Effect of distributional shape on learning a target[*]

Emily Sadlier-Brown
University of British Columbia

Abstract: Labov (2001) showed that vowels that are being pronounced the same from one generation to the next exhibit normal (i.e. bell curve) frequency distributions in their pronunciations, while vowels that are changing diachronically exhibit skewed (i.e. lopsided) frequency distributions. The present paper reports on a production experiment that examines whether the continuation of long-term patterns of vowel behaviour (change over time or stability over time) is partly due to differential learning of the frequency distributions associated with each long-term pattern. The experiment specifically asks whether learners exposed to a skewed distribution of pitches, as opposed to those exposed to a normal distribution of pitches, will learn a target pitch that is different than the mean pitch of the distribution they heard. In weak support of this hypothesis, results indicate that participants exposed to skewed input distributions produced pitches that were further from the input mean pitch than those exposed to normal input distributions, but the difference between groups was not significant. In addition, both groups produced output pitches that tended to be higher than the mean pitch of the input, and both groups' output was strongly influenced by the final note they had heard.

Keywords: sound change, vowel shift, distributional learning, skewness, non-veridical learning

1 Introduction

In any given language, at any given time, some vowels may be changing while others are maintaining stability across generations. In a diachronic vowel shift, a vowel "moves" in the vowel space,[1] subtly changing in pronunciation over time by advancing a little more in the same direction each generation. How this happens, and why it happens with some vowels and not others, is not well-understood. In this paper, I present the results of an experiment that examines vowel shift from the perspective of learning, asking whether different distributional shapes of input have different effects on what learners decide is their production target. In particular, I examine whether learners who hear a more skewed distribution tend to move the mean in their own productions.

[*] I thank Carla Hudson Kam, Michael Rochemont and Joseph Stemberger for their valued comments. Any errors are my own. I also thank Robert Walker for support with the Bounce Metronome software, Michael McAuliffe for help coding in R and Praat, Alexis Black for analysis advice, the Language and Learning Lab for several rounds of input and my three volunteers, who together ran the bulk of the participants, Kelly Chan, Sacha Fonarev and Kaitlyn Martinson.
Contact info: emily.sadlier-brown@ubc.ca

[1] A vowel's location in the vowel space is usually considered a combination of its height (measured by the frequency [in Hz] of its first formant [F1], plotted on the x axis) and backness (measured by the frequency of its second formant [F2], plotted on the y axis). A vowel's mean location refers to its average location in this two dimensional space.

Diachronic vowel shifts are ongoing in many, if not all, varieties of English as well as many other languages (e.g. Montreal French [Yaegor-Dror 1994], Pokcha Russian [Kochetov 2006], Seoul Korean [Kang 2016]). An example of a currently active diachronic vowel shift is the Canadian Shift, which involves the ongoing backing of /æ/ and subsequent backing and/or lowering of /ɛ/ in Canadian English (Clarke, Elms and Youssef 1995). In the Great Lakes area of the United States, the Northern Cities Shift (Labov, Yeager and Steiner 1972) has the opposite effect: /æ/ is raising and tensing in the vowel space, leaving room for the fronting of /ɑ/. Over time, such shifts gradually transform a vowel system and can even come to define a dialect (Labov, Ash and Boberg 2006). Perhaps the most famous example of diachronic vowel shift is the Great English Vowel Shift, which, over approximately 250 years (1350–1600), completely altered the pronunciation of English (e.g. Jesperson 1949). It is these gradual changes in vowel pronunciation over time—attested both currently and historically—that are the focus of the present research question.

A hallmark of ongoing vowel shift is that, for a given vowel involved in a shift (e.g. /æ/ in the Canadian Shift), younger generations of speakers produce successively more advanced mean vowel pronunciations (advanced in terms of the direction of the change).[2] In sound change, of which vowel shift is an example, the speakers with the most advanced means are often teenagers and young adults; older speakers tend to have less advanced means (Labov 2001:110–113, 169–170, Cedergren 1988, Tagliamonte and D'Arcy 2009). In this type of change, after teenagehood/young adulthood, individuals' means remain relatively stable (Cedergren 1988, Labov 1994:105, Cukor-Avila 2000, Bailey 2002, Sankoff 2006, Tagliamonte and D'Arcy 2009). Combined, these facts entail that teenagers/young adults are outputting a mean that is different—that is, more advanced—than the one they heard in their input from older speakers. Teenagers/young adults, have "overshot" the mean they heard, moving it another step in the direction of the change in progress. Here, *overshoot* is defined as 'a difference between a speaker's output and input' (note that the term is not meant to imply that speakers have knowledge of the direction of change).

The fact that adolescents are producing a different vowel mean than the one they heard suggests that a factor in the continuation of vowel shifts may be non-veridical learning, in which what is learned is not the same as the input (Hudson Kam and Newport 2005). Non-veridical learning has been discussed in many linguistic and non-linguistic domains (Singleton and Newport 2004, Hudson Kam and Newport 2005, 2009, Ferdinand, Thompson, Kirby and Smith 2013). Hudson Kam and Newport (2009), in an experiment on morphosyntactic learning, showed that a non-veridical outcome (in the form of overuse of the most common input form compared to less common input forms) depended on the presence of variability in learners' input, and crucially on the *type* of variability. Not all variability led to non-veridical learning; a combination of unpredictability and low frequency of the less common input forms was required. This suggests that the overshoot characteristic of teenagers and young adults' vowels may be related to the variability characteristic of vowels—and in particular to the type of variability specific to *changing* vowels.

Vowels vary slightly from one pronunciation to the next: over the course of several pronunciations of the word *cat*, for example, formant measurements of the vowel /æ/ are likely to reveal slight, though usually unnoticeable, differences in pronunciation. Labov (2001, Ch. 15)

[2] Such a pattern can also apply to forms that are age-graded, that is, forms that change in use over the course of a speaker's lifetime. This type of change is not the focus of the present research question.

describes a difference in the type of variability characteristic of a changing vowel's pronunciations compared to the type of variability characteristic of a stable vowel's pronunciations. Specifically, he shows that stable and changing vowels are characterized by different distributional shapes. A stable vowel's formants exhibit *normal* (bell curve) frequency distributions, with the mean formant frequency (for a given formant) being also the most common formant frequency (for that formant). In a normal distribution, the mean and the mode are the same. In contrast, a changing vowel's formants (F1 and/or F2) exhibit a *skewed* frequency distribution (for the formant/formants that is/are changing) (Labov 2001:480–2). Skewness is a measure of the degree of asymmetry of a distribution: positive skew refers to distributions with longer tails to the right, and negative skew refers to distributions with longer tails to the left. In a skewed distribution, the most frequent value is *not* the mean value; rather, in the unimodal distributions relevant here, the mean is offset from the mode in the direction of the tail of the distribution. Labov further notes a correlation between the degree of skew in a vowel's distribution and how advanced it is in an ongoing change. At the beginning of a change, a vowel's distribution is positively skewed, and as the change advances and comes to completion, the distribution becomes increasingly negatively skewed before a normal distribution is restored (Labov 2001:488). He goes on to suggest that speakers do not aim for the mean in a skewed distribution, but aim for a mean which is shifted in the direction of the change; that is, speakers overshoot (Labov 2001:493).

