Desperately seeking ‘English’ sibilants
Discovering dialect norms and speaker variability for /sʃ/ from large-scale, multi-dialect analysis

Jane Stuart-Smith* Morgan Sonderegger+
Jeff Mielke~ James Tanner+ Vanna Willerton+
The SPADE Consortium
*Glasgow University Laboratory of Phonetics (GULP),
+McGill University, ~North Carolina State University

LabPhon17 (virtual) July 2020
Structured variability in English /sʃ/

• Lexical contrast
e.g. Kleinschmidt (2019)

• Gender
e.g. Flipsen et al (1999); Strand e.g. (2000)

• By dialect? Sibilants in English are assumed be highly similar
e.g. Wells (1982)
Structured variability in English sibilants

• Lexical contrast
e.g. Kleinschmidt (2019)

• Gender
e.g. Flipsen et al (1999); Strand e.g. (2000)

• By dialect? Sibilants in English are assumed be highly similar – with the odd exception

Macafee (1983)
Structured variability in English /s/ʃ/

• Lexical contrast
  e.g. Kleinschmidt (2019)

• Gender
  e.g. Flipsen et al (1999); Strand e.g. (2000)

• By speaker? English /s/ is highly variable, and carries a range of social-indexical meanings
  e.g. Levon/Holmes-Elliott (2017); Stuart-Smith (2007; 2020); Podesva/van Hofwegen 2016)
Are English sibilants really both highly similar (across dialects) and variable (across speakers within dialects)?

Kleinschmidt (2019); cf Chodroff/Wilson (2018)
Research questions

1. How variable is the production of /sʃ/ ... across English dialects? ... across speakers within English dialects?
2. How do degrees of by-speaker and by-dialect variability compare for English?

Kleinschmidt (2019); cf Chodroff/Wilson (2018)
Research questions

1. How variable is the production of /sʃ/ across English dialects?
   ... across speakers within English dialects?
2. How do degrees of by-speaker and by-dialect variability compare for English?
3. Is English /s/ more variable than /ʃ/?

Research questions

1. How variable is the production of /sʃ/
   ... across English dialects?
   ... across speakers within English dialects?

2. How do degrees of by-speaker and by-dialect variability compare for English?

3. Is English /s/ more variable than /ʃ/?

4. What’s the ‘best’ way to acoustically analyse sibilants across multiple datasets and recording conditions?
Software large-scale automatic acoustic analysis of speech datasets

Data from ~40 datasets (socio)linguistic surveys

Research investigate how ‘English’ varies in time and space

Sonderegger et al (in press)
Tanner et al (2020)
McAuliffe et al (2019)
Mielke et al (2019)
Stuart-Smith et al (2019)

Project website: [http://spade.glasgow.ac.uk/](http://spade.glasgow.ac.uk/)
Sample – The SPADE Consortium

- 16 corpora (public/private; 12 spontaneous)
- 34 dialect groupings (‘dialects’)
- 1,925 speakers (female: 1,008)
- [https://spade.glasgow.ac.uk/the-spade-consortium/](https://spade.glasgow.ac.uk/the-spade-consortium/)
US: 286 speakers (150 F)
7 dialects (4 corpora)

Canada: 118 speakers (60 F)
3 dialects (2 corpora)

Scotland: 296 speakers (167 F)
9 dialects (3 corpora)

England: 969 speakers (512 F)
10 dialects (6 corpora)

Ireland: 244 speakers (113 F)
4 dialects (3 corpora)

Wales: 12 speakers (6 F)
1 dialect (1 corpus)

https://spade.glasgow.ac.uk/
/the-spade-consortium/

Details on dialects in Extra Slides
Analysis 1 (LPh abstract)
FFT spectra (in praat)
Centre of Gravity (COG)
1-8kHz
central 50% sibilant

Analysis 2
multitaper spectra (in R)
Spectral peak
s: 2-9kHz
sh: 2-7kHz
25ms window at midpoint

Results presented here (= Analysis 3)
• multitaper spectra (in R)
• Centre of Gravity (COG)
• 1-8kHz
• 25ms window at midpoint

all instances of
word-initial,
stressed,
prevocalic /sʃ/
multitaper spectrum of /s/ from *seven*, middle-class man, Glasgow

Center of Gravity (COG) treats sibilant spectrum as a mean.

**COG** is strongly affected by frequency range (e.g. corpus sampling rate), and is sensitive to factors such as recording setup, e.g. microphone, etc.

![Graph showing multitaper spectrum of /s/ from seven, middle-class man, Glasgow](image)

**COG = 5994Hz**

e.g. Forrest et al 1988; Wrench (1995); Jongman et al (2000)
multitaper spectrum of /s/ from *seven*, middle-class man, Glasgow

**Peak** 5912 Hz  
**COG** 5994 Hz

**Peak** 2437Hz  
**COG** 4379Hz

**Spectral peak** assumes peaks (and troughs) of sibilant spectra arise from acoustic and aerodynamic principles

multitaper spectrum of /ʃ/ from *shelf*, same speaker  
e.g. Shadle 1995; Jesus/Shadle 2002; Koenig et al (2013)
multitaper spectrum of /s/ for seven, middle-class woman, Glasgow

Spectral peak is sensitive to the range from which it is selected.

