

Sound change, priming, salience

Producing and perceiving variation
in Liverpool English

Marten Juskan

Draft
of 9th October 2018, 09:58

Language Variation

Editors: John Nerbonne, Dirk Geeraerts

In this series:

1. Côté, Marie-Hélène, Remco Knooihuizen and John Nerbonne (eds.). The future of dialects.
2. Schäfer, Lea. Sprachliche Imitation: Jiddisch in der deutschsprachigen Literatur (18.–20. Jahrhundert).
3. Juskan, Marten. Sound change, priming, salience: Producing and perceiving variation in Liverpool English.

ISSN: 2366-7818

Sound change, priming, salience

Producing and perceiving variation
in Liverpool English

Marten Juskan

Marten Juskan. 2018. *Sound change, priming, salience: Producing and perceiving variation in Liverpool English* (Language Variation 3). Berlin: Language Science Press.

This title can be downloaded at:

<http://langsci-press.org/catalog/book/210>

© 2018, Marten Juskan

Published under the Creative Commons Attribution 4.0 Licence (CC BY 4.0):

<http://creativecommons.org/licenses/by/4.0/>

ISBN: 978-3-96110-119-1 (Digital)

978-3-96110-120-7 (Hardcover)

ISSN: 2366-7818

DOI:[10.5281/zenodo.1451308](https://doi.org/10.5281/zenodo.1451308)

Source code available from www.github.com/langsci/210

Collaborative reading: paperhive.org/documents/remote?type=langsci&id=210

Cover and concept of design: Ulrike Harbort

Fonts: Linux Libertine, Libertinus Math, Arimo, DejaVu Sans Mono

Typesetting software: Xe_{La}T_EX

Language Science Press

Unter den Linden 6

10099 Berlin, Germany

langsci-press.org

Storage and cataloguing done by FU Berlin

[./langsci/graphics/storage/logo.pdf](#)

To Daniela

Contents

1	Introduction	1
1.1	Intentions – what this study is about	1
1.2	Restrictions – what this study is not about	3
1.3	Structure of the book	4
2	A brief history of Liverpool and Scouse	7
2.1	The first 600 years	7
2.2	19 th century	10
2.3	20 th century	14
2.3.1	Enregisterment and the ‘Scouse industry’	14
2.3.2	Decline	16
2.3.3	Regeneration	19
2.4	21 st century – outlook	21
2.5	Summary	23
3	Variables	25
3.1	General remarks	25
3.2	Supragsegmentals	26
3.3	Consonants	27
3.3.1	/ŋ(g)/	28
3.3.2	Lenition (of /k/)	29
3.4	Vowels	32
3.4.1	happy	33
3.4.2	NURSE – SQUARE	35
3.5	Summary	38
4	A few words on salience and exemplar theory	39
4.1	Salience	39
4.1.1	Salience and circularity	39
4.1.2	Cognitive vs. social salience	41
4.1.3	Salience in this study	43

Contents

4.2	Exemplar theory	47
4.2.1	Basic principles	48
4.2.2	Application in (socio-)linguistics	49
4.2.3	Frequency and salience in exemplar theory	53
4.3	Summary	56
5	Interview method	57
5.1	Interview structure	57
5.2	Participants	59
5.3	Transcription	62
5.4	Measuring	62
5.4.1	Consonants	62
5.4.2	Vowels	65
5.4.3	Normalisation	66
5.5	Phonological context	74
5.6	Statistical analysis	76
6	Vowel production	81
6.1	happy	82
6.1.1	F1 (happy)	82
6.1.2	F2 (happy)	93
6.1.3	Synthesis and Pillai scores (happy)	104
6.2	NURSE	108
6.2.1	F1 (NURSE)	108
6.2.2	F2 (NURSE)	118
6.2.3	Synthesis and Pillai scores (NURSE)	131
7	Consonant production	141
7.1	/ŋ(g)/	142
7.1.1	Overview	142
7.1.2	Phonological context	142
7.1.3	Style and gender	145
7.1.4	Age and social class	147
7.1.5	Style shifting	150
7.2	/k/	151
7.2.1	Overview	151
7.2.2	Phonological context	153
7.2.3	Style and gender	156
7.2.4	Style and social class	158

7.2.5	Age and gender	159
7.2.6	Age and social class	163
7.2.7	Social class and gender	164
7.2.8	Style shifting	165
8	Awareness, comments, evaluation	171
8.1	Scouse and ‘Liverpoolness’	171
8.1.1	Accent and identity	171
8.1.2	Distinctness, geographical spread, and ‘plastic’ Scousers	172
8.1.3	In the north, but not of it?	174
8.2	Features of Scouse	177
8.2.1	Geographical variation	177
8.2.2	Suprasegmentals	178
8.2.3	Phonological variables	179
8.3	Evolution and evaluation of Scouse	185
8.3.1	Old speakers	185
8.3.2	Middle-aged speakers	188
8.3.3	Young speakers	191
9	Discussion (production)	197
9.1	happy: Indicator (of northernness)	197
9.1.1	Overall age differences	197
9.1.2	Gender and class	198
9.1.3	Style shifting and awareness	200
9.1.4	Classification	201
9.2	NURSE: Marker to stereotype and back again	201
9.2.1	Overall age differences	201
9.2.2	Gender and social class	204
9.2.3	Style shifting and awareness	205
9.2.4	Classification	208
9.3	Velar nasal plus: Indicator with prestige option	208
9.3.1	Age, class, and gender	208
9.3.2	Style shifting and awareness	210
9.3.3	Classification	211
9.4	Lenition: From indicator to stereotype	212
9.4.1	Age	212
9.4.2	Gender and class	213
9.4.3	Style shifting	214
9.4.4	Awareness and attitudes	216