The hypothesis I take from Labov's observation is: *when given normally distributed continuous input (mimicking a stable vowel), learners will reproduce the mean, but when given skewed input (mimicking a changing vowel), learners will behave non-veridically, overshooting the input mean.* The claim implicit in the hypothesis is that acquisition and change occur simultaneously, and in some sense may be the same thing. To fully understand why some vowels are changing and some are not, we must understand the forces at work in driving change as well as preventing it. For any given explanation of sound change, we must explain why this does not apply to vowels observed to be stable. For any given explanation of stability, we must explain why this does not apply to vowels observed to be changing. The present hypothesis places the difference between change and stability on the distributional shape of a given vowel, advancing the idea that these different diachronic patterns may partly be explained by how learners react to differing vowel-internal distributions.

In order for such a hypothesis to be plausible, it is necessary that listeners be attuned to fine-grained distributional details within a continuous phonetic distribution. In fact, there is ample evidence that this is the case. Maye (2000) and Maye and Gerkin (2000) showed that adults were more likely to distinguish two categories along a [d]–[t] continuum when they had been exposed to a bimodal distribution along the continuum as opposed to a unimodal one. Other studies have shown that sub-categorical phonetic details affect category goodness ratings (Miller and Volaitis 1989, Miller 1997), lexical access (Andruski, Blumstein and Burton 1994, McMurray, Tanenhaus and Aslin 2002), lexical categorization (Clayards, Tanenhaus, Aslin and Jacobs 2008) and reaction times in phoneme identification and discrimination tasks (Pisoni and Tash 1974). It is therefore quite likely that speakers are sensitive to the distributional differences that Labov describes as characterizing stable vs. changing vowels, and that this sensitivity could result in a difference in what speakers perceive their target to be.

Labov's observation relates skewness to overshoot but is limited in its ability to predict the *direction* of overshoot. Although the bulk of his discussion is focused on positive skew with overshoot in the direction of the tail (as in the beginning of a change), it is also clear that the negative skew characteristic of the end of a change would seem to require a favouring of the mode in order to eventually restore a normal distribution. The small amount of empirical evidence bearing on this question likewise supplies contradictory predictions as to whether learners favour the tail or the centre of a skewed distribution. Olejarczuk and Kapatsinski (2016), examining how learners rated prosodic contours drawing from either a positively or negatively skewed distribution, found that participants' category goodness ratings favoured the distributional tails. On the other hand, van der Ham and de Boer (2015), in an experiment in which participants were required to reproduce a continuous distribution of pitches, found their participants favoured the centre of the distribution. Despite their different outcomes, both of these studies found a relationship between distributional shape and overshoot. However, the different outcomes preclude making a prediction for which direction the learners in the present experiment will favour (distribution tail or centre).

The present research question pertains to vowel formant change, but the stimuli used in the present experiment were non-linguistic pure tones varying in pitch. Tones were used in order to avoid the confounding influence of speakers' pre-existing vowel targets: for most adults, pitch is not identified absolutely in the same way that vowels are (Takeuchi and Hulse 1993). In other ways, however, vowels and pitches share similarities that suggest they might be affected by the same processing mechanisms. For example, vowels and pitches both form categories, and these categories are comprised of internally variable information. (Just like saying the word *cat* will vary slightly from production to production, singing middle C will also vary slightly from production to production.) The type of information being tracked is also similar—fundamental frequency (F0) for pitch, and formant frequency (often F1 and F2), among other things, for vowels. Furthermore, the work of Saffran, Johnson, Aslin and Newport (1999), examining the tracking of transitional probabilities of pitch sequences, found the same results for pitch as in previous experiments with syllables. They suggest that the same mechanism underlies statistical learning in both domains. For current purposes, the use of fine-grained pitch differences as a proxy for the fine-grained vowel formants commonly used to study vowel change appears to be justified (see Trehub and Trainor [1993] for further discussion).

In this study, I exposed learners to input consisting of categories that were either skewed (which mimics a changing vowel) or normal (the control condition, which mimics a stable vowel) in their internal distribution. People heard a sample of attempts at a specific target production drawn from the distribution and then produced a single response indicating what they thought the target was. The predictions were that participants in the skewed distribution condition would output a different mean pitch than the mean pitch of the input they heard, and participants in the normal distribution condition would output the same mean pitch as the mean pitch of the input they heard. This project thus takes an existing observation (the differing distributions characteristic of stable and changing vowels) and attempts to discover whether the relationship between distribution and diachronic patterning is a causal one. The project bears on the continuation of change, but does not bear directly on how a change starts (actuation) or stops (completion). In sum, the focus here is on the relevant learning processes potentially involved in the continuation of diachronic processes over time.

2 Methodology

2.1 Participants

Study participants were 73 adults (16 males) attending the University of British Columbia. Due to strict exclusion criteria, 32 participants' data were eventually excluded. 14 participants were eliminated because they produced more than 5% (n=2) too many or too few responses. Eight additional participants were eliminated for not listening to all the trial tones before producing their response. One participant was eliminated due to experimenter error in reading the instructions, and nine further participants were eliminated because the average of the errors of their 16 copied tones[3] (in Phases 1 and 3 of the experiment, see Section 2.3) was more than 8.5 Hz. This is a quarter tone in the range of the experimental continua, or half a semitone, which is considered the threshold for tone deafness.[4,5] After exclusions, 41 adults (eight males) with a mean age of 22 (range 18–41) were included in the analysis.

