Assessing the effects of "native speaker" status on classic findings in speech perception

Supplemental information

This document contains information about 1) how often papers limit their samples to "native speakers," 2) the flexible coding of participant responses, 3) the demographic questionnaire used, and 4) demographic detail about participants.

How Often do Papers Limit Samples to "Native Speakers"?

To assess how commonly papers on speech processing limit their samples to "native speakers," we conducted a search on the Web of Science (webofscience.com) in the subject areas "Psychology" and "Audiology Speech Language Pathology." We restricted our search to "articles" (i.e., not conference presentations, book chapters, etc.). The search was conducted on April 23, 2024.

Our goal was to identify influential papers in the literature that sought to establish general phenomena about human speech perception in typical adult samples using behavioral methods (e.g., testing models of spoken word recognition, understanding how bottom-up and top-down factors interact during speech processing, etc.). We chose search criteria to fulfill that goal, excluding papers about infants/children or special populations (e.g., people with dyslexia, autism, synesthesia, cochlear implants, etc.), papers that did not report information about participants (e.g., meta-analyses, computational modeling on existing datasets), studies that used neuroimaging methods, and studies that were about speech production rather than speech perception (see OSF page for the full list of articles). Finally, we excluded work that explicitly tested questions related to bilingualism or second language acquisition, because research in these subdisciplines necessitates careful attention to differences in language background across participants.

We used the following search terms:

- "speech perception" or "spoken word recognition" (in "all fields")
- Not: "dyslexia", "cochlear implant*", "infant" "meta-analysis", "bilingual*", "L2", "young children" (in "Topic")

This rendered 4,713 results. The query link can be accessed here: https://www.webofscience.com/wos/woscc/summary/404fba16-6bfa-4f02-83e9-482eb8c1fa43-e2013de1/times-cited-descending/1. Note that this query link searches the same criteria we used, but on the collection of articles on the Web of Science at the time of the search. Therefore, the numbers and ordering of the articles will differ from those reported here.

We sorted the results by number of citations and exported them. Then, searching by descending number of citations, we manually inspected the method sections of the papers until we identified the 100 most cited papers that met our inclusion criteria.

Manual inspection of the Method sections of these 100 papers revealed:

- 62 papers specified that their participants were native speakers
- 3 indicated that participants were English speakers but did not provide information about native speaker status
- 35 did not make any mention of language background

See OSF page for the article coding spreadsheet.

Flexible Coding of Participant Responses

In order to give credit for phonologically plausible misspellings of words, we employed a flexible scoring system based on the freely available tool "Ponto" (Kessler 2017). The process for scoring keywords was identical for the word task (Experiment 1) and the final word of the sentence task (Experiment 2), and was implemented as follows:

