

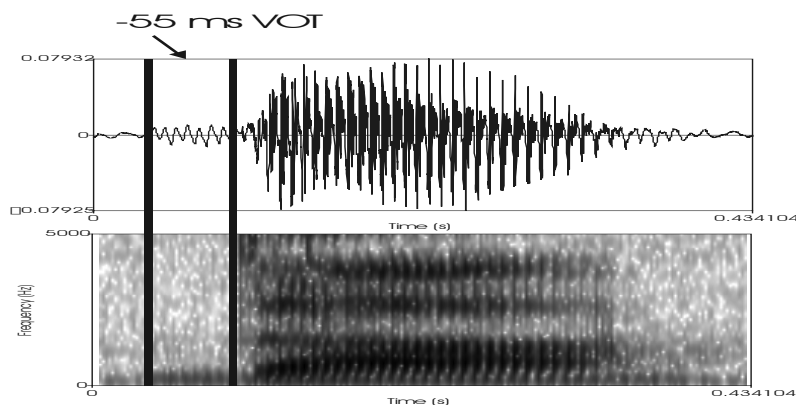
Arkadiusz Rojczyk  
Uniwersytet Śląski  
Katowice

### PARAMETR VOT W JĘZYKU POLSKIM I ANGIELSKIM. BADANIE PERCEPCJI

*Voice Onset Time* (VOT) jest parametrem wprowadzonym przez Leigh Liskera i Arthura Abramsona (1964) w celu klasyfikacji opozycji dźwięczna–bezdźwięczna w nagłosowych spółgłoskach zwarto-wybuchowych. Definiowany jest on jako czas między wybuchem spółgłoski a rozpoczęciem sygnału periodycznego lub też jako „odstępstwo czasowe pomiędzy wybuchem, który sygnalizuje koniec zwarcia spółgłoski, a początkiem periodyczności, która odzwierciedla wibrację krtaniową” (Lisker, Abramson 1964: 422). W badaniu na próbie dwudziestu trzech języków L. Lisker i A. Abramson zauważyli, że nagłosowe zwarto-wybuchowe konsekwentnie wykorzystują maksymalnie dwie z trzech kategorii VOT.

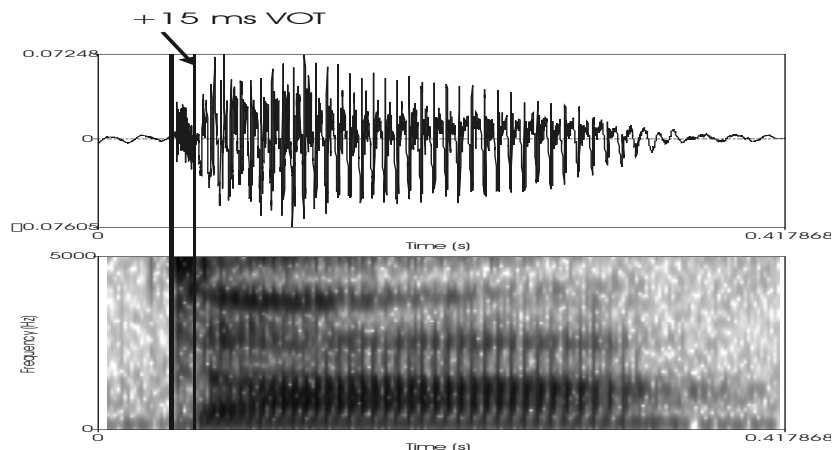
1. Wartości ujemne VOT – periodyczność rozpoczyna się przed wybuchem spółgłoski (około -30 ms lub więcej). Dają one obraz spektralny, w którym widoczne są jako koncentracja harmonicznej energii na niskich poziomach, wskazująca na wibrację strun głosowych (Keating i in. 1981; Lisker 1986). Rycina 1 przedstawia oscylogram i spektrogram polskiej sylaby *da* z wartością parametru VOT -55 ms.

**Rycina 1.** Oscylogram i spektrogram polskiej sylaby *da*. Wartość parametru VOT -55 ms



2. Wartości krótkie VOT – periodyczność rozpoczyna się w momencie lub krótko po wybuchu (około 0 ms do +30 ms (Keating 1984)). Rycina 2 przedstawia oscylogram i spektrogram polskiej sylaby *ta* z wartością parametru VOT +15 ms<sup>1</sup>.

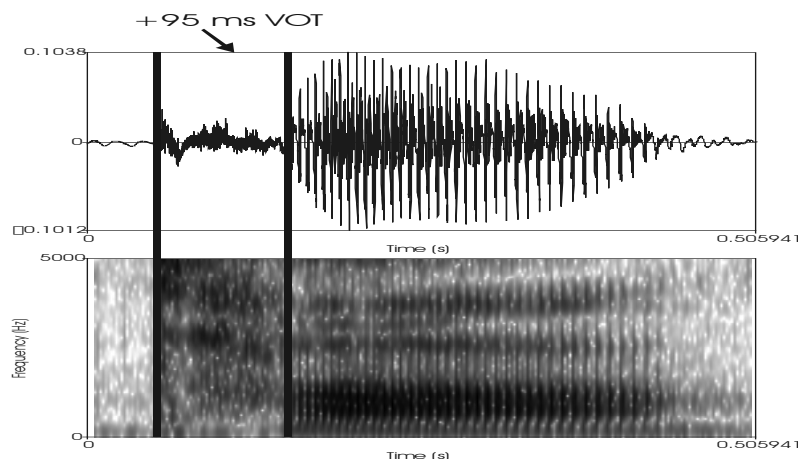
**Rycina 2.** Oscylogram i spektrogram polskiej sylaby *ta*. Wartość parametru VOT +15 ms



3. Wartości długie VOT – periodyczność rozpoczyna się z dużym odstępstwem czasowym od wybuchu (około +50 ms lub więcej). Okres ten może być wypełniony ciszą (Klatt 1975; Lisker 1986) lub aspiracją, jeśli wzburzone powietrze przechodzące przez głośnię daje efekt rezonancyjny (Lisker, Abramson 1964). Aspiracja jest rejestrowana akustycznie jako „szum w częstotliwościach drugiego i trzeciego formantu” (Lisker, Abramson 1964: 386), „abdukcja krtaniowa w momencie wybuchu spółgłoski” (Jansen 2004: 41), „turbulencja w rejonie górnej części aparatu mowy” (Abramson 1977: 296), „turbulencja aerodynamiczna w momencie lekkiego rozwarcia głośni” (Abramson 2000: 8) lub też jako „szum frykcyjny generowany przez rozchyloną głośnię po wybuchu spółgłoski” (Keating 1984: 295). Rycina 3 przedstawia oscylogram i spektrogram angielskiej sylaby *tar* /ta:/ z wartością parametru VOT +95 ms.

<sup>1</sup> W wartościach krótkich VOT język angielski lokuje swe dźwięczne spółgłoski zwarto-wybuchowe, o czym w dalszej części wywodu.

**Rycina 3.** Oscylogram i spektrogram angielskiej sylaby *tar* /ta:/. Wartość parametru VOT +95 ms



## 1. Parametr VOT w artykulacji

Próby stworzenia całościowego fonetycznego opisu fenomenu dźwięczności i bezdźwięczności spółgłosek zwarto-wybuchowych w językach świata (np. Keating 1984; Kohler 1984; Kingston, Diehl 1994, 1995) ukonstytuowały podział na języki oparte na dźwięczności i języki oparte na aspiracji. Języki oparte na dźwięczności wykorzystują kontrast między wartościami ujemnymi VOT a wartościami krótkimi VOT. Są to głównie języki wschodniej i południowej Europy, z rodziny języków romańskich, słowiańskich, bałtyckich oraz węgierski (Jansen 2004). Jeśli język wykorzystuje tylko jedną kategorię dźwięczności (tylko dźwięczne bądź tylko bezdźwięczne), należy on prawie zawsze do tej kategorii. 49 z 50 języków tylko z jedną kategorią, zbadanych przez Maddiesona (1984), lokuje swoje zwarto-wybuchowe w granicach wartości krótkich VOT<sup>2</sup>. Natomiast do grupy języków opartych na aspiracji należą języki takie, jak: duński, farski, islandzki, norweski, szwedzki (Jessen 1998; Jansen 2004) oraz standardowe odmiany angielskiego i niemieckiego<sup>3</sup>.