Contents

9.4.5	Classification	217
9.5	Summary	218
10	Experiment design	221
10.1	Stimuli	221
10.1.1	Stimuli sentences and frequency of keywords	221
10.1.2	Vowel stimuli	225
10.1.3	Consonant stimuli	226
10.2	Presentation	228
10.2.1	Online platform	228
10.2.2	Reaction times	231
10.3	Participants	232
10.4	Statistical analysis	235
11	Perception results	237
11.1	happy	238
11.1.1	Overview	238
11.1.2	Prime	239
11.1.3	Position of carrier word	240
11.1.4	Geographical distance	241
11.1.5	Frequency	242
11.1.6	Age	244
11.2	NURSE	245
11.2.1	Overview	245
11.2.2	Prime	245
11.2.3	Position of carrier word	247
11.2.4	Stimulus order	248
11.3	/ŋ(g)/	251
11.3.1	Overview	251
11.3.2	Prime	252
11.3.3	Phonological context	253
11.3.4	Age	255
11.4	/k/	256
11.4.1	Overview	256
11.4.2	Prime and social class	257
11.4.3	Phonological context	261
11.4.4	Geographical distance	263
11.4.5	Frequency	265

12 Discussion (Perception)	267
12.1 Priming and salience	267
12.1.1 Salient vs. non-salient variables	267
12.1.2 Degree of priming and accuracy	268
12.1.3 Salience as likelihood of remembrance	269
12.2 Social factors	272
12.2.1 Social class	272
12.2.2 Gender	273
12.2.3 Age	274
12.2.4 Geographical distance	276
12.3 Non-social factors	277
12.3.1 Time held in memory	277
12.3.2 Stimulus order	279
12.3.3 Phonological environment	280
12.4 Issues and limitations	283
12.4.1 The problem of velar nasal plus	283
12.4.2 Comparability with previous research	285
12.4.3 The issue of frequency	286
12.4.4 Size of the priming effect	287
12.5 Direction of priming	287
12.5.1 The problem of replicating priming experiments	287
12.5.2 Hay et al.'s explanation for inverted effects	289
12.5.3 Assimilation and contrast effects	290
12.6 Summary and implications	293
13 Conclusion	295
Appendix A: Questionnaire	299
Appendix B: Reading passage	303
Appendix C: Word list	305
Appendix D: Stimuli for perception test	307
References	309

10 Experiment design

Based on a detailed analysis of production data, the first part of this study has independently established the sociolinguistic salience of happy-tensing, velar nasal plus, the NURSE-SQUARE merger, and lenition of /k/. Knowing they can be ordered in this way (from least to most salient) it is now possible to test the hypothesis that salience is a crucial factor in exemplar priming experiments. Perception of these variables was analysed with the help of an online test, the detailed methodology of which is described below.

10.1 Stimuli

10.1.1 Stimuli sentences and frequency of keywords

To ensure comparability, stimuli were designed in a way similar to that of Hay, Nolan, et al. 2006 and Niedzielski 1999. Six keywords for each of the four variables were taken from the word list that had been used to elicit production data (see appendix C). In the test, the keywords appeared twice within complete sentences, once sentence-finally and once sentence-medially. This way it was possible to investigate whether having to hold the relevant sound in memory had an impact on subjects' responses (cf. Hay, Nolan, et al. 2006). Hay, Nolan, et al. (2006) noted a problem in their methodology, namely the fact that their stimuli confounded the position of the keyword and the total length of the sentence it was embedded in. Since sentences with the keyword in the middle were always (considerably) longer than the corresponding sentences with the keyword at the end there was no way of determining whether differences found were due to the fact that participants had to hold the sound in memory or that the longer sentence just contained more phonetic material, which might have activated further (or different!) exemplars.