1. Participant responses were stripped of extraneous punctuation and numbers (e.g., the response "cat;" would be changed to "cat")
2. Responses that are orthographically identical to the target are scored as correct
3. Responses that are homophones of the target word are scored as correct (e.g., "tied" for "tide"). Homophones were identified as words with different orthography but identical phonology according to the English Lexicon Project database (Balota et al. 2007).
4. Incorrect responses that form real English words (defined as appearing in the English Lexicon Project's list of 40,000 words) are counted as incorrect (e.g., the response "cast" substituted for "cat"). We decided not to change responses that form real English words because it is impossible to know whether the incorrect responses were the result of mishearing or mistyping, and we wanted to limit our corrections to words that were likely to have been mistyped.
5. The remaining incorrect, nonword responses were entered into the Ponto scoring web interface. Ponto compares the phonological transcription of the target word with possible pronunciations of the orthography of the participant's response. This is intended to answer the question: "Might this orthographic participant response reasonably be expected to be pronounced the same as the target?" The possible pronunciations of each letter or letter string (i.e., phoneme to orthography correspondences) were derived for Ponto from the Carnegie Mellon University Pronouncing Dictionary (CMUPD). The correspondences were created by identifying all the ways a given phoneme could be spelled in words that appear in the CMUPD. For example, /k/ can be spelled "k" (as in "kiss"), "ck" (as in "back"), "ch" (as in "chaos"), che (as in "headache"), and more. In Ponto, the difference between the target and the participant's response is given an "edit distance" that represents the changes necessary to transform the target to the response. All "legal" phoneme corrections according to the CMUPD were given an edit distance of 0. For example, if the target word was "headache," responses of "headak," "headack," and "headach" would all have an edit distance of 0, indicating that the final phoneme or phoneme clusters can be pronounced the same as the final phoneme cluster of the target (/k/). Other phoneme insertions or deletions (e.g., adding an "l" in the response "headlache") were given an edit distance of 1, and substitutions were given an edit distance of 1.4 (e.g., "headat").[1]
6. Words with an edit distance of 0 (e.g., "nat" for "gnat", "mil" for "mill" "tor" for "tore") were automatically counted as correct. Those with an edit distance of more than 2 (e.g., "luff" for "laugh") were counted as incorrect. Those with an edit distance between 0 and 2 were manually checked by two coders with training in phonetics and phonology. The coders were given the participant's orthographic transcription and the correct answer and asked to determine whether the most likely pronunciation of the orthographic response was the target. Typos in which the target word could be formed by a one-letter substitution formed from an adjacent key (e.g., "douby" for "doubt") were counted as correct. Given that only nonwords are entered into the Ponto scoring system and manual scoring systems, inflections (e.g., "cats" for "cat") were not counted as correct. Manual scoring was done without knowledge of condition (e.g., L1 or LX speaker, noise or quiet). Only items that both coders independently deemed correct were counted as correct.

Demographic and Language Background Questionnaire

The questions about language background were adapted from the LEAP-Q (Marian et al. 2007).

[1] Note that if a participant response has multiple routes to be corrected (e.g., different combinations of insertions and substitutions), the final Ponto score will be the smallest distance.

- What is your age? [free field]
- Are you of Hispanic/Latino origin?
    - Yes (Hispanic / Latino)
    - No (not Hispanic / Latino)
- Which of the following best describe(s) your race? You may pick more than one.
    - American Indian / Alaskan Native
    - Asian
    - Black / African American
    - Native Hawaiian / other Pacific Islander
    - White
- What is your biological sex? [free field]
- Which of the following best characterizes your experience with English?
    - I am a native English speaker (meaning I learned English before any other language)
    - I learned English simultaneously with another language
    - I learned another language before I learned English
- Please list all the languages you know, in order of dominance
- Please list all the languages you know, in order of acquisition (your native language first)
- Please list what percentage of the time you are currently and on average exposed to each language (your percentages should add up to 100%)
- Please check your highest education level (or the approximate US equivalent to a degree obtained in another country)
    - Less than high school
    - High school
    - Professional training
    - Some college
    - College
    - Some Graduate
    - Masters
    - PhD/MD/JD
- What age did you:
    - Begin acquiring English?
    - Become fluent in English? (write NA if you don't consider yourself fluent)
    - Begin reading in English?
    - Became fluent reading in English? (write NA if you don't consider yourself fluent)
- Please list the number of years and months you spent in each language environment:

	Years	Months
A country where English is spoken		
A family where English is spoken		
A school and/or working environment where English is spoken		

- On a scale from zero to ten, please select your *level of **proficiency*** in speaking, understanding, and reading English

	No Proficiency 0	5	High Proficiency 10
Speaking			
Understanding spoken language			
Reading			

Writing

- Please rate to what extent you are currently exposed to English in the following contexts:
    - Interacting with friends
    - Interacting with family
    - Watching TV
    - Listening to radio/music
    - Reading
    - Language-lab/self-instruction

  (scale)
  0 - never
  1 - almost never
  2
  3
  4
  5 - half of the time
  6
  7
  8
  9
  10 - always

- What country do you currently live in?
- What form of English are you most familiar with: American English or another form of English (e.g., British English, Australian English, etc).