Na ogół trzy wyżej opisane kategorie VOT są wystarczające do opisu kontrastu dźwięczności w większości języków świata. Spośród 51 języków zbadanych przez Keating i współpracowników (1983) wszystkie lokowały kontrast dźwięczności w dwóch z trzech kategorii VOT. Liczba języków, które kontrastują wartości ujemne VOT z wartościami krótkimi VOT, była porównywalna do liczby języków, które kontrastują wartości krótkie VOT z wartościami długimi VOT. Warto jednak wspomnieć o przypadkach, które nie dopasowują się do tego schludnego podziału na ję-

<sup>2</sup> Język aleucki jest wyjątkiem od tej reguły. Ma tylko jedną kategorię dźwięczności i lokuje ją w granicach wartości długich VOT (Cho, Ladefoged 1999).

<sup>3</sup> Wartości parametru VOT w niektórych dialektach angielskiego są charakterystyczne dla grupy języków opartych na dźwięczności (Wells 1982; Docherty 1992; Gimson 1994; Hughes i in. 2005).

zyki oparte na dźwięczności i języki oparte na aspiracji. Na przykład tajski i wschodnioarmieński wykorzystują wszystkie trzy kategorie VOT (Lisker, Abramson 1964). Ostatnio Riney i współpracownicy (2007) zwrócili uwagę na fakt, że japoński może potrzebować dodatkowej kategorii pomiędzy wartościami krótkimi a wartościami długimi VOT. Podobny postulat został zgłoszony dla hebrajskiego (Raphael i in. 1995). Hindi nie tylko wykorzystuje trzy kategorie VOT, ale potrzebuje również dodatkowej kategorii określanej jako dźwięczna aspiracja (Keating 1984). W związku z tym niektórzy badacze proponują bądź podzielenie wartości krótkich VOT na odrębne kategorie dla języków opartych na dźwięczności i dla języków opartych na aspiracji (Raphael i in. 1995), bądź też wprowadzenie czterech kategorii w zakresie wartości dodatnich VOT (Cho, Ladefoged 1997, 1999; Ladefoged, Cho 2000). Propozycje te nie zastąpiły jednakże nigdy oryginalnej trzystopniowej skali, zaproponowanej przez L. Liskera i A. Abramsona (1964).

Wartości VOT zależą od miejsca artykulacji. Dla spółgłosek wargowych są zdecydowanie krótsze niż dla spółgłosek dźwięcznych czy welarnych. To, czy dźwięczne mają wartości krótsze niż welarne (Lisker, Abramson 1967; Zue 1976; Weismer 1979; Nearey i Rochet 1994), czy też są takie same (Crystal, House 1988; Cooper 1991, za: Whalen i in. 2007), wydaje się idiosynkratyczną cechą mówiącego. Docherty (1992) w badaniu wartości VOT w brytyjskiej odmianie angielszczyzny zaobserwował różnicę między spółgłoskami wargowymi a spółgłoskami dźwięcznymi i welarnymi, ale nie znalazł różnicy między samymi dźwięcznymi i welarnymi. Wśród najnowszych badań warto zauważyć pracę Whalena i współpracowników (2007), w której analizowano gaworzenie niemowlaków uczących się języka angielskiego i francuskiego. Pomiar parametru VOT w analizowanym materiale stworzyły hierarchię, w której wargowe mają najkrótszy czas VOT, następnie są dźwięczne, a najdłuższy czas VOT występuje w artykulacji spółgłosek welarnych. Identyčną hierarchię zaobserwowali Cole i współautorzy (2007) w analizie parametru VOT w mowie amerykańskich prezenterów radiowych. Wszystkie te różnice w długości VOT jako funkcja miejsca artykulacji wydają się wynikać z mechaniki funkcjonowania aparatu mowy (dyskusja w: Cho, Ladefoged 1999). Jej reguły można ująć następująco: im bardziej tylne zwanie spółgłoski, tym dłuższe wartości VOT (Fischer-Jørgensen 1954; Peterson, Lehiste 1960); im większy obszar zwania, tym dłuższe wartości VOT (Stevens i in. 1986); im krótszy czas potrzebny na rekonfigurację artykulacyjną (od zwania do samogłoski), tym krótsze wartości VOT (Hardcastle 1973).

Dodatkowym czynnikiem, który został wskazany w licznych badaniach i który wpływa na kształtowanie się wartości parametru VOT, jest tempo artykulacji (Miller i in. 1986; Miller, Volaitis 1989; Volaitis, Miller 1992; Kessinger, Blumstein 1997, 1998; Lim i in. 2001; Waniek-Klimczak 2005). Zmiana prędkości artykulowania wpływa szczególnie na wartości mierzone dla bezdźwięcznych /p, t, k/. Wartości VOT zmniejszają się odwrotnie proporcjonalnie do wzrastającego tempa mowy.

## 2. Kategoryzacja parametru VOT w percepcji

Podział parametru VOT na trzy kategorie: wartości ujemne, wartości krótkie oraz wartości długie, zaproponowany przez L. Liskera i A. Abramsona (1964) jako narzędzie analizy kontrastu dźwięczna–bezdźwięczna w artykulacji spółgłosek zwarto-wybuchowych, znalazł zastosowanie w badaniach percepcji. Manipulowanie długością VOT powoduje, że spółgłoski bezdźwięczne słyszane są jako dźwięczne i odwrotnie (Abramson, Lisker 1965, 1967, 1970, 1973; Lisker, Abramson 1967, 1970). Co więcej, percepcja w tym wypadku ma charakter kategoriyczny, a więc zmiany wartości w obrębie kategorii nie są słyszane, natomiast przekroczenie granicy powoduje nagłą zmianę w rozpoznaniu dźwięku. Klasyczne badania prowadzone przez L. Liskera i A. Abramsona polegały na tym, iż za pomocą syntezatora mowy generowano continuum VOT od -150 ms do +150 ms dla każdego z trzech miejsc artykulacji (np. sylaby /ba – pa/, /da – ta/, /ga – ka/). Angielscy uczestnicy badania nie słyszeli różnicy, jeśli manipulowano wartościami VOT np. z +10 ms na +20 ms, ponieważ wartości te znajdują się w obrębie kategorii wartości krótkich VOT, a więc kategorii dźwięcznej. Natomiast analogiczna zmiana o 10 ms z +40 ms na +50 ms powodowała nagłą zmianę kategoryzacji i badani słyszeli spółgłoskę bezdźwięczną zamiast dźwięcznej. Miller i Volaitis (1989) pokazali, że percepcja kategoriyczna parametru VOT zależy od tempa artykulacji. Autorzy zsyntezowali continuum VOT między /bi/ i /pi/, a następnie pytali uczestników badania, czy słyszą w nagłosie /b/ czy /p/. Granica kategoryzacji zmieniała się wraz z szybkością prezentacji sylaby. Im szybsze tempo, tym niższe wartości VOT wystarczały do rozpoznania spółgłoski jako bezdźwięcznej /p/. Zmiana tempa artykulacji nie musi dotyczyć samego bodźca, wystarczy, że manipuluje się tempem zdania, w którym jest osadzony (Summerfield 1981).

Ciekawych i zaskakujących wyników dostarczają badania percepcji z wykorzystaniem parametru VOT ze zwierzętami. Wskazują one na niemal identyczną czułość na granicę pomiędzy wartościami krótkimi a wartościami długimi VOT wśród zwierząt i ludzi<sup>4</sup>. Kuhl i Miller (1975) badali kategoryzację parametru VOT wśród szynszyli. Autorzy zsyntezowali sylaby od /da/ (0 ms VOT) do /ta/ (+80 ms VOT) i znaleźli granicę kategoryzacji na poziomie +33 ms dla szynszyli i +35 ms dla ludzi. W kolejnym badaniu Kuhl i Miller (1978) zauważyli, że wartości, w których zmienia się kategoria, zależą od miejsca artykulacji zarówno wśród ludzi, jak i szynszyli. Granica między /ba/ i /pa/ wynosiła +25 ms VOT, a między /ga/ i /ka/ +42 ms VOT zarówno dla ludzi, jak i szynszyli. Podobne wyniki otrzymano w badaniu z papużkami falistymi (Dogling i in. 1989 za: Hawkins 1999), rezusami (Waters, Wilson 1976) oraz japońskimi przepiórkami (Holt i in. 2001).