In this study, care was therefore taken to make sure that all stimuli sentences were (roughly) equal in length (measured in words). Total duration (in seconds and milliseconds) of the final stimuli was also measured and a paired t-test revealed that stimuli sentences with the keyword in the middle were not significantly longer than sentences where the keyword was at the end ($t(23) = -0.129$,

$p = 0.898$). For /ŋ(g)/ and /k/, a further criterion was to have an equal number of words with the variable in intervocalic and in word-final position. This was to match the most prominent phonological contexts that these variables were investigated in in production and to be able to discover any potential differences in priming related to this criterion. Occurrences of the other three test variables were avoided in the stimuli sentences of the fourth variable. An example pair is given below (see appendix D for the complete list):

1. People in that town almost never went to *church*.
2. In that town *church* was not popular with people.

Key words from the original list were selected in such a way that there was a continuum from very high to very low frequency words. Frequency categorisation was initially based on occurrences in the BNC (spoken language section only). Table 10.1 provides an overview of the keywords used. All figures are absolute occurrences in the BNC (spoken), i.e. relative occurrences per 10 million words. These figures were later replaced by frequencies based on SUBTLEX-UK, a 200+ million word corpus of subtitles collected from BBC broadcasts (taken from 9 different channels) in the years 2010–2012. SUBTLEX-UK frequencies were preferred because they have been found to explain more variance than, for example, BNC frequencies in psycholinguistic experiments focussing on things such as lexical decision reaction times (Van Heuven et al. 2014). According to Van Heuven et al. (2014), one of the most popular frequency measures, frequency per million words, comes with a number of problems, especially for low frequency words. They propose an alternative measure called the Zipf scale, the values of which are calculated as follows:

$$\text{Zipf score} = \log(n) + 3 \quad (10.1)$$

Where n is the absolute frequency in a 1 million word corpus. This logarithmic scale produces values of ‘1’ for words that occur once in 100 million words, ‘2’ for words that occur once in 10 million words, ‘3’ for words that occur once in 1 million words, etc., with a range of about 1 (very low frequency words) to 6 or 7 (very high frequency words). The actual equation used in practice is slightly more complex in order to also assign Zipf scores to words that have an absolute frequency of 0 in the corpus, but this does not drastically change the interpretation of the Zipf scores as outlined just above (cf. Van Heuven et al. 2014). The Zipf scores used in this study are based on the frequencies in SUBTLEX-UK as a whole (instead of one of the two sub-corpora ‘CBeebies’ (pre-school children)

and ‘CBBC’ (primary school audience), for which separate Zipf values are also available).

Table 10.1: BNC and SUBTLEX-UK frequencies of keywords in the perception test

NURSE			happy		
word	BNC	Zipf	word	BNC	Zipf
turn	2572	5.45	happy	1820	5.56
word	2222	5.29	city	1486	5.40
girl	1613	5.29	pretty	1455	5.50
church	1149	5.02	baby	962	5.29
shirt	218	4.47	lazy	101	4.07
fur	40	4.03	stingy	5	3.02
/k/			/ŋg/		
like	38153	6.53	young	1890	5.51
book	2396	5.21	song	386	5.10
pack	425	4.62	gang	119	4.39
chicken	479	4.82	singer	109	4.48
snooker	71	4.25	hanger	27	3.13
hacker	3	3.88	longish	4	2.13

The author is aware of the fact that this list probably leaves much to be desired from the point of view of linguistic frequency research. Ideally, there would have been two words each in more clearly (and more consistently) defined high, middle and low frequency ranges for every variable investigated: [Van Heuven et al. \(2014\)](#) draw the line between low frequency and high frequency words somewhere between Zipf values 3 and 4, which would classify only 1 in 6 of the carrier words as low frequency and the remainder as high frequency. This was difficult, however, since the group of possible keywords was already restricted by the word list that had been used in the interview. It should be borne in mind, however, that frequency is only of secondary interest in this study, and the opportunity to possibly find connections between use and perception of particular words was deemed more relevant to the task at hand. The whole design of the experiment would have looked different if frequency had been of central importance. For instance, more keywords would be needed, because with only 6 words (per variable) chances are high that any effects of frequency that might be found

are, at least in part, *lexical* effects of individual words rather than (generalisable) frequency effects. Anything this study might find about frequency effects should therefore be considered a bonus rather than one of the main points.

There were thus a total of 48 stimuli sentences in the perception test, with the 12 test tokens of one variable simultaneously functioning as dummies for the other three. These sentences were recorded by a phonetician in her late twenties who is originally from East Central Manchester. She used her native Mancunian accent, meaning that *NURSE* was realised as [ɜ:], *happy* as [ɪ] (phrase-internally) or [ə] (phrase-finally), *velar nasal plus* as [ŋg]¹ and /k/ as [k].

The stimuli participants had to choose from were created in Praat. A script extracted the key word from every sentence, manipulated the relevant sound and saved the resulting four different versions of the word. This study differs from many previous studies, e.g. Hay, Nolan, et al. 2006 or Niedzielski 1999, in that participants were presented with whole words *containing* manipulated sounds instead of just individual vowels. This choice was made because (a) it made the experimental situation ever so slightly less artificial, and (b) the manipulation of /k/ was based on differing ratios of silence and friction, and therefore depended on the /k/ having some immediately preceding phonetic material. A second deviation from the seminal studies concerns the number of answer tokens. Both Niedzielski and Hay, Nolan, et al. presented their participants with a 6-step continuum, but the present study had only 4 options that listeners could choose from. This reduction was done for two reasons. Firstly, creating more than 4 different realisations that can be reliably distinguished by non-experts would have been rather challenging in the case of *velar nasal plus* and /k/ lenition. Since it was deemed essential to ensure comparability between the consonantal and the vocalic variables, the number of answer tokens was therefore limited to 4 for the whole experiment. Secondly (and more importantly), it is not at all clear whether non-linguists are even able to make full use of an answer continuum that consists of 6 tokens which differ only in rather fine phonetic details. Actually, Niedzielski 1999 ended up limiting her analysis to the 3 ‘middle’ tokens of her continuum because subjects had almost never chosen one of the other 3 options (cf. Niedzielski 1999: 64–65). As far as the present study is concerned, it was therefore decided that a 4-step continuum was difficult enough, and that in inflating the amount of answer tokens one would run the risk of potentially asking too much of the