2.2 Stimuli

The stimuli consisted of samples from four 6-tone pitch continua: two continua (one high and one low) sampled from a normal distribution and two continua (one high and one low) sampled from a positively skewed distribution. Distributions were created in R (R Core Team 2016). The normal distribution was set to have a mean of 298 Hz (D4 + 25 cents[6], range 288–308 Hz, or about 1.16 semitones) in the low range and 337 Hz (E4 + 38 cents, range 327–347 Hz, or 1.03 semitones) in the high range. The skewed distribution was set to have a mean of 294 Hz (D4 + 2 cents, range 288–308 Hz) in the low range and 333 Hz (E4 + 18 cents, range 327–347 Hz) in the high range. (Note that the normal and skewed distributions have identical pitch ranges.) Thus, the high and low continua are actually spaced quite closely together, their centres being only about two semitones apart. Since pitch is perceived logarithmically, two intervals identical in perceptual distance will be different when measured in Hz. However, the low and high continua are located close enough together in pitch space to ensure that this difference is almost negligible: the high and low continua have identical pitch ranges (in Hz), while in perceptual terms, the high continuum is only 0.13 semitones wider than the low continuum.

Among the variety of skewed distributions described by Labov (2001:481–3), the most common shape was unimodal and positive (meaning the mean is to the right of the mode) and characterized by a skewness of about 0.5. This is therefore the shape that was chosen for the skewed distribution in the current experiment. Kurtosis, which is a measure of the heaviness of a distribution's tails, was set to 0 for all distributions. Tones occurred in 4 Hz steps. This distance was chosen for two reasons. First, it is slightly more than 3 Hz, which is the just noticeable difference in this pitch range (Kollmeier, Brand and Meyer 2008). This should help to ensure that differences between neighbouring tones are perceptible. (That is, the trials are being perceived as distributions and not as repeated tones.) Second, tones spaced at intervals wider than this would widen the pitch range beyond that which would be typical of a single pitch category. In this experiment, it is crucial that participants perceive the tones as being members of a single category.

[3] One "mistake", defined as an error of more than 17 Hz, was permitted, per participant, and not included in the average.
[4] Psyche Loui (2016), p.c.
[5] As the number of participants eliminated for this reason seems quite high, it is likely that participants excluded on the grounds of copying inaccuracy also included those who had difficulty *producing* the tone they intended (as opposed to perceiving it) as well as those who were inaccurate for other reasons, e.g. inattention.
[6] Cents are a perceptual measure of pitch. 1 semitone = 100 cents.

To create a sample sequence, 20 pitch frequencies were randomly sampled from a given distribution (that is, normal or skewed). Each sample sequence was to become an individual trial. Skewness limits on the sample sequences were imposed: normal sample sequences were discarded if their skewness was >0.25 or <-0.25; skewed sample sequences were discarded if their skewness was <0.35 or if the difference between the mean and mode of that set was <3 Hz. 20 sample sequences (that is, 20 sequences consisting of 20 tones each) were created from the low normal distribution, yielding 400 tones total. 20 sample sequences were then created from the high normal distribution (=another 400 tones total), and so on for the high and low skewed distributions. This resulted in four sets of sample sequences (normal high, normal low, skewed high, skewed low). For the purpose of post-hoc tests to confirm the sample sets had the desired properties, all normal samples were combined and all skewed samples were combined (low samples were transposed into the high range for the purpose of these post-hoc tests). Using t-tests, the complete set of normal samples was then compared to the complete set of skewed samples to confirm that the two sets' degree of skewness was significantly different and degree of kurtosis was not significantly different. T-tests were performed to confirm that in the normal sample sequences, skewness and kurtosis was not significantly different than 0, and that in the skewed trial sets, kurtosis was not significantly different than 0. T-tests were also used to check that skewness was not significantly different in the low and high continua (both normal and skewed). Figure 1a is a histogram of the entire set of 800 input pitch frequencies (black bars) sampled from the high normal and low normal distributions combined. (Low continuum pitches were transposed into the high range for display purposes.) Figure 1b shows the same for the skewed distributions. Note that the probability of selecting the highest tone in the skewed distribution was so low that this tone did not occur in any of the sample sets. (However, the highest tone in the normal condition was rarely sampled: it occurs only 12 times in the 800 samples, a frequency of 1.5%.) To create the sample tones, each sample set was synthesized as a sequence of pure tones in Praat (Boersma and Weenink 2014). Tones were 0.75 seconds in length, with 0.55 seconds between them.

2.3 Procedure

Participants were seated in front of two computer screens, one directly in front of them and one slightly to their right. The experiment contained three phases. In the first and third phases, participants listened to a single tone and were instructed to copy it. To output tones, participants used a mouse to control a digital theremin on the right-hand screen (Bounce Metronome program [Walker 2016]). A theremin is an instrument that can play a continuous range of pitches. The theremin covered a range of nine semitones centred on 340 Hz (approximately F4). Equal movements of the mouse resulted in equal distances in perceptual space. The theremin screen was completely blank. Once a participant had selected an output tone, they released the mouse button and pressed "R", which logged the pitch frequency of the tone (in Hz) and the time it was played. This exercise allowed for multiple attempts at output pitch selection while avoiding the recording of unintended responses. In the second phase of the experiment, participants were told they would hear a series of tones that had been produced by 20 amateur musicians, all of whom were attempting to play the same note. They were told the notes were slightly variable, and that their task was to listen to all of the attempts and then "play the note [they thought] the musicians were asked to play". The experiment consisted of 40 trials, alternating between the low pitch continuum and the high pitch continuum of the relevant condition. Distractor trials, consisting of totally different tone sequences, were included every few trials to help keep participants' attention. On each trial, the participant's selected tone was recorded in the same manner as for Phases 1 and 3. The resulting raw values were later used to compute a difference score for each trial, comprising the participant's response minus the mean pitch of that trial.

3 Results

3.1 Basic results

Figure 1 shows histograms of all the raw output tone frequencies overlaid over all the input frequencies in the normal (1a) and skewed (1b) conditions. Output pitches covered a wider range than the input, and a tendency towards outputting higher pitches is also apparent. Figure 2 shows histograms of the difference scores in each condition. Difference scores are the variable of interest; a participant's difference score on a given trial is the amount they overshot the input mean of that trial. A difference score of 10, for example, indicates a participant output a tone that was 10 Hz higher than the mean of the tones they heard on that trial. A difference score of 0 indicates a participant's response was the same as the input mean on that trial. The bias towards outputting higher pitches is evident in the overrepresentation of positive difference scores in both the normal and the skewed condition. T-tests revealed that difference scores were significantly higher than 0 in both conditions ($p<.001$ for the normal condition; $p<.0001$ for the skewed condition). Therefore, under the current definition, participants overshot the mean in both conditions, consistently outputting higher pitches than the mean of the pitches they had heard. A post-hoc examination of a random sample of eight participants (20% of the total number of participants) revealed the pitch bias was also evident in copied tones, the single tones that participants were instructed to copy at the beginning and the end of the experiment ($p<.05$).