Badania nad percepcją parametru VOT wśród niemowlaków wskazują, że kategoryzacja jest niezależna od doświadczenia językowego. Eimas i współautorzy (1971)

<sup>4</sup> Wszystkie przywołane poniżej badania dotyczyły anglojęzycznych uczestników badania. Warto podkreślić, że moment kategoryzacji continuum VOT otrzymany w tych badaniach jest właściwy dla języka angielskiego, który lokuje granicę między /b, d, g/ a /p, t, k/ w wartościach dodatnich VOT. Język polski korzysta z granicy między wartościami ujemnymi a wartościami krótkimi, o czym w kolejnym rozdziale.

badali, czy czteromiesięczne niemowlęta są w stanie rozpoznać różnicę między angielskimi dźwięcznymi /b, d, g/ a bezdźwięcznymi /p, t, k/. Autorzy zsyntezowali kontinuum VOT między -20 ms a +80 ms w odstępach co 10 ms. Badane niemowlęta miały granicę kategoryzacji na poziomie +25 ms VOT, co jest wynikiem niezwykle zbliżonym do wyniku w grupie dorosłych użytkowników języka angielskiego. Lasky i współpracownicy (1975) pokazali, że niemowlaki urodzone w środowisku hiszpańskojęzycznym, a więc słuchające języka, który, w przeciwieństwie do języka angielskiego, kontrastuje wartości ujemne i wartości krótkie dla /b, d, g/ i /p, t, k/, rozpoznają granicę VOT właściwą dla języka angielskiego, mimo iż nie mają kontaktu z tym językiem. Identyczne wyniki otrzymał Streeter (1976) w badaniach z niemowlakami urodzonymi w środowisku używającego języka Kikuyu. Niektórzy badacze twierdzą, że do okresu ok. sześciu miesięcy niemowlaki są w stanie rozróżnić trzy kategorie VOT – wartości ujemne, wartości krótkie oraz wartości długie, mimo iż tylko dwie z nich są wykorzystywane w języku ich środowiska (Lasky i in. 1975; Aslin i in. 1981; Werker, Tees 1984). Inni badacze sugerują, że w trakcie narodzin ludzki system percepcji ustawiony jest domyślnie na rozróżnianie między wartościami krótkimi VOT a wartościami długimi VOT (Eimas i in. 1971; Eimas 1975; Serniclaes 2005), a granica między wartościami ujemnymi a wartościami krótkimi zostaje utworzona, jeśli język środowiska korzysta z takiej właśnie implementacji kontrastu dźwięczności (Keating i in. 1981).

### 3. Parametr VOT w języku angielskim i polskim

Język angielski należy do grupy języków aspirujących, a więc korzysta z kontrastu pomiędzy wartościami krótkimi dla /b, d, g/ i wartościami długimi dla /p, t, k/<sup>5</sup>. Pierwsze badania przeprowadzone przez L. Liskera i A. Abramsona (1964) wykazały, że angielski lokuje kategorię dźwięczności w wartościach krótkich, a bezdźwięczności w wartościach długich VOT.

**Tabela 1.** Średnie wartości VOT w angielskich nagłosowych spółgłoskach zwarto-wybuchowych (Lisker, Abramson 1964: 394)

	DŹWIĘCZNE	BEZDŹWIĘCZNE
wargowe /b, p/	+1 ms	+58 ms
dziąsłowe /d, t/	+5 ms	+70 ms
welarne /g, k/	+21 ms	+80 ms

Andrzej Kopczyński (1977) notuje wartości nieznacznie wyższe w kategorii dźwięcznej, co jeszcze lepiej uwidacznia, że angielski lokuje swój kontrast dźwięczności w kategoriach dodatnich VOT.

<sup>5</sup> Czasami Anglicy artykułują /b, d, g/ z wartościami ujemnymi VOT. Zależy to m.in. od miejsca artykulacji spółgłoski, kontekstu samogłoskowego czy też płci mówiącego. Jest to jednak zjawisko rzadkie i spotykane głównie w wymowie hiperpoprawnej (Keating 1984; Miller i in. 1986; Kessinger, Blumstein 1997; Magloire, Green 1999).

**Tabela 2.** Średnie wartości VOT w angielskich nagłosowych spółgłoskach zwarto-wybuchowych (Kopczyński 1977: 72)

	DŹWIĘCZNE	BEZDŹWIĘCZNE
wargowe /b, p/	+18 ms	+82,5 ms
dziąsłowe /d, t/	+14 ms	+84 ms
welarne /g, k/	+31 ms	+71 ms

W przeciwieństwie do angielskiego, język polski lokuje swój kontrast dźwięczności pomiędzy wartościami ujemnymi a wartościami krótkimi VOT (Keating 1980, 1984; Keating i in. 1981). Jak pokazują poniższe tabele z pomiarami, zwarto-wybuchowe dźwięczne mają wartości ujemne, natomiast bezdźwięczne mają stosunkowo krótkie wartości dodatnie.

**Tabela 3.** Średnie wartości VOT w polskich nagłosowych spółgłoskach zwarto-wybuchowych (Keating i in. 1981: 1262)

	DŹWIĘCZNE	BEZDŹWIĘCZNE
wargowe /b, p/	-88,2 ms	+21,5 ms
zębowe /d, t/	-89,9 ms	+27,9 ms
welarne /g, k/	-66,1 ms	+52,7 ms

**Tabela 4.** Średnie wartości VOT w polskich nagłosowych spółgłoskach zwarto-wybuchowych (Kopczyński 1977: 72)

	DŹWIĘCZNE	BEZDŹWIĘCZNE
wargowe /b, p/	-78 ms	+37,5 ms
zębowe /d, t/	-72 ms	+33 ms
welarne /g, k/	-61 ms	+49 ms

Podczas gdy w angielskim granica pomiędzy wartościami krótkimi dla /b, d, g/ a wartościami długimi dla /p, t, k/ może być rozmyta w szybkiej mowie potocznej – bezdźwięczne mogą być produkowane z wartościami VOT stosunkowo krótkimi (Lisker, Abramson 1967; Moslin 1978, za: Keating 1980) – język polski charakteryzuje się konsekwencją w utrzymywaniu sztywnej granicy między wartościami ujemnymi dla /b, d, g/ a wartościami krótkimi dla /p, t, k/ niezależnie od szybkości artykulacji (Keating 1980).

Polacy uczący się języka angielskiego stoją przed trudnym zadaniem nauczenia się artykulacji wartości długich VOT dla angielskich spółgłosek bezdźwięcznych. Dlatego też podręczniki do wymowy angielskiej dla Polaków (np. Jassem 1973, 1974; Bałutowa 1974; Reszkiewicz 1981; Arabski 1987; Sobkowiak 2001) zachęcają uczących się do artykulacji nagłosowych /p, t, k/ w angielskim z mocnym przydechem krtaniowym. Ćwiczenie to powoduje wydłużenie czasu odsunięcia periodyczności przez wprowadzenie aspiracji pomiędzy wybuch spółgłoski a rozpoczęcie samogłoski. Nie jest to jednak zadanie łatwe w trakcie spontanicznej rozmowy w ję-

zyku angielskim. Nawet zaawansowani językowo Polacy produkują wartości VOT dla angielskich /p, t, k/, które są znacznie krótsze niż wartości mierzone dla rodzimych użytkowników języka angielskiego (Waniek-Klimczak 2005)<sup>6</sup>.

Badania percepcji w języku angielskim pokazują, że granica pomiędzy wartościami krótkimi i długimi, ustalona w badaniach wymowy, funkcjonuje również jako granica percepcyjna. Krótkie wartości VOT powodują, że spółgłoska słyszana jest jako dźwięczna, natomiast długie wartości VOT powodują, że ta sama spółgłoska percypowana jest jako bezdźwięczna.