¹Using *velar nasal plus* as a variable in a priming test where the conditions are ‘Liverpool’ and ‘Manchester’ is problematic, because *velar nasal plus* is actually present in both accents. See §12.4 for an explanation why I still think it was justified to use this variable in this priming experiment.

average participant. An even number of stimuli was chosen to ‘force’ listeners to one side of the spectrum (more Mancunian or more Liverpoolian).

10.1.2 Vowel stimuli

Vowel stimuli were created using a script that was heavily inspired by one written by Styler (2008). First of all, the recording was downsampled to 11 kHz. The vowels in the key words were then extracted individually and run through source-filter synthesis in Praat (Boersma & Weenink 2015). An LPC (prediction order 12, window length 25 milliseconds, time step 5 milliseconds, pre-emphasis frequency 50 Hz) and a formant object (maximum number of formants 5, maximum formant frequency 5500 Hz, window length 25 milliseconds, pre-emphasis frequency 50 Hz) were created, both using the Burg algorithm. The downsampled vowel was then inverse-filtered with the LPC, which results in a ‘reconstruction’ of the pure source sound as generated by the vocal folds of the speaker. By applying the formant filter (which simulates vocal tract shape) to this reconstructed source a vowel can be re-synthesised that is almost identical to the natural one in the recording (token 2 was created this way). If the filter (i.e. the formant structure) is manipulated, a synthesised vowel is generated that is acoustically and perceptually different from the original, but most of the time still fairly natural sounding. The re-synthesised vowel was then upsampled again and spliced back into the carrier word. F1 and F2 of NURSE and happy were manipulated using this method. The relevant values were chosen in such a way that there was roughly equal perceptual distance (auditory judgement by the author) between any two adjacent vowels.

Table 10.2: Filter settings used in vowel resynthesis (in Hz)

	NURSE		happy	
	F1	F2	F1	F2
stimulus 1	-150	-250	+100	-200
stimulus 2	0	0	0	0
stimulus 3	+100	+150	-100	+200
stimulus 4	+200	+300	-200	+300

Table 10.2 summarises the final parameters, and Figure 10.1 illustrates the result for one set of NURSE tokens. There was thus a continuum from stimulus 1 (hyper-Mancunian), to 2 (actual) and 3 (Scouse), to stimulus 4 (hyper-Scouse).



Figure 10.1: Formant tracks of *fur* tokens

The formant structure of stimulus 2 was not modified, but still re-synthesised so as not to stick out as the only completely natural recording. For *NURSE* stimulus 1 was thus higher and further back than the speaker’s natural realisation, while stimuli 3 and 4 were both fronted and lowered to varying degrees. With respect to *happy*, stimulus 1 was lowered and centralised, while stimuli 3 and 4 were higher and fronter than the vowel that actually occurred in the recordings. Participants always heard the re-synthesised words in the same order (1 to 4). Unfortunately, it was not possible to synthesise stimuli of satisfactory quality from the sentences that had *happy* in phrase-final position (where it was typically realised as [ə], and often articulated with breathy voice). For this reason, answer tokens for these sentences were taken from the equivalent recordings where *happy* occurred in the middle of the sentence (and in the same carrier word). Crucially, however, the realisation of *happy* in the stimulus sentence itself was not altered. With hindsight, this should have been avoided as this procedure created a confound with the independent variable ‘position of keyword in sentence’ (cf. §11.1).

10.1.3 Consonant stimuli

An equivalent procedure was developed for the consonants /ŋ(g)/ and /k/. The link between those two in the context of this study is that the proportion of friction/aspiration is higher in the Scouse than in the standard British English variants, and this criterion was used to build a continuum similar to the one created for the two vowels. Once more, the goal was to create roughly equal perceptual

distance (again checked auditorily by the author) between any two tokens, just as for the re-synthesised vowels. For every keyword a TextGrid was prepared which marked phases of aspiration, burst, and silence (for /k/) or nasality (for /ŋ(g)/) respectively (cf. Figure 10.2). For /ŋ(g)/, a script written by the author of this study then

1. cut away the aspiration for stimulus 3,
2. additionally cut away the burst for stimulus 2 (leaving only the nasal),
3. and shortened the (often rather long) nasal by 25% to arrive at a more standard-like length for stimulus 1.