A lower frequency-bounded peak may not always coincide with the sibilant spectral prominence (PEAK).

multitaper spectrum of /s/ for seven, middle-class woman, Glasgow

Spectral peak is sensitive to the range from which it is selected.

A lower frequency-bounded peak may not always coincide with the sibilant spectral prominence (PEAK).

Peak (3-5kHz): 4354Hz
Peak (2-7kHz): 6591Hz

multitaper spectrum of /s/ for seven, middle-class woman, Glasgow

Spectral peak is sensitive to the range from which it is selected.

A lower frequency-bounded peak may not always coincide with the sibilant spectral prominence (PEAK).

Peak (3-5kHz): 4354Hz
Peak (2-7kHz): 6591Hz
PEAK (2-9kHz): 8149Hz

frequency-bounded peaks give distributions with sharp right-skew, less amenable to modelling

Spectral peak (2-9kHz) multitaper spectra

Peak needs careful specification to capture the right peak.

Spectral peak (2-7kHz) multitaper spectra

n = 179,693, from Analysis 2
We used COG from multitaper spectra. COG provided sibilant distributions more amenable to modelling.

COG (1-8kHz) **multitaper** spectra

$\text{COG(1-8kHz) } \textbf{praat (FFT) spectra}$

$r = 0.91; n = 199,851$

cf Reidy (2015)
Analysis using Integrated Speech Corpus ANAlysis (ISCAN)


ISCAN script selected tokens, took praat measures -> Rscript to run and add multitaper measures
Data filtering

Removed tokens with COG or peak < 2400

By-speaker plots identified spectral problems, e.g. this male speaker from ICE-Canada

compare this female speaker from the same corpus, with better energy pattern
• Bayesian linear mixed-effects model
  • Stan/brms (Carpenter et al., 2017; Bürkner 2018)

• Dependent variable:
  • COG from multitaper spectra

• Random effects:
  • Dialect (34)
  • Speaker, within dialect (1,925)
  • Word (1,401)
• Dependent variable:
  • COG from multitaper spectra, 1-8kHz

• Predictors:
  • Gender (M/F)
  • gender:onset
  • Phone duration

\[
\begin{array}{c|c}
S & \int \\
\end{array}
\]

- varies by dialect
- varies by speaker (within dialect)
‘English’ sibilants by onset and gender

• /s/ is higher than /ʃ/
• female speakers show higher COGs than males
• the s/ʃ difference is very slightly smaller for male than female speakers
the s/ʃ difference varies by dialect, but this variability is fairly small compared to the overall ‘English’ difference between /s/ and /ʃ/

the average sibilant mean, and for /s/, /ʃ/ alone, vary more by dialect
• anomalous lowered COG (Can.various, Ire.various) and higher COG (Hastings) seem to reflect actual spectral patterns
There is more variability across speakers within dialects, than across dialects

COG: /s/-/ʃ/ difference

COG: /s//ʃ/ average

Speaker variability (Hz)

Dialect variability (Hz)

98% post. prob.

76% post. prob.

75% post. prob.

98% post. prob.
/s/ is more variable than /ʃ/ both across dialects, and across speakers within dialects.
How do ‘English’ sibilants look on scaling up?

- Lexical contrast: /s/ always distinct from /ʃ/

- Gender: female speakers show higher COGs than male speakers

  - confirms assumptions from ‘current-scale’ studies
  - begins to provide some overall ‘norms’ for ‘English’ sibilants
  - and to define frame of reference for structured variability (cf Kleinschmidt 2019)
How do ‘English’ sibilants look on scaling up?

- Social gender: female speakers show slightly more separation of /s/ and /ʃ/ than males
  
  - this result weakened as the dialect sample increased, and we changed measure from spectral peak to COG
  - will this evaporate if we use an auditorily-transformed measure, e.g. Bark/ERB?
How variable is the production of ‘English’ /s ʃ/?

- Sibilant production varies a bit across English dialects

- the difference between /s/ and /ʃ/ varies rather little by dialect, compared to overall ‘English’ /s/ vs /ʃ/

- variance for overall sibilant mean, and /s/ and /ʃ/ separately, shows more differences across dialects

- some exceptions, e.g. Irish is low; Hastings is high
How does by-speaker and by-dialect variability compare for ‘English’ sibilants?

• There does seem to be an asymmetry, such that sibilants look more variable by speaker than by dialect

➢ strong evidence for variability by-speaker > by-dialect for s/ʃ difference, and /ʃ/ alone
➢ weaker evidence for overall sibilant mean and just /s/
➢ the asymmetry weakened as the dialect sample increased, and we changed measure from praat-derived to multitaper-derived COG
How does by-speaker and by-dialect variability compare for ‘English’ sibilants?