**Tabela 5.** Percepcyjna granica VOT między angielskimi nagłosowymi dźwięcznymi i bezdźwięcznymi spółgłoskami zwarto-wybuchowymi (Lisker, Abramson 1970: 565)

wargowe	+22 ms
działkowe	+37 ms
welarne	+40 ms

**Tabela 6.** Percepcyjna granica VOT między angielskimi nagłosowymi dźwięcznymi i bezdźwięcznymi spółgłoskami zwarto-wybuchowymi (Kuhl, Miller 1978: 910)

wargowe	+27 ms
działkowe	+35 ms
welarne	+42 ms

Nieliczne badania nad percepcją parametru VOT w języku polskim sugerują, że granica między wartościami ujemnymi a wartościami krótkimi VOT zaobserwowana w wymowie pełni również funkcję percepcyjną (Keating 1980). Przykładowo Mikoś i współpracownicy (1978) zsyntezowali kontinuum VOT od wartości ujemnych dla /b, d, g/ do wartości krótkich dla /p, t, k/. Polacy uczestniczący w badaniu słyszeli konsekwentnie dźwięczność w zakresie wartości ujemnych i bezdźwięczność w zakresie wartości dodatnich. Podobne wyniki dostarcza badanie Keating i współautorów (1981), w którym granica percepcyjna została ustalona na poziomie 0 ms VOT; wszystkie wartości poniżej rozpoznawane były jako dźwięczne, natomiast wszystkie wartości powyżej jako bezdźwięczne.

Powyższe porównanie percepcji parametru VOT w języku angielskim i polskim pozwala wyciągnąć następujące wnioski. Polski lokuje dźwięczne zwarto-wybuchowe w granicach wartości ujemnych, natomiast bezdźwięczne zwarto-wybuchowe w granicach wartości krótkich. Angielski natomiast wykorzystuje wartości krótkie dla dźwięcznych zwarto-wybuchowych, a wartości długie dla bezdźwięcznych. Powoduje to komplikacje percepcyjne wynikające z nachodzenia na siebie kategorii

<sup>6</sup> Warto zauważyć, że polskie bezdźwięczne, które lokują się w wartościach krótkich VOT, są zarówno artykulacyjnie prostsze niż angielskie bezdźwięczne, których parametr VOT znajduje się w wartościach długich (Westbury, Keating 1980, za: Keating 1984), jak i łatwiejsze do przyswojenia (Kewley-Port, Preston 1974; Scobbie i in. 1996). Jednakże sama kategoria wartości ujemnych VOT, w której język polski lokuje swoje /b, d, g/, jest przyswajana stosunkowo późno (Macken, Barton 1980; Eilers i in. 1984; Allen 1985; Gandour i in. 1986).



VOT. Te same krótkie wartości VOT będą percypowane przez Polaków jako /p, t, k/, a przez Anglików jako /b, d, g/. Ponieważ język angielski wykorzystuje kontinuum wartości dodatnich do rozróżnienia między spółgłoskami dźwięcznymi i bezdźwięcznymi, Polacy uczący się angielskiego będą musieli utworzyć nową granicę percepcyjną, jako że język polski ma tylko jedną kategorię, bezdźwięczną, w obszarze wartości dodatnich VOT. Poniższe badanie ma na celu ustalić, jak rodowici użytkownicy języka angielskiego kategoryzują kontinuum wartości dodatnich VOT oraz czy Polacy są w stanie rozpoznać kontrast dźwięczna–bezdźwięczna w zakresie wartości niewykorzystywanych przez ich język rodzimy.

#### **4. Badanie**

Celem niniejszego badania jest ustalenie, jak kontinuum dodatnich wartości parametru VOT jest percypowane przez rodowitych użytkowników angielskiego, których język lokuje tutaj swoją granicę między dźwięcznymi a bezdźwięcznymi zwarto-wybuchowymi, a jak przez Polaków, których język nie wykorzystuje granicy między wartościami krótkimi a wartościami długimi w implementacji kontrastu dźwięczności i bezdźwięczności.

##### **4.1. Uczestnicy badania**

###### **4.1.1. Rodowici użytkownicy języka angielskiego**

W badaniu wzięło udział 11 rodowitych użytkowników języka angielskiego: 6 Amerykanów oraz 5 Brytyjczyków w wieku od 23 do 56 lat (średnia: 32, odchylenie stand.: 9,05). Wszyscy uczestnicy wzięli udział w badaniu dobrowolnie i nie zostali wynagrodzeni finansowo. Żaden z badanych nie był świadomy celu badania. Wstępny wywiad ujawnił, że żaden z badanych nie znał biegle języka obcego. Nikt nie miał historii chorób słuchu lub mowy oraz nie cierpiał na żadne choroby słuchu w momencie badania.

###### **4.1.2. Rodowici użytkownicy języka polskiego**

W badaniu wzięło udział 26 Polaków, którzy uczestniczyli w kursie języka angielskiego na poziomie początkującym po 6 miesiącach nauki. Uczestnicy byli uczniami kursu języka angielskiego sponsorowanego przez Europejski Fundusz Społeczny i skierowanego do osób, które nigdy wcześniej nie uczyły się języka angielskiego. Osoby w tej grupie były w wieku od 26 do 47 lat (średnia: 39, odchylenie standardowe: 6,51). Wszyscy uczestnicy wzięli udział w badaniu dobrowolnie i nie zostali wynagrodzeni finansowo. Żaden z badanych nie był świadomy celu badania. Wstępny wywiad ujawnił, że żaden z badanych nie znał biegle języka obcego. Nikt nie miał historii chorób słuchu lub mowy oraz nie cierpiał na żadne choroby słuchu w momencie badania.

## 4.2. Nagranie materiału i kryteria manipulacji

Materiał do badania percepcji został nagrany w stosownie wyciszonym pomieszczeniu. Bodziec wyjściowy został nagrany przez rodzowego użytkownika języka angielskiego. Nagrywający został poinstruowany, by artykułować z intonacją neutralną.

Nagranie zostało dokonane za pomocą mikrofonu Media Tech MT385 USB o czułości pomiędzy 100Hz a 16000Hz, ulokowanego 20 centymetrów od nagrywającego. Sygnał został przetworzony za pomocą zewnętrznej karty dźwiękowej Sound Blaster X-Fi X-MOD z 24-bitową rozdzielczością, zakresem między 140Hz a 20000Hz i czułością 112 dB +/-3 dB. Częstotliwość próbkowania nagrania wynosiła 22,05 kHz (16-bitowa rozdzielczość). Nagranie zostało przetworzone do pliku wav i przechowane do dalszej manipulacji na twardym dysku.

Analiza i manipulacja zostały wykonane za pomocą programu do analizy mowy Praat 4.6.18 (Boersma 2001; Boersma, Weenink 2007). Długość parametru VOT mierzona była poprzez analizę oscylogramu i spektrogramu od momentu wybuchu spółgłoski, widocznego jako nieperiodyczne wzburzenie sygnału, do momentu rozpoczęcia samogłoski, widocznego w postaci pojawienia się drugiego formantu (np. Abramson 1977; Lisker 1978; Keating 1980; Keating i in. 1981; Cho i in. 2002; Cole i in. 2007).

## 4.3. Materiał

Z nagranych logatomów *keef* /ki:f/ (+70 ms VOT w nagłosowym /k/)<sup>7</sup> utworzonych zostało 8 bodźców od +70 ms do 0 ms VOT w odstępach 10 ms. Nagłosowa sekwencja /ki:/ została wybrana, ponieważ spółgłoska welarna poprzedzająca samogłoskę wysoką daje najdłuższy przedział długości VOT (Cho, Ladefoged 1999; Chang i in. 2001). Otrzymane 8 przedziałów dodatnich wartości VOT obejmuje kategorię dźwięczną i bezdźwięczną w języku angielskim oraz tylko kategorię bezdźwięczną w języku polskim.