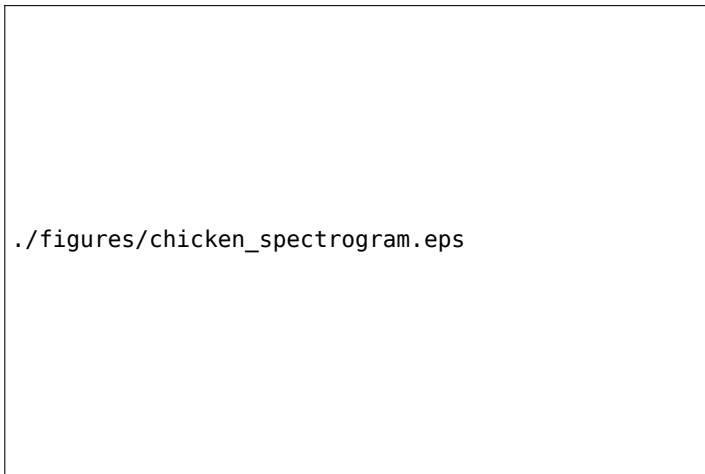


Figure 10.2: Waveform and TextGrid for *chicken* stimuli

Stimulus 4 was the full, unaltered velar nasal plus realisation [ŋg]. To make the velar nasal plus sentences comparable to those of the other variables, stimulus 2 was afterwards copied back into the sentence to replace the natural [ŋg] realisation. There was thus no voiced velar plosive present in the sentences subjects heard, and the continuum of stimuli was structured in the same way as those for the vowels and for /k/ in the sense that stimulus number 2 was the objectively most accurate choice that corresponded best to the sound actually used in the sentence.

Stimulus 2 for /k/ was the actual released plosive [k] as it occurred in the sentence recordings. The hyper-Mancunian stimulus was created from this by

cutting the aspiration, but leaving in the burst. The speaker had also recorded all the /k/ sentences using the Scouse fricative variants. The frication from these variants was pasted into the place of the original aspiration to form an affricate stimulus 3 (to make the result more natural sounding half of the closure phase was also deleted). Stimulus 4, finally, had the whole plosive (silence, burst, and aspiration) replaced by the fricative by way of removing remaining silence and burst from stimulus 3.

For both /ŋ(g)/ and /k/, asymmetrical cross-fading and intensity adaptation was applied in intervocalic environments to create a smoother and less artificial transition from the nasal or the aspiration-less plosive to the following vowel in the hyper-Mancunian tokens (using a 25 milliseconds fade-in interval for the vowel and 50 milliseconds of overlap between phonemes when concatenating). Table 10.3 provides an overview of the consonantal stimuli.

Table 10.3: Structure of consonant stimuli

	/ŋg/	/k/
stimulus 1	nasal only (shortened)	plosive with burst
stimulus 2	nasal only	plosive plus burst & aspiration
stimulus 3	nasal plus burst	plosive plus burst & frication
stimulus 4	nasal plus burst & aspiration	fricative

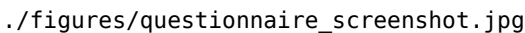
A small pilot study was run among linguists of the English Seminar at the University of Freiburg to make sure the stimuli (both the vocalic and the consonantal ones) were (a) sufficiently natural-sounding, and (b) equally distant from one another in perceptual terms. This pilot study did not reveal any problems, so the investigation proper was carried out using these stimuli.

10.2 Presentation

10.2.1 Online platform

The actual test was administered online using *SoSciSurvey.de* (2015), a professional tool for academic online surveys and questionnaires. The platform uses flash to play audio and video files. Although declining in importance, flash is still installed on most desktop and laptop computers (though not on tablets and smartphones) so the vast majority of potential subjects should have had no technical problems related to the website. Nevertheless, participants first of all had

to answer a filter question to make sure they had flash installed and activated and could actually play the sound recordings. Subjects were given the hint that they were going to hear the name of a type of bird and the questionnaire then used the flash plug-in to play a short sound file of the speaker saying *raven*. Participants then typed in the word they had heard. Wrong answers to this test question prevented the user from progressing in the questionnaire. Participants were randomly assigned to one of two groups. The control group was (correctly) told that the speaker they were going to listen to was from Manchester. The other group was led to believe that the speaker was from Liverpool (about 1 in 3 of the participants who lived outside of Liverpool actually believed this). Depending on which group subjects had been assigned to, 'Manchester' or 'Liverpool' was displayed at the top of every page as a reminder (cf. Figure 10.3).



./figures/questionnaire_screenshot.jpg

Figure 10.3: Online questionnaire - training item

After a couple of practice items (the results of which did not enter the analysis) participants were presented with 6 groups of 8 test tokens each. Both the

10 Experiment design

order of the groups and the order of the items within each group was randomised. For every sentence, subjects were asked to pay special attention to the (part of the) word that was underlined. The sound files that were played had been created using another Praat script written by the author which pasted the stimulus sentence together with the resynthesised keywords and added bits of silence in between (Figure 10.4 visualises the structure).