(a) (b)

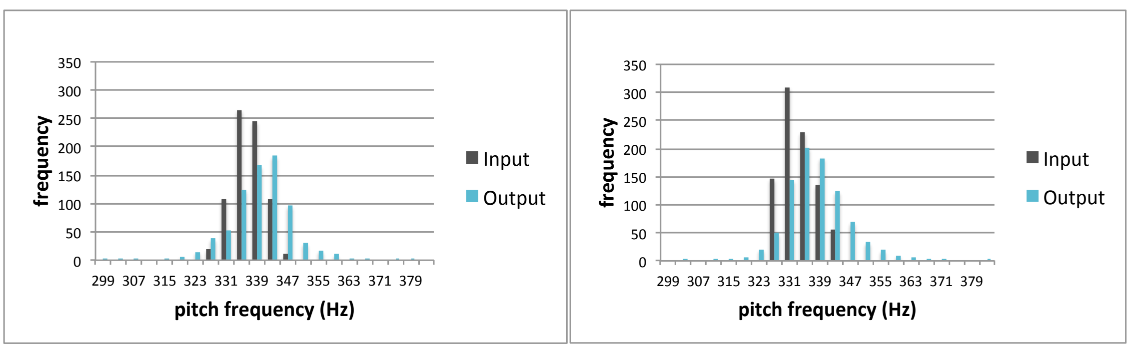

Figure 1: Raw input (black) and output (blue) tones, normal condition (a) and skewed condition (b).

The effect of distribution shape on difference scores was measured with a linear mixed effects model using the lme4 package in R (Bates, Maechler, Bolker and Walker 2015, R Core Team 2016). Fixed effects included: *condition*, *final-mean* (a measure of the difference between a trial's final note and its mean), *age*, *gender*, *trial* and *pitch region* (which refers to the low pitch continuum vs. the high pitch continuum). Two additional fixed effects were included to mitigate possible experimenter-induced variability. These were *instruction type* (some participants received a slight variation on the instructions wording) and +/- *responses* (referring to data in which a participant either forgot to record up to two responses or recorded up to two extra responses, which were adjusted by the experimenter to account for the missing or extra line(s) of data).[7] By-participant random slopes were included for the effect of trial. A second model included the fixed effect *mode-mean* (a measure of skewness obtained by subtracting input mean from input mode) in place of condition. This model was only used to report on the effect of mode-mean. Mode-mean and condition were not included in the same model due to the collinearity of these two alternate measures of skewness. P-values were obtained using the likelihood ratio test described by Winter (2013). This test uses an ANOVA to compare two models, one complete and one excluding the factor of interest.

(a) (b)

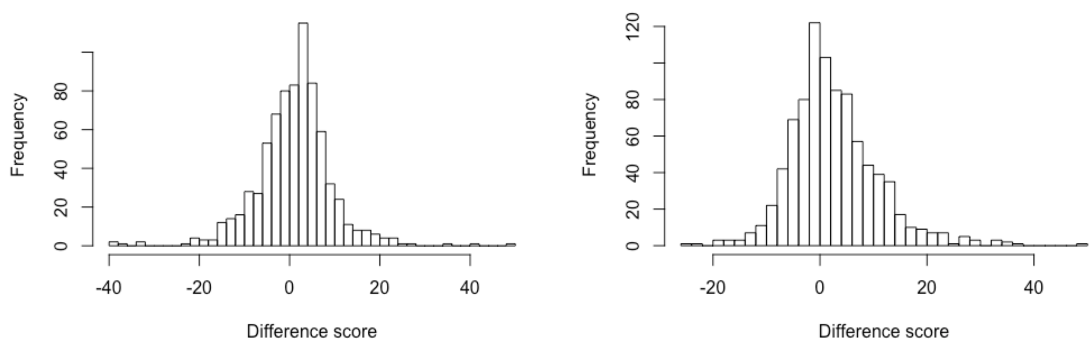

Figure 2: Difference scores (difference between a participant's response on a trial and the mean of the input for that trial), all trials, normal condition (a) and skewed condition (b).

The model yielded two significant predictors of overshoot: *final-mean* (χ^2 (1) = 12.26, *p*<.001, B = 0.18 ± 0.05 [standard errors]) and *pitch region* (χ^2 (1) = 7.63, *p*<.01, B = 1.07 ± 0.39 [standard errors]). Clear tendencies emerged for *condition* and *mode-mean*; however, neither condition (*p*=.09) nor mode-mean (*p*=.08) was significant. All other factors were non-significant: *age*, *gender*, *trial*, as well as the two factors included to mitigate the possibility of unintentional variability introduced in the experimental procedure and analysis, *instruction type* and *+/- responses*. There were no significant interactions. Section 3.2 presents the results for the main factors of interest, condition and mode-mean. Section 3.3 presents the results for final note and Section 3.4 presents the results for pitch region.

3.2 Condition and mode-mean

Condition and mode-mean are reported together because they are essentially two different measures of distributional shape. Condition is a discrete variable, with each condition containing 40 trials that vary on their degree of skewness, but whose average skewness is characteristic of the condition. In contrast, mode-mean is a continuous variable capturing the difference between

[7] As mentioned in Section 2.1, if there were more than two missing or two extra responses, the participant was eliminated.

the mode and the mean on each trial, regardless of condition. Comparing conditions, the skewed condition was associated with a non-significant tendency towards more overshoot, which was more pronounced in the low pitch region (see Section 3.4). In the skewed condition, overshoot was 2.6 Hz, whereas in the normal condition, overshoot was 1.0 Hz.

The tendency exhibited by the difference in condition is equally clear when looking at the continuous factor mode-mean. Recall that the current experiment contains normally distributed and *positively* skewed (or right-skewed) distributions. Negative skews are only present as part of the natural variability inherent in the normal distribution samples. A positively skewed unimodal distribution has a mean to the right of the mode (that is, the mean pitch is higher than the mode pitch); therefore, in the current model, more *negative* values of mode-mean indicate a more *positively* skewed distribution. Mode-mean was associated with a tendency towards more overshoot; specifically, lower values of mode-mean (=higher degrees of positive skew) were associated with a tendency to increase overshoot (Figures 3a and 3b). The amount of overshoot decreased as distributions approached normal, and decreased even more as skews became negative. The overall bias towards higher pitches, however, is always present: normal and even negatively skewed distributions are still associated with a small degree of overshoot.