1. *keef* /ki:f/, /k/ +70 ms VOT
2. *keef* /ki:f/, /k/ +60 ms VOT
3. *keef* /ki:f/, /k/ +50 ms VOT

<sup>7</sup> Użycie logatomów w badaniach percepcji fonetycznej motywowane jest faktem, że jakiegokolwiek czynniki semantyczne mogą zakłócić przetwarzanie sygnału na poziomie fonetycznym. Literatura psycholingwistyczna obfituje w wyniki badań, które wskazują, że słowa o większej częstotliwości użycia są rozpoznawane szybciej (Soloman, Postman 1952; Oldfield, Wingfield 1965; Luce, Pisoni 1998), słowa konkretne łatwiej niż abstrakcyjne (Strain i in. 1995), a słowa przewidywalne z kontekstu łatwiej niż nieprzewidywalne (Morton, Long 1976). Konkretnym przykładem dla parametru VOT jest efekt Ganonga (1980), który powoduje, że granica percepcji kategorii dźwięczności zmienia się zależnie od tego, czy badani znają słowo lepiej czy gorzej. Na przykład badanie kontinuum VOT w polskich słowach *kot* – *Got* dałoby zniekształcone wyniki, ponieważ słowo *kot* byłoby aktywowane znacznie częściej z racji swej wysokiej frekwencyjności. Tak więc najlepszym sposobem na uniknięcie tak dużej liczby zmiennych jest wyłączenie poziomu przetwarzania leksykalnego.

4. *keef* /ki:f/, /k/ +40 ms VOT
5. *keef* /ki:f/, /k/ +30 ms VOT
6. *keef* /ki:f/, /k/ +20 ms VOT
7. *keef* /ki:f/, /k/ +10 ms VOT
8. *keef* /ki:f/, /k/ 0 ms VOT

#### 4.4. Procedura

Eksperyment odbywał się w wyciszonym pokoju, gdzie badani słuchali materiału przez wysokiej jakości głośniki audio. Siła dźwięku została ustawiona na poziomie zapewniającym optymalną słyszalność dla wszystkich uczestników. Pojedyncza sesja zawierała średnio 6–7 osób i trwała około 25 minut. Kolejność prezentacji materiału została automatycznie wymieszana. Każdy bodziec prezentowany był 4 razy, co dało w sumie 32 próby. Każda poszczególna prezentacja bodźca składała się z 2 powtórzeń rozdzielonych dwusekundową pauzą.

Przed eksperymentem badani zostali poinformowani, że usłyszą nieistniejące, ale fonologicznie potencjalne słowa angielskie nagrane przez rodzowitego użytkownika języka angielskiego.

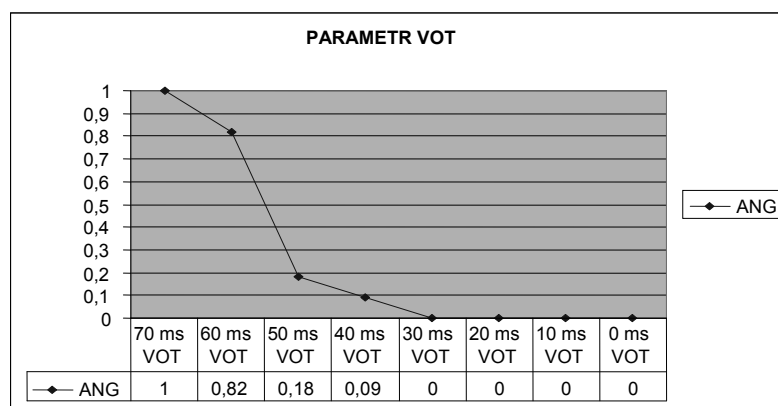
Badani korzystali z listy zawierającej 34 przykłady sekwencji *k / g eef*, na której zakreślali albo bezdźwięczne /k/, albo dźwięczne /g/.

Przed rozpoczęciem badania uczestnicy byli zachęceni do zadawania pytań, które rozwiązywałyby wszelkie wątpliwości i zagwarantowały poprawne przeprowadzenie procedury. Przed prezentacją właściwego materiału badawczego uczestnicy wykonali 5 prób na słowach nieistotnych dla celów badania.

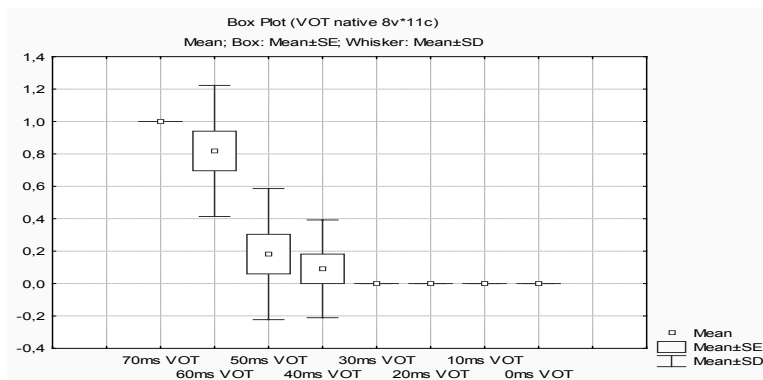
#### 4.5. Wyniki i analiza

##### 4.5.1. Rodowici użytkownicy języka angielskiego

**Rycina 4.** Percepcja nagłosowej spółgłoski jako bezdźwięcznej /k/ wraz z malejącymi wartościami VOT przez rodowitych użytkowników języka angielskiego



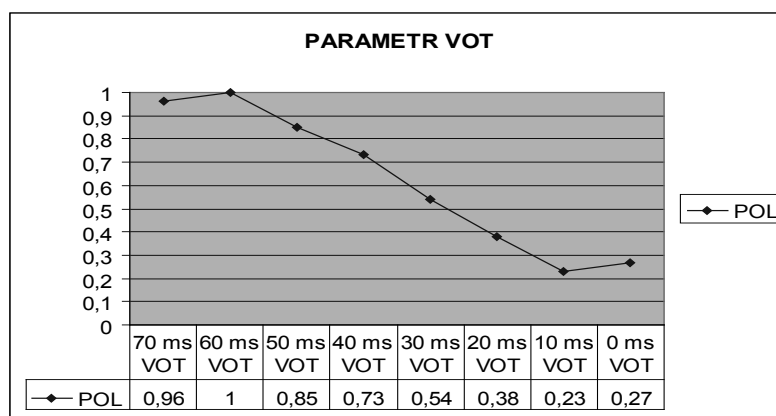
**Rycina 5.** Wykres analizy statystycznej dla grupy rodowitych użytkowników języka angielskiego: średnia, odchylenie standardowe, błąd standardowy



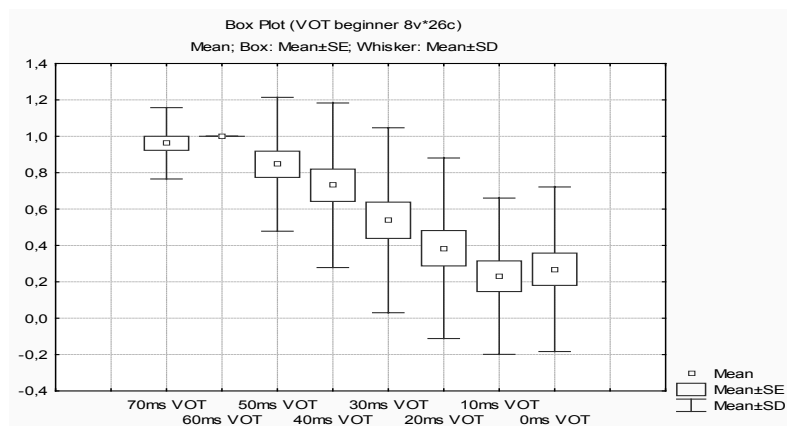
Analiza statystyczna w grupie rodowitych użytkowników języka angielskiego pokazała, że zmniejszająca się długość parametru VOT była istotnym czynnikiem w percepcji dźwięczności nagłosowej spółgłoski zwarto-wybuchowej. Nieparametryczny test dla zależnych prób nominalnych Cochran Q wykazał statystycznie wysoce istotny wpływ czasu VOT na zmianę kategorii dźwięczności ( $Q = 84,620$ ,  $p = .000^{**}$ ). Moment redukcji o 20 ms do poziomu +50 ms VOT powodował kategoryczną zmianę percepcji z /k/ na /g/. Wszystkie wartości poniżej +30 ms percypowane były jako dźwięczne /g/.