Demographic and Language Background Information

This section contains information from the demographic and language experience survey (see accompanying R script at https://osf.io/tvwur/ for more details). We did not run any statistical analyses on these data, so the accompanying text is simply descriptive.

Language background

	Exp 1 (Mean, SD)		**Exp 2 (Mean, SD)**	
	L1	**LX**	**L1**	**LX**
Speaking proficiency	9.90 (0.36)	7.02 (1.85)	9.91 (0.48)	7.77 (1.89)
Understanding proficiency	9.91 (0.32)	8.21 (1.37)	9.89 (0.49)	8.45 (1.67)
Reading proficiency	9.87 (0.49)	9.03 (1.06)	9.89 (0.47)	9.08 (1.21)
Writing proficiency	9.80 (0.60)	7.65 (1.74)	9.78 (0.98)	7.88 (1.90)
Current exposure to English with friends	9.72 (1.02)	4.02 (2.72)	9.86 (0.56)	5.54 (3.22)
Current exposure to English with family	9.54 (1.49)	1.23 (1.88)	9.66 (1.00)	2.39 (2.79)
Current exposure to English with TV	9.24 (1.64)	6.76 (2.86)	9.37 (1.36)	7.70 (2.54)
Current exposure to English with	9.24 (1.35)	7.97 (2.39)	9.29 (1.39)	8.22 (1.92)

radio/music				
Current exposure to English with reading	9.72 (0.93)	6.96 (2.30)	9.83 (0.55)	7.14 (2.39)
Current exposure to English with self instruction	8.69 (3.20)	5.18 (3.09)	8.63 (3.16)	5.21 (3.26)
Age began acquiring English	0.98 (1.54)	7.84 (3.98)	1.11 (3.13)	7.65 (4.33)
Age of English fluency	4.86 (2.91)	17.18 (3.97)	3.90 (2.81)	17.06 (5.82)
Age began reading English	4.37 (1.49)	10.46 (4.27)	3.90 (1.51)	10.56 (4.78)
Age of reading fluently in English	6.41 (2.62)	15.70 (3.85)	5.81 (2.93)	15.21 (4.67)
Years spent in English-speaking country	35.21 (10.51)	2.51 (6.65)	34.65 (9.05)	5.42 (9.75)
Years spent in English-speaking family	34.57 (10.66)	1.87 (5.94)	34.00 (9.70)	4.21 (9.49)
Years spent in English-speaking school/working environment	30.16 (12.04)	5.96 (6.94)	29.07 (11.19)	7.18 (7.91)

Note: Proficiency and current exposure are self-rated values out of 10.

Education Level

	Exp 1		**Exp 2**	
	L1 (%)	**LX (%)**	**L1 (%)**	**LX (%)**
Less than high school	2	0.00	2.00	0.00
High school	17	23	20.67	16.00
Professional training	1	2	1.33	5.33
Some college	26	15	18.67	10.67
College	40	20	44	19.33
More than college	14	40	13.33	48.67

Note. Values represent percentage of participants reporting each education level

Form of English participants are most familiar with (e.g., American, British)

	Exp 1		**Exp 2**	
	L1 (%)	**LX (%)**	**L1 (%)**	**LX (%)**
American English	99	67	93.33	49.33
British English	0	23	0	36.00

	Exp 1		Exp 2	
Both American and British English	0	10	0	11.33
Other	1	0	0.67	3.33

Note. Values represent percentage of most familiar form(s) of English reported by participants

Country of Residence

For Experiment 1, the majority of LX participants resided in Poland (25%), Portugal (21%), Mexico (12%) and Italy (11%). The remaining participants (fewer than 10% of each) were from Belgium, Czech Republic, Finland, Germany, Greece, Hungary, Slovenia, South Africa, Spain, and the UK.

For Experiment 2, the majority of LX participants resided in the UK (27.33%), Portugal (18.67%), and Poland (14%). The others (fewer than 10% of each) were from Canada, Chile, Czech Republic, Finland, France, Germany, Greece, Hungary, Italy, Mexico, the Netherlands, Norway, Slovenia, South Africa, Spain, and the US.