The negative correlation between mode-mean and overshoot means that positive skew had the effect of drawing output towards the tail of the distribution. (However, there may be a brief favouring of overshoot towards the mode when mode-mean is approximately 1 Hz; in that slightly negatively skewed situation, overshoot is about .75 Hz, which is towards the mode rather than the tail.)

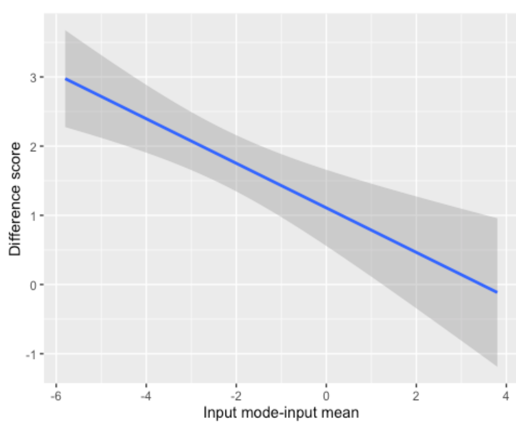

Figure 3a: LOESS-smoothed visualization of the effect of mode-mean on participants' difference scores. The lower the input mode is from the input mean (that is, the more positively skewed the distribution), the greater the difference score (overshoot).

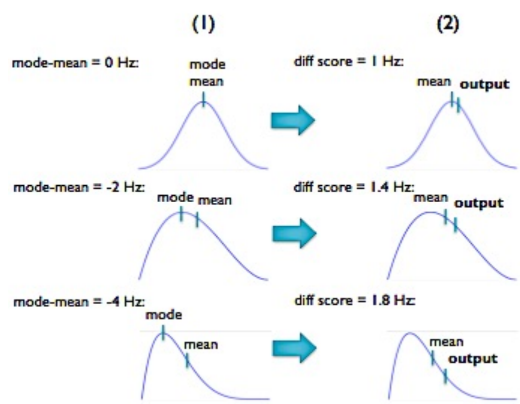

Figure 3b: Visualization of the effect of an example distributional shape (column 1) on overshoot (column 2).

3.3 Final-mean

More positive values of final-mean were associated with increased positive difference scores (Figures 4a and 4b). This is to say that the higher the final note of a trial was compared to the trial's mean pitch, the more participants' output diverged from the mean in the direction of the final note. Participants' outputs were drawn to the final note they heard. Another way to measure the effect of final note on output is to use the raw final note value in place of final-mean. This alternate measure was expected to produce the same result, and indeed this was the case. A model replacing the factor final-mean with raw final note indicated that the final note a participant heard was a significant predictor of a participant's output ($\chi^2(1) = 7.63$, $p<.01$, B = 0.15 ± 0.05 [standard errors]). Due to the pitch bias discussed in Section 3.1, people still overshot the mean regardless of the location of the final note. (That is, even when final note = mean, participants still produced a higher note.) The effect of location of the final note with respect to the mean therefore had the effect of reining in the amount of overshoot due to pitch bias. Without this pitch bias, it seems that final-mean would have had a more direct effect on output, perhaps even reversing overshoot direction when the final note was lower than the mean.

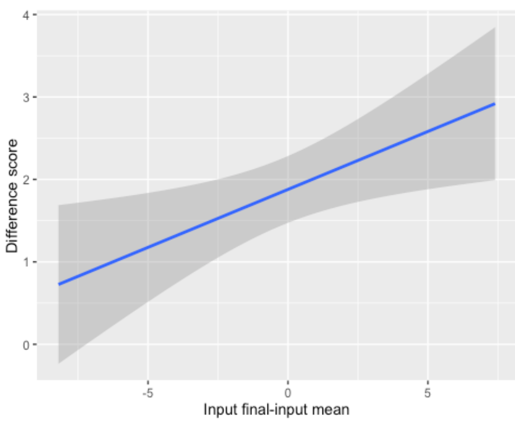

Figure 4a: LOESS-smoothed visualization of the effect of final-mean on participants' difference scores. The higher the input final note is from the input mean, the greater the difference score (overshoot).

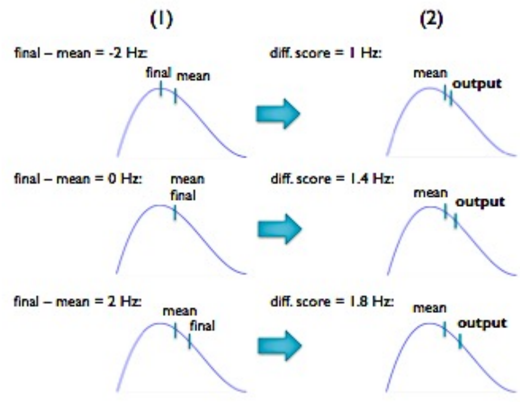

Figure 4b: Visualization of the effect of an example final note (column 1) on overshoot (column 2)

3.4 Pitch region

Participants diverged more from the mean on low pitch trials (where they overshot by 2.3 Hz) than on high pitch trials (where they overshot by 1.4 Hz). Although the test for an interaction between pitch region and condition was not significant (p=.07) the difference nevertheless appears to be limited to the skewed condition, in which the high and low overshoot means were different: 3.4 Hz on the low pitch trials and 1.8 Hz on the high pitch trials. (In the normal condition, overshoot was 1.0 Hz in both pitch regions).

4 Discussion

4.1 Distribution shape

The effect of distribution shape on participants' output was measured in two ways: by the discrete variable *condition*, and by the continuous variable *mode-mean*. Counter to prediction, participants overshot the mean in the normal as well as the skewed condition; however, in weak support of the prediction for the effect of distribution shape, both condition and mode-mean yielded clear though non-significant trends pointing towards an influence of skew on overshoot. People overshot *more* in skewed distributions than in normal distributions. In this experiment then, skewness seems to have caused a "mini sound change", as participants in the skewed condition output a different mean than the one they heard (more so, at least, than those in the normal condition). Beyond being a potential driver of sound change, skew might have an effect on the *rate* of change, with distributions that are initially more skewed resulting in more overshoot and thus potentially a faster rate of change. Besides the implication for sound change, the result suggests that pitch category learning, at least, seems to be affected by the sub-categorical details of an input distribution. As was especially obvious in the skewed condition, participants did not learn the mean nor mode, but a pitch shifted towards the distributional tail. The results are expected to be relevant for other continuous distributions, including vowel distributions. The failure to achieve significance at the $p<.05$ level warrants a cautious interpretation of the results; however, the large amount of variability in this type of data combined with the fact that the trend is in the predicted direction (skewness producing more overshoot) suggests that the results should be approached as indicative of a trend that merits further investigation, rather than dismissed.