#### 4.5.2. Rodowici użytkownicy języka polskiego

**Rycina 6.** Percepcja nagłosowej spółgłoski jako bezdźwięcznej /k/ wraz z malejącymi wartościami VOT przez rodowitych użytkowników języka polskiego



**Rycina 7.** Wykres analizy statystycznej dla grupy rodowitych użytkowników języka polskiego: średnia, odchylenie standardowe, błąd standardowy

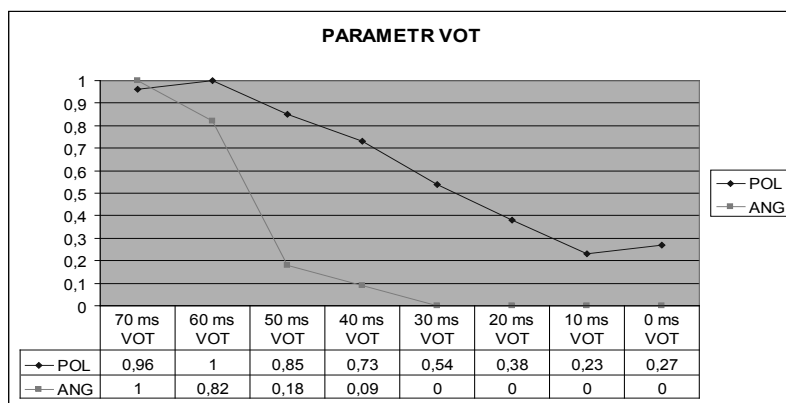


Nieparametryczny test dla nominalnych prób zależnych Cochran Q wykazał, że zmniejszające się wartości VOT były również czynnikiem statystycznie wysoce istotnym w percepcji dźwięczności przez Polaków ( $Q = 85,997$ ,  $p = 0.000^{**}$ ). Grupa polska reagowała na zmniejszające się wartości VOT, rozpoznając bodźce proporcjonalnie częściej jako dźwięczne /g/, w przeciwieństwie jednak do grupy rodowitych użytkowników języka angielskiego nie można zaobserwować zmiany katagorycznej w percepcji kontrastu dźwięczności. W grupie polskiej percepcja dźwięczności nie jest kompletna nawet dla wartości 0 ms VOT. Podczas gdy rodowici użytkownicy języka angielskiego percypują wartości poniżej 30 ms VOT jako dźwięczne, Polacy w ok. 20% przypadków słyszą spółgłoskę bezdźwięczną na poziomie +10 ms VOT. Warto również zwrócić uwagę na trudny do wytłumaczenia wzrost percepcji bezdźwięczności dla wartości 0 ms VOT w porównaniu do wartości +10 ms VOT.

## 5. Dyskusja

Zestawienie wyników percepcji dla obu grup wskazuje na różnicę w percepcji kontrastu dźwięczności w zakresie wartości dodatnich parametru VOT.

**Rycina 8.** Zestawienie percepcji dla grupy rodowitych użytkowników języka angielskiego oraz użytkowników języka polskiego



Zgodnie z wynikami pomiarów parametru VOT w artykulacji badanie percepcji pokazuje, że Polacy mają trudności z kategoryzowaniem kontrastu dźwięczna–bezdźwięczna w zakresie wartości dodatnich VOT. Jako że polski kontrastuje wartości ujemne VOT dla spółgłosek dźwięcznych oraz wartości dodatnie krótkie dla spółgłosek bezdźwięcznych, rozpoznanie angielskiego kontrastu między wartościami dodatnimi krótkimi dla /b, d, g/ a wartościami długimi dla /p, t, k/ jest dla Polaków zadaniem trudnym. Mimo iż można zaobserwować stopniowy wzrost percepcji dźwięcznego /g/ wraz z malejącą długością VOT, nie jest to zmiana kategoryczna, jaką można zaobserwować u anglojęzycznych uczestników badania. W grupie anglojęzycznej następuje przeskok z kategorii bezdźwięcznej do kategorii dźwięcznej na poziomie +50 ms VOT, a kompletne rozpoznanie spółgłoski jako /g/ następuje dla wartości poniżej +30 ms. Przeskok ten odzwierciedla czułość systemu percepcyjnego rodowitych użytkowników języka angielskiego na granicę kategorii dźwięczności między wartościami krótkimi a wartościami długimi VOT. Brak pełnego i kategorycznego przeskoku z kategorii dźwięcznej do bezdźwięcznej wśród Polaków sugeruje, że użytkownicy języka polskiego oczekują wartości ujemnych VOT, by sklasyfikować z pełnym przekonaniem spółgłoskę jako dźwięczną.

Różnica między językiem angielskim i polskim w implementacji kontrastu dźwięczna–bezdźwięczna wzdłuż parametru VOT ma oczywisty wpływ na problematykę uczenia Polaków poprawnej wymowy i skutecznej percepcji języka angielskiego. Dydaktyka języka angielskiego musi zwrócić uwagę zarówno na artykulację odpowiednio długich wartości VOT dla angielskich /p, t, k/, jak i na ćwiczenia percepcji kontrastu między dźwięcznymi a bezdźwięcznymi spółgłoskami zwarto-wybuchowymi w pozycjach nagłosowych. O ile percepcja angielskich /p, t, k/ nie stwarza problemu, ponieważ wartości długie VOT, niewykorzystywane w polskim, są łatwo rozpoznawane przez Polaków jako bezdźwięczne, o tyle angielskie /b, d, g/ mają wartości VOT bardzo zbliżone do polskich /p, t, k/ i, jak wskazuje niniejsze badanie, mogą być rozpoznawane jako bezdźwięczne. Odpowiedni trening fonetycz-

ny pozwala jednak Polakom na stworzenie kategorii dźwięczności wzdłuż wartości dodatnich VOT (Rojczyk w druku). Uczący się języka angielskiego na poziomie zaawansowanym są w stanie kategoryzować wartości dodatnie VOT w sposób zbliżony do rodowitych użytkowników języka angielskiego.

Różnice w implementacji opozycji dźwięczna–bezdźwięczna wzdłuż wartości VOT między polskim a angielskim pozwalają na wykorzystanie tego mechanizmu w badaniach psycholingwistycznych. Pozwala on na przykład na obserwowanie procesu przyswajania struktury dźwiękowej języka obcego w różnych etapach zaawansowania. Nabywanie języka wymaga tworzenia nowych kategorii kontrastu, które są obce w języku ojczystym. Pomiar parametru VOT daje możliwość uchwycenia kolejnych stadiów rozwoju nowej organizacji kontrastu między dźwięcznością i bezdźwięcznością. Przyszłe badania powinny zwrócić również uwagę, jak anglojęzyczni uczniowie języka polskiego kategoryzują polski kontrast między wartościami ujemnymi a wartościami krótkimi VOT; kontrast, który jest niewykorzystany w języku angielskim.