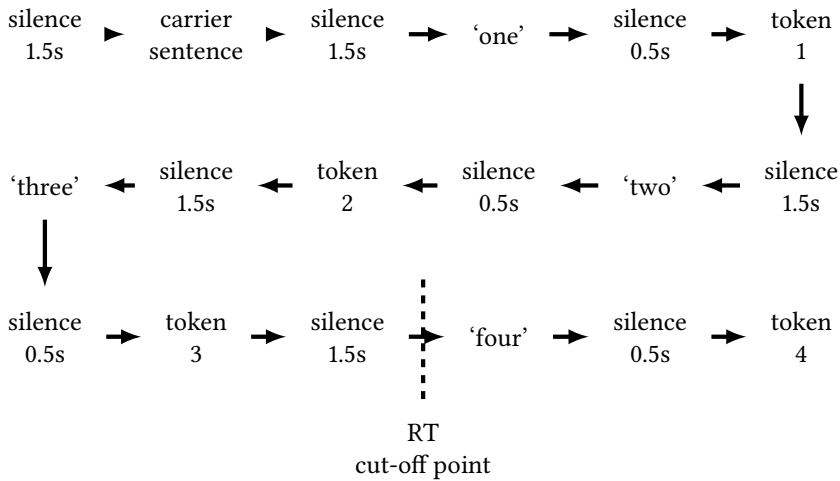


Figure 10.4: Timing of perception stimuli

Each sentence was presented on a separate page of the questionnaire. Participants would see the sentence first for 1.5 seconds before the playback started automatically. After the recording of the sentence had finished playing, participants heard the four resynthesised words (introduced by “one”, “two”, “three”, “four” to avoid confusion²) with 1.5 seconds of silence in between words, and were asked to choose the one that they thought corresponded most closely to the sound in the test sentence. Participants could either choose a sound by clicking on a button or by pressing “1”, “2”, “3”, or “4” on their keyboard. Both the sentence and the resynthesised sounds were only played once. As soon as the subject had clicked on the button of their choice (or pressed the relevant key) the next sentence was automatically presented.

²The token numbers were recorded by a different female speaker (aged 26). This was not ideal as it is possible that the accent of the other speaker might have influenced participants. Incidentally, however, Hay & Drager were faced with the same problem and correctly point out that any effect the second speaker *might* have had would manifest itself in identical fashion in *both* experimental conditions and should therefore not be able to confound the priming effect (cf. Hay & Drager 2010: 871 and 889).

10.2.2 Reaction times

Reaction times were also recorded. These have to be taken with a grain of salt in the context of an online survey, as a number of external factors (skill in using a mouse etc.) add a lot of variance. The potentially most important factor – speed of internet connection – can be ruled out as a confounding variable, however. This is because the platform uses JavaScript to send each subset of stimuli to the computer of the subject. Playback of the sound files does not start before all bits of the question have been downloaded. RT measurements (with an accuracy of about 10 milliseconds) are then taken locally on the participant’s computer before being bundled and sent back to the server. No loading from the server takes place in between stimuli (cf. *SoSciSurvey.de* 2015). While there is still some room for variation between different hardware configurations in terms of timing accuracy and the like, there is no reason to assume that one subgroup of participants in particular would be affected in a significantly different way, so overall such unwanted effects can be hoped to cancel out. Furthermore, reaction times were not analysed as a dependent variable in its own right, but ‘merely’ served to filter responses in the way described below.

The platform *SoSciSurvey.de* does come with some technical limitations, however, since it is not primarily designed for tests requiring RT measurements. Unfortunately, it is neither possible to define a time window outside of which participants cannot enter a response nor to specify when the reaction time clock starts. Technically speaking, RT measurements are really measurements of how long a subject spent on a particular page of the questionnaire. Measuring thus starts automatically once the page is loaded and stops when the next page is accessed (which, in this study, happened automatically once an answer had been selected). “Real” reaction times are arrived at by subtracting the total duration of each audio file from the time spent on the respective page.

As a consequence of these restrictions, it was possible for subjects to make their choice at any time, including *before* the recording had actually finished even though the instructions spelled out that participants should listen to all answer options first. It would have been quite easy to filter out all of these premature answers by simply eliminating all observations with negative RTs from the dataset. This course of action was not taken for two reasons. The first one is comparability with previous research. Neither *Niedzielski 1999* nor *Hay, Warren, et al. 2006*; *Hay & Drager 2010* even recorded reaction times and responses were given on a physical answer sheet so it is quite possible that these studies were, in part, based on answers that had been given before subjects had listened to all of their options. The second, more important reason, is that it does not necessarily

make sense to exclude an answer just because the subject did not listen to all resynthesised sounds first. After all, this study is interested in finding out how (stereotypical) expectations influence people's perception. The claim that at least some expectations and attitudes will prime people to perceive particular sounds *implies* that at least to a certain extent the choice is already made before subjects can really process the physical signal and this might well show in negative RTs. Many people might simply be reluctant to wait for the end of the recording if they already 'know' the answer or if they have already heard the option that they think is the best match. Ignoring negative RTs might then result in involuntarily eliminating (parts of) the priming effect one is interested in from the dataset – which might be a reason why previous studies have not even bothered with reaction times in the first place.