Since the current experiment contained positive skews with tails towards the right, overshoot in the direction of the tail meant that people tended to play higher pitches than they heard. Overshoot was therefore amplified by the overall pitch bias towards higher tones, and it seems likely that without this bias, the direction of overshoot might have reversed towards lower pitches in the few negatively skewed trials. (This would still mean overshoot towards the tail, but a tail in the opposite direction.) Even if the pitch bias were to persist in a future experiment, the addition of more negatively skewed distributions would help to strengthen the current finding regarding the direction of overshoot.

The idea that skew is associated with overshoot in learners was due to Labov's observation that changing vowels exhibit a skewed distribution in the community at large, whereas stable ones exhibit a normal distribution (Labov 2001:482). The current results may help to explain this relationship by suggesting that learners continue a change specifically when their input is skewed. The participants in the current experiment acted in accordance with Labov's predictions for the *beginning* of a change (bias towards tail), as opposed to the end of a change, where a bias towards the mean or mode might be expected (which leads to the idea that perhaps there is a different balance of factors at work at the end of a change). Labov suggested the relationship between change and skew has to do with the physical limitations of the vowel space, the perceptual boundaries between vowels, and most importantly, the quality of the tokens that lead in a shift (stressed monosyllables). However, the current results found that learners overshoot in the *absence* of all of these vowel-specific explanatory factors. Instead, the current results point to a potential driver of change that is relevant to all continuous distributions, both linguistic and non-linguistic: distribution shape.

It is promising that the current suggestion for an effect of distributional shape is in line with other experimental results. Two experimental studies that have examined skew have both found an effect of distributional shape, albeit in opposite directions. Van der Ham and De Boer (2015) exposed participants to an /a:/ sound that varied in pitch across a continuous positively skewed, negatively skewed, or uniform (flat) distribution of 100–243 Hz. Participants were then asked to output the entire distribution from memory using up and down arrows on a computer to adjust a reference pitch that was given to them. Results showed a bias towards the centre of the distribution in participants' output—the opposite to what was found in the current experiment. However, several major differences between Van der Ham and de Boer (2015) and the current experiment suggest that the two are not comparable. First, although their variability was continuous, it encompassed a wide pitch range (16 semitones, or more than an octave), which is arguably not representative of a single category. It is possible that this type of variability is perceived differently than the fine-grained sub-categorical distributions in the present experiment. Second, the participants in Van der Ham and De Boer (2015) were instructed to attempt to reproduce the entire distribution from memory, which again is arguably not a realistic proxy for what happens when speakers learn a vowel category. The task, therefore, may have been more of a test of memorizing pitch sequences than a test of generalization behaviour.

The current results are more comparable to those of Olejarczuk and Kapatsinski (2016), who exposed listeners to either a positively or negatively skewed training set of LHL prosodic contours that differed with respect to pitch excursion magnitude. As in the current experiment, participants' goodness ratings indicated a favouring of the distributional tails. The authors interpret the results as being consonant with their *log frequency hypothesis*, under which a listener's representation of category typicality reflects not linear, but logarithmic, tracking of token frequency. The log frequency hypothesis, motivated in part by studies of word recognition (e.g. Kreuz 1987), predicts that an infrequent or surprising token contributes more to the listener's perception of category typicality than a frequent or expected one. The reasoning behind the tracking of log frequency rather than raw frequency is that unsurprising stimuli do not require us to update our beliefs; they do not provide much new information and therefore are paid less attention and encoded more weakly in memory compared to surprising tokens (Palmeri and Nosofsky 1995, Tulving and Kroll, 1995). In experiments such as Olejarczuk and Kapatsinski's, the log frequency hypothesis predicts that greater weighting of the infrequent/surprising tokens in a distribution's tail results in their overrepresentation in a calculation of category typicality, causing a shift in what people perceive to be the mean. The current results accord well with this explanation. In fact, if the overweighting of lower frequency tokens plays a role in the diachronic change of continuous variables (such as pitch and formant frequency), the same mechanism may be at work in the diachronic change of categorical variables (e.g. phonologically abrupt changes, lexical changes, etc.). For the current experiment, a second, related, explanation for participants' favouring of the distributional tail involves the "perceptual magnet effect", described as "a shrinking of perceptual space" near a category prototype or centre, and an "expansion of perceptual space" in regions further from a category prototype/centre (Kuhl 1991, Burns and Ward 1978 [for musical intervals], MacMillan, Goldberg and Braida 1988, Iverson and Kuhl 1995, Guenther and Gjaja 1996, Feldman, Griffiths and Morgan 2009). Averaging over such a warped space might produce just the type of overshoot found here. These two possible explanations, in fact, may be pointing to the same perceptual phenomenon in emphasizing the greater influence of more distant, low frequency tokens.

The current findings suggest that the effect of distributional shape may be an important contributing factor to sound change and may ultimately deserve inclusion in sound change models. Models that rely strictly on the concept of the raw mean (or equal weighting of exemplars) may be missing important details about the effect of category-internal variation on speakers' determination of target. Systematic biases in the perception of vowels and other continuous data are well-studied (e.g. Repp and Williams 1987, Kuhl 1991, Vallabha and Tuller 2004, Stadler, Richter, Pfaff, and Kruse 1991, Albrecht and Scholl 2010), yet the potential role of such biases in driving long-term sound change has not, to my knowledge, been explored (although they have been implicated in maintaining *stability* [Wedel 2004, 2012]). The current findings have three other implications for future models of sound change. (1) The effect of skewness is likely not restricted to pitch (and vowel formant) continua, but may be generalizable to other continuous and perhaps even categorical variables. Many models of sound change use an appeal to phonetic bias[8] (e.g. Garrett and Johnson 2013) or stressed monosyllables (e.g. Labov 2001) to drive change and are thus restricted to phonetic change or even just vowel change, when many other types of change seem to proceed in a similar manner (e.g. Taglimante and D'Arcy 2009). (2) Distributional shape may at once play a role in change as well as stabilization. Many models of sound change suffer from overgeneration of change: the prediction that change will always happen, or the failure to stop it once it starts (e.g. Pierrehumbert 2001). Skewness may help to continue change, but the normal distribution may have a stabilizing effect. (3) A perceptual sensitivity to distributional shape means that tokens do not need to initially *move* in order to actuate a sound change. Many models introduce a physical bias to begin a change, but it is stipulative to introduce physical bias for some vowels and not others. In contrast to this, a skewed distribution of vowels can result from factors external to the vowel system, thereby eliminating the need to stipulate actuation. For example, skewness could result from an increase in the amount of words containing a vowel in a particular phonological conditioning context.[9] That is, actuation can come from outside the sound system, and the effect of distributional shape plus other system-internal factors can take over from there.