## Literatura

- ABRAMSON A. S., 1977, *Laryngeal timing in consonant distinctions*, „Phonetica” 34, s. 295–303.
- ABRAMSON A. S., 2000, *The perception of voicing distinctions*, [w:] Burnham, D., Luksaneeyanawin, S., Davis, C., Lafourcade, M. (ed.) *Interdisciplinary Approaches to Language Processing*, Bangkok, s. 25–31.
- ABRAMSON A. S., LISKER L., 1965, *Voice onset time in stop consonants: Acoustic analysis and synthesis*, [w:] *Proceedings of the 5th International Congress of Acoustics*, Liège, A51.
- ABRAMSON A. S., LISKER L., 1967, *Discrimination along the voicing continuum: Cross-language tests*, [w:] *Status Report on Speech Research*, Haskins Laboratories, s. 17–22.
- ABRAMSON A. S., LISKER L., 1970, *Discriminability along the voicing continuum: Cross language tests*, [w:] *Proceedings of the Sixth International Congress of Phonetic Sciences, Prague 1967*, Prague, s. 569–573.
- ABRAMSON A. S., LISKER L., 1973, *Voice-timing perception in Spanish word-initial stops*, „Journal of Phonetics” 1, s. 1–8.
- ALLEN G. D., 1985, *How the young French child avoids the pre-voicing problem for word-initial voiced stops*, „Journal of Child Language” 12, s. 37–46.
- ARABSKI J., 1987, *Wymowa amerykańska*, Warszawa.
- ASLIN R. N., PISONI D. B., HENNESSY B. L., PERREY A. V., 1981, *Discrimination of voice onset time by human infants: New findings and implications for the effect of early experience*, „Child Development” 52, s. 1135–1145.
- BALUTOWA S., 1974, *Wymowa angielska dla wszystkich*, Warszawa.
- BOERSMA P., 2001, *Praat, a system for doing phonetics by computer*, „Glott International” 10, s. 341–345.
- BOERSMA P., WEENINK D., 2007, *Praat: Doing phonetics by computer*. (Wersja 4.6.18) [Program komputerowy]. Ściągnięty ze strony <http://www.praat.org> 31 sierpnia 2007.
- CHANG S. S., PLAUCHÉ M. C., OHALA J. J., 2001, *Markedness and consonant confusion asymmetries*, [w:] E. Hume, K. Johnson (ed.), *The Role of Speech Perception in Phonology*, New York.
- CHO T., JUN S. A., LADEFOGED P., 2002, *Acoustic and aerodynamic correlates to Korean stops and fricatives*, „Journal of Phonetics” 30, s. 193–229.
- CHO T., LADEFOGED P., 1997, *Variations and universals in VOT: Evidence from 17 endangered languages*, „UCLA Working Papers in Phonetics” 94, s. 18–40.

- CHO T., LADEFOGED P., 1999, *Variations and universals in VOT: Evidence from 18 languages*, „Journal of Phonetics” 27, s. 207–222.
- COLE J., KIM H., CHOI H., HASEGAWA-JOHNSON M., 2007, *Prosodic effects on acoustic cues to stop voicing and place of articulation. Evidence from Radio News speech*, „Journal of Phonetics” 35, s. 180–209.
- COOPER A., 1991, *An Articulatory Account of Aspiration in English*, rozprawa doktorska, Uniwersytet Yale.
- CRYSTAL T., HOUSE A., 1988, *Segmental durations in connected-speech signals: Current results*, „Journal of the Acoustical Society of America” 83, s. 1553–1573.
- DOCHERTY G., 1992, *The Timing of Voicing in British English Obstruents*, Dordrecht.
- DOGLING R. J., OKANOYA K., BROWN S. D., 1989, *Speech perception by budgerigars (Melopsittacus undulatus): The voiced-voiceless distinction*, „Perception and Psychophysics” 46, s. 65–71.
- EILERS R. E., OLLER D. K., BENITO-GARCIA C. R., 1984, *The acquisition of voicing contrasts in Spanish and English learning infants and children: A longitudinal study*, „Journal of Child Language” 11, s. 313–336.
- EIMAS P. D., 1975, *Speech perception in early infancy*, [w:] L. B. Cohen, P. Salapatek (ed.), *Infant Perception: From Sensation to Cognition*, vol. II, New York, s. 193–231.
- EIMAS P. D., SIQUELAND E. R., JUSZCZYK P., VIGORITO J., 1971, *Speech perception in infants*, „Science” 171, s. 303–306.
- FISCHER-JØRGENSEN E., 1954, *Acoustic analysis of stop consonants*, „Miscellanea Phonetica” 2, s. 42–59.
- GANDOUR J., PETT S. H., DARDARANANDA R., DECHONGKIT S., MUKNGOEN S., 1986, *The acquisition of the voicing contrast in Thai: A study of voice onset time in word-initial stop consonants*, „Journal of Child Language” 13, s. 561–572.
- GANONG W. F., 1980, *Phonetic categorization in auditory word perception*, „Journal of Experimental Psychology: Human Perception and Performance” 6, s. 110–125.
- GIMSON A., 1994, *Gimson's Pronunciation of English*, 5th ed. rev. by A. Cruttenden, London.
- HARDCASTLE W. J., 1973, *Some observations on the Tense-Lax distinction in initial stops in Korean*, „Journal of Phonetics” 1, s. 263–271.
- HAWKINS S., 1999, *Looking for invariant correlates of linguistic units: Two classical theories of speech perception*, [w:] J. M. Pickett (ed.), *The Acoustics of Speech Communication: Fundamentals, Speech Perception Theory, and Technology*, Boston, s. 198–231.
- HOLT L. L., LOTTO A. J., KLUENDER K. R., 2001, *Influence of fundamental frequency on stop-consonant voicing perception: A case of learned covariation or auditory enhancement?* „Journal of the Acoustical Society of America” 109, s. 764–774.
- HUGHES A., TRUDGILL P., WATT D., 2005, *English Accents and Dialects: An Introduction to Social and Regional Varieties of English in the British Isles, 4th edition*, London.
- JANSEN W., 2004, *Laryngeal Contrast and Phonetic Voicing: A laboratory Phonology Approach to English, Hungarian, and Dutch*, rozprawa doktorska, Uniwersytet Groningen.
- JASSEM W., 1973, *Podręcznik wymowy angielskiej*, Warszawa.
- JASSEM W., 1974, *Fonetyka języka angielskiego*, Warszawa.
- JESSEN M., 1998, *Phonetics and Phonology of Tense and Lax obstruents in German*, Amsterdam.
- KEATING P. A., 1980, *A Phonetic Study of a Voicing Contrast in Polish*, rozprawa doktorska, Uniwersytet Brown.
- KEATING P. A., 1984, *Phonetic and phonological representation of stop consonant voicing*, „Language” 60, s. 286–319.
- KEATING P. A., LINKER W., HUFFMAN M., 1983, *Patterns in allophone distribution for voiced and voiceless stops*, „Journal of Phonetics” 11, s. 277–290.
- KEATING P. A., MIKOŚ M. J., GANONG III W. F., 1981, *A cross-language study of range of voice onset time in the perception of initial stop voicing*, „Journal of the Acoustical Society of America” 70, s. 1261–1271.
- KESSINGER R. H., BLUMSTEIN S. E., 1997, *Effects of speaking rate on voice-onset time in Thai, French, and English*, „Journal of Phonetics” 25, s. 143–168.