• There does seem to be an asymmetry, such that sibilants look more variable by speaker than by dialect

➢ the social-indexical informativity of sibilants may have the potential to be more important than group-level speaker attributes like dialect

➢ English sibilants may be different from, e.g. English stops, which show asymmetry in the other direction, i.e. more variable by dialect, less by speaker (Kleinschmidt 2019, Tanner et al 2020)
Is English /s/ more variable than /ʃ/?

• Yes: /s/ is more variable than /ʃ/ across dialects, and across speakers within dialects

➢ Chodroff/Wilson also note less evidence for ‘contrast uniformity’ (place of articulation) than ‘target uniformity’, in their expansion of principles of uniformity (e.g. Chodroff/Wilson 2017)

➢ Does this asymmetry occur because /s/ usually carries more social-indexical information than /ʃ/, or because inherently it can? If so, why?

➢ (Is this asymmetry partly an artefact of using a non-auditory scale?)
Next steps in (desperately) seeking ‘English’ sibilants

- missing factors: e.g. **vowel rounding, speech style, age, ethnicity**...
- how much does changing measure matter, e.g. spectral estimator, or from peak to COG?
- is there a better measure for sibilant characteristics for scaled-up analyses?
Next steps in (desperately) seeking ‘English’ sibilants

- missing factors: e.g. **vowel rounding, speech style, age, ethnicity**...
- how much does changing measure matter, e.g. spectral estimator, or from peak to COG?
- is there a better measure for sibilant characteristics for scaled-up analyses?
- How much does increasing/changing dialect sample influence the norms, and hence the margins, of ‘English’ sibilants?

The search continues...
Thank you!

Data Guardians in the SPADE Consortium

Rachel Macdonald, project manager and our teams of assistants; plus our other PIs

http://spade.glasgow.ac.uk/

Try out our prototype Shiny app – to look at time/space visualisation of SPADE data:

http://152.1.64.33/spade/latest
Stacey Harkin
Kirsty McCahill
Mitchell McGee
Edward Marshall
Julia Moreno
Jo Pearce
Niamh Walker
Ewa Wanat

Arlie Coles
(U. de Montréal)

Elias Stengel-Eskin (Johns Hopkins)

Michael Goodale,
Sarah Mihuc
(McGill)

Vanna Willerton
(McGill)

Jordan Holley
Peter Andrews
Kaylynn Gunter
Extra slides
Columbus (Buckeye; white)
US.N.Cities (Santa Barbara corpus: Northern Cities, New England, NYC, Lower South; ethnicity varies)
US.West (Santa Barbara corpus: California, Oregon; ethnicity varies)
US.DC.AAVE (CORAAL: Washington DC)
US.NC.AAVE (CORAAL: Princeville NC)
US.NY.AAVE (CORAAL: Rochester NY)
Raleigh (white)

286 speakers (150 female)

Can.Prair.Urb (Canadian Prairies) Can.Prair.Rur (Canadian Prairies) Can.various (ICE Canada) (ethnicity varies in all three corpora) 118 speakers (60 female)

https://spade.glasgow.ac.uk/the-spade-consortium/
https://spade.glasgow.ac.uk/the-spade-consortium/

read and spontaneous speech

Glasgow (SOTC; white)
Sc.Centrl (SCOTS; not recorded)
Sc.Edin (SCOTS; not recorded)
Sc.Gla (SCOTS; not recorded)
Sc.North (SCOTS; not recorded)
Sc.Isles (SCOTS; not recorded)
SSE (Doubletalk; white)
Scotland.EDA (EDA; ethnicity varies)
Scot.North.EDA (EDA; ethnicity varies)

296 speakers (167 female)

RP (Modern RP; ethnicity not recorded)
SSBE (DiVyS; ethnicity not recorded)
Hastings (white)
Bradfd.Pjabi (IViE; Panjabi-English)
Cambridge (IViE; white)
Leeds (IViE; white)
London.Jamaican (IViE)
Newcastle (IViE; white)
Eng.North.EDA (EDA; ethnicity varies)

969 speakers (512 female)

Ir.various (Irish; ethnicity not recorded)
Belfast (IViE; white)
Dublin (IViE; white)
Ireland.EDA (EDA; ethnicity varies)

244 speakers (113 female)

Cardiff (IViE; white)

12 speakers (6 female)
Datasets (speech corpora, lexicons) → Database → Set of linguistic objects → Data file (CSV)

*import* → *querying* → *export*

*add measures & structure*

**Implementation**

- Python API: [https://polyglotdb.readthedocs.io/](https://polyglotdb.readthedocs.io/)
- GUI: [https://iscan.readthedocs.io/](https://iscan.readthedocs.io/)
English speech over time and space

http://152.1.64.33/spade/latest/

Try out our Shiny app!