The last point above provides an example of the ways in which distributional shape could interact with lexical frequency, a factor that has been found to play a role in sound change (Bybee 2002, Hay, Pierrehumbert, Walker and LaShell 2015). Indeed, it is important to emphasize that distributional shape can be only one factor of the many that interact to affect the diachronic patterning of vowels. Other factors specific to vowels—such as their positioning relative to other vowels and their susceptibility to articulatory biases—continue to be extremely important. The question is how do these other factors mediate the effect of distributional shape, and vice versa? If two vowels are close together, for example, will the effect of skewness be counteracted by the need to prevent the vowels from moving too close together? Furthermore, general and specific factors may operate in tandem. For example, if stressed monosyllables are further ahead in a change and are focused upon by learners (Labov 2001, Jacewicz, Fox and Salmons 2006), it is conceivable that both vowel specific factors (stress) and the current general perceptual factor reinforce each other in moving vowel shifts along at certain stages of a change. This was illustrated in the present experiment, as distributional shape affected lower pitched stimuli more than higher pitched stimuli, demonstrating the simultaneous effects of a potentially general mechanism on the specific choice of stimulus.

[8] Furthermore, phonetic bias may be overly deterministic in its predictions for the direction of change. In fact, some vowels move in opposite directions in different dialects (e.g. [æ] in the Canadian Shift and the Northern Cities Shift [Boberg 2010:147]).

[9] See Soskuthy (2013) for a detailed complex systems account of sound change, including the role of factors outside the vowel system.

4.2 Recency

The final note of each trial had a "pulling" effect on a participant's output. When final-mean was higher, people overshot more than when final-mean was lower. It is possible that participants simply attempted to copy the final note. It cannot be discounted that this was the strategy adopted by some participants either implicitly or explicitly; however, t-tests on the eight participant sub-sample showed that 4/8 participants differed significantly (at the p<.05 level or lower) on their mean raw output-input *copied* note error (in Phases 1 and 3 of the experiment) vs. their mean raw output-input *final* note error, suggesting there is at least some individual variation in the amount of influence exacted by the final note.

The effect of final note can be seen as a recency effect, a common finding in studies of vowel perception (Cole 1973, Crowder 1973, Hellstrom 1985, Repp and Crowder 1990). The recency effect here is present even though the task was about target determination based on information in the entire trial and did not require any particular attention to the final note. This is different than previous experiments in which attention to single tokens has been much more explicit (e.g. memory tasks and same/different tasks). In previous studies, the recency effect has been attributed to memory decay over time or interference from subsequent events (Pisoni 1975). However, one explanation is that it is due to the loss of specific information for remembered stimuli, and the replacement of this with generic information (Repp and Crowder 1990:123, citing Hellstrom 1985). It is interesting to note the similarity of this explanation to the outcome predicted by explanations emphasizing atypicality such as the log frequency hypothesis. If tones prior to the final tone have lost their individual pitch characteristics, then the final note—provided it is higher or lower than the perceived pitch of the earlier tones—is also effectively the most atypical. Therefore, the effect of recency and distribution shape could have a similar perceptual source. The effect of final note underscores the importance of including recency in models of sound change. The results of this study suggest that this effect is a non-trivial aspect of listeners' category representations, at least in the very short term.

The effect of final note had one further consequence of interest: its consistent pull on participants' output seemed to have the effect of replicating the input distribution shape. The raw output shown in Figure 1 (blue lines) is indeed more skewed in the skewed condition (which has a skew of .71) than in the normal condition (which has a skew of -.07). Thus, in the current experiment, it appears that the effect of final note could help perpetuate the effect of distributional shape by setting up the right distributional conditions for change or stability to continue in the next generation of participants.

4.3 Task/mode specific effects: Pitch bias and pitch region

Participants consistently output pitches that were higher than the mean of the pitches they heard (by 1.9 Hz on average). This was the case even when the distribution was close to normal and even when the final note and the mean were close to the same. The source of the pitch bias is unknown; it could be either task-specific or mode-specific, but does not appear to have any immediate implications for vowel shift. The post-hoc analysis of eight participants (n=4 in each condition) revealed that 7/8 participants output higher tones (by 1.3 Hz on average) even when simply trying to copy a single note. 3/4 of the normal condition participants in this sample overshot the mean more in the experimental trials, which contained variability, than in the tone-copying exercise, which did not contain variability. Therefore, it is possible that a task or mode-specific bias towards higher pitches may have been amplified by the presence of variability *in itself* (separate from the effect of distributional shape discussed in Section 4.1).

The results revealed, in the skewed condition, more overshoot in the lower pitch region than the higher pitch region. It is not the case that the amount of overshoot in the low range *looks* more pronounced but is perceptually the same as the amount of overshoot in the high range: converted into cents, the amount of overshoot in the skewed condition is 20 cents in the low range, vs. only 10 cents in the high range. As with the overall pitch bias, the source of the pitch region difference is unknown. However, the interaction of distributional shape with pitch range highlights the fact that a factor proposed to be general (distributional shape, which could in principle apply to any continuous data) might have effects mediated by features specific to the type of data. In real sound changes, vowel formants and other vowel specific factors are expected to interact with the effect of distributional shape (see Section 4.1, final paragraph). Higher and lower values of a particular formant—that is, different vowel qualities or different vowels—might be more or less susceptible to overshoot, in parallel to the effect of higher vs. lower pitch.