- KESSINGER R. H., BLUMSTEIN S. E., 1998, *Effects of speaking rate on voice-onset and vowel production: Some implications for perception studies*, „Journal of Phonetics” 26, s. 117–128.
- KEWLEY-PORT D. K., PRESTON M. S., 1974, *Early apical stop production: A voice onset time analysis*, „Journal of Phonetics” 2, s. 195–210.
- KINGSTON J., DIEHL R. L., 1994, *Phonetic knowledge*, „Language” 70, s. 419–454.
- KINGSTON J., DIEHL R. L., 1995, *Intermediate properties in the perception of distinctive feature values*, [w:] B. Connel, A. Arvaniti (ed.), *Papers in Laboratory Phonology IV: Phonology and Phonetic Evidence*, Cambridge, s. 7–27.
- KLATT D. H., 1975, *Voice-onset time, friction, and aspiration in word-initial consonant clusters*, „Journal of Speech and Hearing Research” 18, s. 687–703.
- KOHLER K. J., 1984, *Phonetic explanation in phonology. The feature fortis lenis*, „Phonetica” 41, s. 150–174.
- KOPCZYŃSKI A., 1977, *Polish and American English Consonant Phonemes: A Contrastive Study*, Warszawa.
- KUHL P. K., MILLER J. D., 1975, *Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants*, „Science” 190, s. 69–72.
- KUHL P. K., MILLER J. D., 1978, *Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli*, „Journal of the Acoustical Society of America” 63, s. 905–917.
- LADEFOGED P., CHO T., 2000, *Linking linguistic contrasts to reality: The case of VOT*, „UCLA Working Papers on Phonetics” 98, s. 1–9.
- LASKY R. E., SYRDAL-LASKY A., KLEIN R. E., 1975, *VOT discrimination by four to six and a half month old infants from Spanish environments*, „Journal of Experimental Child Psychology” 20, s. 215–225.
- LIM B. J., JONG K. D., NAGAO K., 2001, *Cross-language perception of syllable affiliation: Effects of voicing and language background*, „Berkeley Linguistic Society” 27, s. 43–67.
- LISKER L., 1978, *In qualified defense of VOT*, „Language and Speech” 21, s. 375–383.
- LISKER L., 1986, „Voicing” in English: *A catalogue of acoustic features signalling /b/ versus /p/ in trochees*, „Language and Speech” 29, s. 3–11.
- LISKER L., ABRAMSON A. S., 1964, *A cross language study of voicing in initial stops: Acoustic measurements*, „Word” 20, s. 384–422.
- LISKER L., ABRAMSON A. S., 1967, *Some effects of context on voice onset time in English stops*, „Language and Speech” 10, s. 1–28.
- LISKER L., ABRAMSON A. S., 1970, *The voicing dimension: Some experiments in comparative phonetics*, [w:] *Proceedings of the Sixth International Congress of Phonetic Sciences, Prague 1967*, Prague, s. 63–567.
- LUCE P. A., PISONI D. B., 1998, *Recognizing spoken words: The Neighborhood Activation Model*, „Ear and Hearing” 20, s. 19–36.
- MACKEN M. A., BARTON D., 1980, *The acquisition of the voicing contrast in Spanish: A phonetic and phonological study of word-initial stop consonants*, „Journal of Child Language” 7, s. 433–458.
- MADDIESON I., 1984, *Patterns of Sounds*, Cambridge.
- MAGLOIRE J., GREEN K., 1999, *A cross-language comparison of speaking rate effects on the production of Voice Onset Time in English and Spanish*, „Phonetica” 56, s. 158–185.
- MIKOŚ M. J., KEATING P. A., MOSLIN B. J., 1978, *The perception of voice onset time in Polish*, „Journal of the Acoustical Society of America” (S1), 63, S19.
- MILLER J. L., GREEN K. P., REEVES A., 1986, *Speaking rate and segments: A look at the relation between speech production and speech perception for the voicing contrast*, „Phonetica” 43, s. 104–115.
- MILLER J. L., VOLAITIS L. E., 1989, *Effects of speaking rate on the perceptual structure of a phonetic category*, „Perception and Psychophysics” 46, s. 505–512.
- MORTON J., LONG J., 1976, *Effects of word transition probability on phoneme identification*, „Journal of Verbal Learning and Verbal Behavior” 15, s. 43–52.
- MOSLIN B. J., 1978, *The Role of Phonetic Input in the Child’s Acquisition of the Voiced-Voiceless Contrast in English Stops: A Voice Onset Time Analysis*, rozprawa doktorska, Uniwersytet Brown.

- NEAREY T. M., ROCHET B. I., 1994, *Effects of place of articulation and vowel context on VOT production and perception in French and English stops*, „Journal of the International Phonetic Association” 24, s. 1–19.
- OLDFIELD R. C., WINGFIELD A., 1965, *Response latencies in naming objects*, „Quarterly Journal of Experimental Psychology” 17, s. 273–281.
- PETERSON G. E., LEHISTE I., 1960, *Duration of syllable nuclei in English*, „Journal of the Acoustical Society of America” 32, s. 693–703.
- RAPHAEL L. J., TOBIN Y., FABER A., MOST T., KOLLIA H. B., MILSTEIN D., 1995, *Intermediate values of voice onset time*, [w:] F. Bell-Berti, L. J. Raphael (ed.), *Producing Speech: Contemporary Issues. For Katherine Safford Harris*, New York.
- RESZKIEWICZ A., 1981, *Correct your English Pronunciation*, Warszawa.
- RINEY T. J., TAKAGI N., OTA K., UCHIDA Y., 2007, *The intermediate degree of VOT in Japanese initial voiceless stops*, „Journal of Phonetics” 35, s. 439–443.
- ROJCZYK A. (w druku), *Temporal and spectral parameters in perception of the voicing contrast in English and Polish*, Katowice.
- SCOBIE J. M., GIBBON F., HARDCASTLE W. J., FLETCHER P., 1996, *Covert contrast as a stage in the acquisition of phonetics and phonology*, „QMC Working Papers in Speech and Language Sciences”, vol. 1, s. 43–62.
- SERNICLAES W., 2005, *On the invariance of speech percepts*, „ZAS Papers in Linguistics” 40, s. 177–194.
- SOBKOWIAK W., 2001, *English Phonetics for Poles: A Resource Book for Learners and Teachers*, Poznań.
- SOLOMAN R. L., POSTMAN L., 1952, *Frequency of usage as a determinant of recognition thresholds for words*, „Journal of Experimental Psychology” 43, s. 195–201.
- STEVENS K. N., KEYSER S. J., KAWASAKI H., 1986, *Toward a phonetic and phonological theory of redundant features*, [w:] J. S. Perkell, D. H. Klatt (ed.) *Invariance and Variability in Speech Processes*, Hillsdale, s. 426–429.
- STRAIN E., PATTERSON K., SEIDENBERG M. S., 1995, *Semantic effects in single-word naming*, „Journal of Experimental Psychology: Learning, Memory, and Cognition” 21, s. 1140–1154.
- STREETER L. A., 1976, *Language perception of two-month old infants shows effects of both innate mechanisms and experience*, „Nature” 259, s. 39–41.
- SUMMERFIELD A. Q., 1981, *Articulatory rate and perceptual constancy in phonetic perception*, „Journal of Experimental Psychology: Human Perception and Performance” 7, s. 1074–1095.
- VOLATIS L. E., MILLER J. L., 1992, *Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories*, „Journal of the Acoustical Society of America” 92, s. 723–735.
- WANIEK-KLIMCZAK E., 2005, *Temporal Parameters in Second Language Speech: An Applied Linguistic Phonetics Approach*, Łódź.
- WATERS R., WILSON W., 1976, *Speech perception by rhesus monkeys: The voicing distinction in synthesised labial and velar stops consonants*, „Perception and Psychophysics” 19, s. 285–289.
- WEISMER G., 1979, *Sensitivity of voice onset measures to certain segmental features in speech production*, „Journal of Phonetics” 7, s. 194–204.
- WELLS J., 1982, *Accents of English*, vol. 2: *The British Isles*, Cambridge.
- WESTBURY J. R., KEATING P. A., 1980, *A model of stop consonant voicing and a theory of markedness* [Artykuł zaprezentowany na zjeździe Amerykańskiego Towarzystwa Językoznawczego].
- WHALEN D. H., LEVITT A. G., GOLDSTEIN L. M., 2007, *VOT in the babbling of French- and English-learning infants*, „Journal of Phonetics” 35, s. 341–352.
- WERKER J. F., TEES R. C., 1984, *Cross-language speech perception: Evidence for perceptual reorganization during the first year of life*, „Infant Behavior and Development” 7, s. 49–63.
- ZUE V. W., 1976, *Acoustic Characteristics of Stop Consonants: A Controlled Study (Technical Report 523)*, Lexington.

### **The Voice Onset Time in Polish and English: perception experiments Summary**

The Voice Onset Time (VOT) introduced by Lisker and Abramson (1964) is defined as the single production dimension, the time interval between the release of a stop occlusion and the onset of vocal cord vibration. Languages generally fall into two of the three broad categories that show little cross-linguistic variation: voicing lead, short lag, and long lag. English and Polish exploit the VOT continuum differently. While English contrasts short lag vs. long lag for voiced and voiceless stops, Polish exploits voicing lead vs. short lag for its voiced and voiceless stops. This acoustic difference makes an interesting cross-linguistic scenario for perception studies in an identification paradigm. From a naturally obtained nonword *keef*, we generated 8 stimuli with the VOT values of an initial stop ranging in 10ms-steps from 0ms to +70 ms. These values span across the English VOT boundary which separates short lag (voiced) vs. long lag (voiceless) categories. In a forced-choice format, we asked two groups of subjects – native speakers of English and Polish beginner learners of English – to recognise and initial segment in each stimulus. The analysis of the results shows that the two groups performed differently in that native speakers categorised short lag into voiced /g/ and long lag into voiceless /k/. Polish subjects, on the other hand, did not exhibit a categorical shift from a voiceless into voiced category.