On the other hand, it is fortunate if one is able to cleanse the dataset of nonsense answers which are particularly likely to occur in an online test where the physical presence of the researcher cannot act as an incentive to take the task at hand seriously. Responses with RTs that indicate the subject did not even listen to the carrier sentence, let alone the answer options, are clearly nothing but noise and should be eliminated from the sample. As a sort of compromise it was decided to keep all responses that were not given more than 2000 milliseconds before the end of the stimulus. This threshold ensures that the participant has listened to at least three of the four answer options (cf. Figure 10.4). Responses that were given more than 4000 milliseconds after the end of the recording were likewise eliminated.

10.3 Participants

Finding participants for the online test proved rather difficult. Subjects were recruited through a number of channels. A call for participants was distributed through Hope, Liverpool, and Manchester University. Announcements posted in Liverpool and Manchester related groups on Facebook resulted in some (but very few) responses. Some exchange students were recruited with the help of public notice boards on the campus of Freiburg University. Several friends and colleagues were kind enough to spread the word via e-mail and social media, and this friend-of-a-friend approach proved to be comparatively fruitful. Finally, personal contacts in Liverpool and Manchester and some of the people who had been interviewed about a year earlier were contacted and asked if they would like to participate. All participants were required to

1. be British
2. be native speakers of English
3. have normal hearing

The necessity of the last requirement should be obvious, as subjects were supposed to choose between audio stimuli that differed in rather subtle ways. Requirements 1 and 2 were set to make sure participants were at least likely to have some degree of experience with or knowledge of the accents of Liverpool and Manchester. After all, exemplar priming can only work if there are exemplars that can be activated by the prime. Given what [Montgomery \(2007a\)](#) found with respect to the status of Scouse in particular it seems not too far fetched to assume that people with British nationality (and native competence of English) have at least some exemplars indexed for ‘Liverpool’, even if these are only derived from (stereotypical) media performances. People from, say, the U.S. or Australia, on the other hand, will probably be a lot less familiar with the British accent landscape and might not have the slightest idea of what a Scouse accent sounds like – either because they have never listened to someone from Liverpool, or because they have not done so *knowingly*, meaning that the relevant exemplars will be indexed for a more general category such as ‘British’. In both cases, priming subjects for ‘Liverpool’ (or ‘Manchester’, for that matter) would not be possible.

At the end of the questionnaire, participants had to indicate their age, gender, educational level, profession (profession of parents for students), geographical origin and current town/city of residence (both via the first half of UK postcodes), and whether they self-identified as working or middle class. The occupation scale was the simplified version of the National Statistics Socio-economic Classification, which is used, for instance, by the [Office for National Statistics](#) and classifies jobs into lower, intermediate, and higher, primarily based on how much routine (at the lower end) and responsibility over others (at the upper end) is involved. Just as for the interview data, levels of occupation and education were then used to classify subjects as belonging to one of the two broad categories ‘working class’ or ‘middle class’. Since this classification correlated strongly with the social class participants had explicitly chosen anyway, self-reported social class membership was used as a predictor for statistical modelling (see below).

On the basis of the postcodes participants provided, geographic coordinates were obtained from <http://xposition.co.uk/geopostcode/>, and a euclidean distance value was calculated for every subject using the following formula:

$$d = \sqrt{(\text{lon}_L - \text{lon}_s)^2 + (\text{lat}_L - \text{lat}_s)^2} \quad (10.2)$$

Where lon_L and lat_L are the longitude and latitude of central L1 (Liverpool city centre), lon_s and lat_s the (central) coordinates of the subject's postcode, and d is the resulting distance value (the higher, the further the subject lives from Liverpool). These figures are not, strictly speaking, directly comparable. This is because 1 degree of longitude translates to a different absolute distance in kilometres or miles depending on the latitude. Since the north-south extent of the UK is comparatively small, however, the amount of distortion introduced was considered negligible. Figure 10.5³ shows the geographical distribution of participants by representing every individual with a grey dot. Two clusters are clearly visible, one in the north-west and one in the south-east of England. The first is due to the fact that recruitment initially focussed on Liverpool and Manchester, while the London bias is to a degree even representative, given that between 15 and 20% (depending on where one draws the boundaries) of the UK population live in the metropolitan area. The remaining participants seem to come more or less from all over the country, although Wales and Northern Ireland are clearly under-represented.

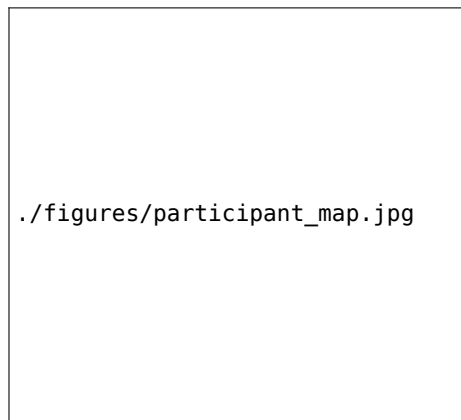


Figure 10.5: Geographical distribution of subjects (perception)

Subjects were given the opportunity to receive the results of the experiment after its completion and to participate in a lottery for a £100 gift card from a big online retailer. E-mail-addresses were collected for these purposes, but kept strictly separate from the questionnaire responses (using an in-built function of the online platform specifically designed for that purpose).

