Between 6 and 12 Months. *Developmental Science*, 9(2), F13–F21.

- Lenneberg E.H. (1967), *Biological Foundations of Language*. New York: John Wiley and Sons.
- Lewis M., Braginsky M., Tsuji S., Bergmann C., Piccinini P.E., Cristia A., Frank M.C. (2016), A Quantitative Synthesis of Early Language Acquisition Using Meta-Analysis [Preprint]. *PsyArXiv*, <https://psyarxiv.com/htsjm>.
- Makropoulos A., Robinson E.C., Schuh A., Wright R., Fitzgibbon S., Bozek J. [...] Rueckert D. (2018), The Developing Human Connectome Project: A Minimal Processing Pipeline for Neonatal Cortical Surface Reconstruction. *NeuroImage*, 173, 88–112.
- Meltzoff A.N., Kuhl P.K., Movellan J., Sejnowski T.J. (2009), Foundations for a new science of learning. *Science* (New York, N.Y.), 325(5938), 284–288. <https://doi.org/10.1126/science.1175626>
- Moshontz H., Campbell L., Ebersole C.R., IJzerman H., Urry H.L., Forscher P.S. [...] Chartier C.R. (2018), The Psychological Science Accelerator: Advancing Psychology Through a Distributed Collaborative Network. *Advances in Methods and Practices in Psychological Science*, 1(4), 501–515.
- Nelson L.D., Simmons J., Simonsohn U. (2018), Psychology’s Renaissance. *Annual Review of Psychology*, 69(1), 511–534.
- Newman R.S., Hussain I. (2006), Changes in Preference for Infant-Directed Speech in Low and Moderate Noise by 4.5- to 13-Month-Olds. *Infancy*, 10(1), 61–76. https://doi.org/10.1207/s15327078in1001_4.
- Posner M.I., Rothbart M.K. (2007), Research on Attention Networks as a Model for the Integration of Psychological Science. *Annual Review of Psychology*, 58(1), 1–23.
- Reynolds G.D., Romano A.C. (2016), The Development of Attention Systems and Working Memory in Infancy. *Frontiers in Systems Neuroscience*, 10, 15.
- Saxton M. (2010), *Child Language: Acquisition and Development*. Los Angeles: SAGE Publications Ltd.
- Schirmer M., McIntosh A.R., Jirsa V., Deco G., Ritter P. (2018), Inferring Multi-scale Neural Mechanisms with Brain Network Modelling. *ELife*, 7, <https://doi.org/10.7554/eLife.28927>.
- Simmons J.P., Nelson L.D., Simonsohn U. (2011), False-positive Psychology: Undisclosed Flexibility in Data Collection and Analysis allows Presenting Anything as Significant. *Psychological Science*, 22(11), 1359–1366. <https://doi.org/10.1177/0956797611417632>.
- Stanley D.J., Spence J.R. (2014), Expectations for Replications: Are Yours Realistic? *Perspectives on Psychological Science*, 9, 305–318.
- Stefanovski L., Ghani A., McIntosh A.R., Ritter P. (2016), Linking Connectomics and Dynamics in the Human Brain. *E-Neuroforum*, 7(3), 64–70.
- Tsuji S., Cristia A. (2014), Perceptual Attunement in Vowels: A Meta-analysis. *Developmental Psychobiology*, 56(2), 179–191.
- Werker J.F., Curtin S. (2005), PRIMIR: A Developmental Framework of Infant Speech Processing. *Language Learning and Development*, 1(2), 197–234.
- Werker J.F., Tees R.C. (1984), Cross-Language Speech Perception: Evidence for Perceptual Reorganization during the First Year of Life. *Infant Behavior & Development*, 7, 49–63.
- Werker J.F., Polka L., Pegg J.E. (1997), The Conditioned Head Turn Procedure As a Method for Testing Infant Speech Perception. *Early Development and Parenting*, 6(34), 171–178.
- Wright B.A., Zhang Y. (2009), A Review of the Generalization of Auditory Learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1515), 301–311.