³Map tiles by Stamen Design, under CC BY 3.0. Data by OpenStreetMap, under CC BY SA.

In total, 67 subjects participated in the experiment and provided 2508 data points. In addition to the more detailed placement via the postcode, participants were also assigned to one of the broad geographical categories “internal” (Liverpool and Merseyside, “L” postcode area; 9 subjects) and “external” (everything else, 58 subjects). It is possible that some of subjects in the Liverpool sample had already participated in one of the sociolinguistic interviews conducted by the author, but since these interviews had taken place (at least) a year earlier it is highly unlikely that this could have a distorting effect on results in the perception test. Table 10.4 provides a more detailed overview of the ‘external’ sub-sample (3 subjects in this group declined to indicate their gender). One thing that is immediately obvious is that the middle class is heavily over-represented, only 7 subjects were classified as belonging to the working class. Among middle class subjects, female and male participants are not really evenly distributed across priming conditions: women dominate in the group that was primed for Liverpool, while men do in the other. However, Prime and Gender are still not collinear in the dataset ($\kappa = 1.54$), and mixed-effects ordinal models regressing reported percept on Gender did not find a significant effect of the latter, neither in the ‘Liverpool’ ($p = 0.801$) nor in the ‘Manchester’ condition ($p = 0.597$). In other words, women and men did not behave differently in this experiment so the fact that they are unevenly distributed across conditions is unproblematic. Incidentally, Niedzielski (1999: 69 and 79–80) also found that “there was essentially no difference between what male and female respondents selected”.

Table 10.4: Gender and social class distribution of subjects (perception, external)

prime	‘Liverpool’		‘Manchester’	
	F	M	F	M
working class	2	3	1	1
middle class	17	6	9	16

10.4 Statistical analysis

Unlike in the case of the interview data where there were fewer participants and the role of the individual subject was different (cf. §5.6), it actually does make sense to include random effects for the statistical analysis of the perception data.

However, the data collected by the experiment described above are not ratio or interval scaled. While answers are expressed in numbers (token 1, 2, 3, or 4), these are really labels for 4 distinct categories rather than measurements on a continuous scale. There is no guarantee, for instance, that the (perceptual) distance from token 1 to token 2 is identical to that between tokens 2 and 3 (although this *was* the aim when creating the tokens; cf. §10.1.2 and §10.1.3) – which is what the values suggest when they are being treated as numerical measurements. On the other hand, the data *are* ordered in a way (from most standard/Mancunian to most Liverpool), so while choosing token 4 instead of token 3 might not be the same as choosing 3 instead of 2 in terms of perceptual distance, higher values do in all cases represent a more Liverpool-like percept than lower values – much in the same way that, say, a higher RT always mean that a subject took more time to respond than another one with a lower RT. It was therefore decided not to follow Hay, Nolan, et al. 2006; Hay & Drager 2010 in treating the answers as numerical data for the purposes of statistical modelling. Instead, the R package ‘ordinal’ (Christensen 2015) was used to calculate cumulative link mixed ordered regression models via the Laplace approximation.

Subject was entered as a random intercept (cf. §5.6), and a random by-subject slope for stimulus order was also included to counter any individual training or fatigue effects. Carrier word was not included as a random effect because frequency and, for consonants, phonological environment were of interest as fixed effects. Since there were only 6 keywords per variable filtering out any lexical effects would almost certainly also have eliminated effects of frequency and/or phonological environment. Prime was entered as a fixed factor, along with gender, social class (self-reported), age of the subject, and geographical distance from Liverpool. Furthermore, the Zipf score of the carrier word was included, as was the position of the carrier word in the stimulus sentence (sentence-medially or -finally), the position of the stimulus sentence within the experiment (when was the sentence played to the subject), and, for consonants, the phonological environment (V_V or _#). All two-way interactions of the prime with these linguistic and extra-linguistic factors were also considered. Sum coding was used for all models in order to be able to identify main effects and interactions. Model selection was based on AIC scores and F-tests comparing nested models. Collinearity was investigated applying the functions written by Austin Frank (cf. §5.6) to corresponding mixed-effects logistic regression models.

Sound change, priming, salience

This volume investigates the realisation and perception of four phonological variables in Liverpool English (Scouse), with a special focus on their sociolinguistic salience. Younger speakers' speech is found to be more local, but only for the two salient variables in the sample (NURSE-SQUARE and /k/ lenition), which appear to carry considerable amounts of covert prestige. Local variants of non-salient happy-tensing and velar nasal plus, on the other hand, are actually found to be receding, so at least to a certain extent Scouse also seems to be participating in regional dialect levelling.

The importance of salience is also obvious in the perception data, with only the two highly salient stereotypes generating robust effects in a social priming experiment (albeit in the unexpected direction). These results indicate that the investigated variables differ measurably not only in their use in production, but also in terms of how central they are to mental sociolinguistic representations of Scouse. They also tell us more about the way we process, store, and (re-)use sociolinguistic variation in perception. By defining likely contexts for significant priming effects they might finally even help in coming up with a more elaborate 'theory of priming' in the realm of sociophonetics.

ISBN 978-3-96110-119-1



9 783961 101191