4.4 Non-significant factors of interest

Age and gender were included in the model because they are known to play major roles in sound change. The advancement of sound change is commonly attributed to teenagers and young adults, with older adults being increasingly conservative (e.g. Labov 2001: Ch. 9). As for gender, it is females who lead in the majority of sound changes (e.g. Labov 2001: Ch. 8). It was possible, therefore, that younger speakers and females would exhibit more overshoot in the current experiment;[10] however, this was not the case. This likely reflects the fact that the complex interaction of social factors that brings about age and gender patterns in real sound changes were simply not present in this non-social, experimental setting.

5 Conclusion

This study found that skewness had a (non-significant) effect on increasing mean overshoot towards a distribution's tail. The finding of a potential causal relationship between the distribution a person hears and the output they produce could partly explain how changing vowels continue to change and stable ones continue to be stable. Results also showed a recency effect evident in the pull exerted by a trial's final note, an effect which served to replicate the distributional shape of the input. Two other significant effects, specific to the task or mode, were an overall bias towards higher pitches, and increased overshoot in the lower pitch range of the skewed condition. This is the first study to examine the effect of sub-categorical distributional shape on a production target, and the results provide new information on how variability and perception/learning may interact to catalyze change. The results are consonant with the idea that language is a complex, adaptive system and that the patterns we see are a product of multiple interacting pieces. The challenge now is to discover how this new piece fits in to the sound change puzzle, specifically how distributional shape interacts with other, better known factors (e.g. social factors, phonetic bias, etc.).

---

[10] Although, arguably, even the youngest speakers in this experiment were "too old". Future work would do well to include high school aged speakers.

The most immediate need for follow-up lies in adding negative skewness as a condition in order to confirm the result of the bias towards the tail. Manipulating other aspects of the distributions, including standard deviation and kurtosis, may have further effects on learners' determination of a target. Adding a measure of token quality (an important part of Labov's model) would be interesting for this same reason. However, all of these possibilities should ideally be informed by what the actual input to learners looks like. To this end, it will be important to undertake more precise descriptions of the actual distributional shapes exhibited by different vowels in different dialects. This is true at the individual level as well as the community level. One of the most interesting areas for future study will be in uncovering the reason for the (potential) bias towards a distribution's tail and to see whether the same perceptual mechanism(s) produce different targets, given different distributions. This knowledge would bring us a lot closer to understanding the role this one particular piece, distributional shape, plays in patterns of vowel change and stability.

References

Albrecht, A. R., & Scholl, B. J. (2010). Perceptually averaging in a continuous visual world: extracting statistical summary representations over time. *Psychological Science*, *21*(4), 560–567.

Andruski, J. E., Blumstein, S. E., & Burton, M. (1994). The effect of subphonetic differences on lexical access. *Cognition*, *52*(3), 163–187.

Bailey, G. (2002). Real and apparent time, In J. K. Chambers, P. Trudgill, and N. Schilling-Estes (Eds.) *The Handbook of Language Variation and Change* (pp. 312–32). Malden, MA: Blackwell Publishing.

Bates, D., Maechler, M., Bolker, B. & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi:10.18637/jss.v067.i01.

Boberg, C. (2010). *The English language in Canada: Status, history and comparative analysis*. New York: Cambridge University Press.

Boersma, P. & Weenink, D. (2014). Praat: doing phonetics by computer [Computer program]. Version 5.3.78, retrieved Feb. 2014 from http://www.praat.org/

Burns, E. M., & Ward, W. D. (1978). Categorical perception—phenomenon or epiphenomenon: Evidence from experiments in the perception of melodic musical intervals. *The Journal of the Acoustical Society of America*, *63*(2), 456–468.

Bybee, J. (2002). Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language variation and change*, *14*(03), 261–290.

Cedergren, H. J. (1988). The spread of language change: Verifying inferences of linguistic diffusion. In P. Lowenberg (Ed.) *Language Spread and Language Policy: Issues, Implications, and Case Studies* (pp. 45–60). Washington, DC: Georgetown University Press.

Clarke, S., Elms, F., & Youssef, A. (1995). The third dialect of English: Some Canadian evidence. *Language Variation and Change*, *7*(02), 209–228.

Clayards, M., Tanenhaus, M. K., Aslin, R. N., & Jacobs, R. A. (2008). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition*, *108*(3), 804–809.

Cole, R. A. (1973). Different memory functions for consonants and vowels. *Cognitive Psychology*, *4*(1), 39–54.

Crowder, R. G. (1973). Precategorical acoustic storage for vowels of short and long duration. *Perception & Psychophysics*, *13*(3), 502–506.

Cukor-Avila, P. (2000). Revisiting the observer's paradox. *American Speech*, *75*(3), 253–254.

Feldman, N. H., Griffiths, T. L., & Morgan, J. L. (2009). The influence of categories on perception: explaining the perceptual magnet effect as optimal statistical inference. *Psychological review*, *116*(4), 752–782.

Ferdinand, V., Thompson, B., Kirby, S., & Smith, K. (2013). Regularization behavior in a non-linguistic domain. In M. Knauff, N. Sebanz, M. Pauen, & I. Wachsmuth (Eds.) *Proceedings of the 35th Annual Meeting of the Cognitive Science Society* (pp. 436-441). Austin, TX: Cognitive Science Society.

Garrett, A., & Johnson, K. (2013). Phonetic bias in sound change. In Yu, A. (Ed.), *Origins of sound change: Approaches to phonologization* (pp. 51–97). OUP Oxford.

Guenther, F. H., & Gjaja, M. N. (1996). The perceptual magnet effect as an emergent property of neural map formation. *The Journal of the Acoustical Society of America*, *100*(2), 1111–1121.

Hay, J. B., Pierrehumbert, J. B., Walker, A. J., & LaShell, P. (2015). Tracking word frequency effects through 130 years of sound change. *Cognition*, *139*, 83–91.

Hellström, Å. (1985). The time-order error and its relatives: Mirrors of cognitive processes in comparing. *Psychological Bulletin*, *97*(1), 35.

Hudson Kam, C. L., & Newport, E. L. (2005). Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language, Learning and Development*, *1*(2), 151–195.

Hudson Kam, C. L., & Newport, E. L. (2009). Getting it right by getting it wrong: When learners change languages. *Cognitive psychology*, *59*(1), 30–66.

Iverson, P., & Kuhl, P. K. (1995). Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *Journal of the Acoustical Society of America*, *97*(1), 553–562.

Jacewicz, E., Fox, R. A., & Salmons, J. (2006). Prosodic prominence effects on vowels in chain shifts. *Language Variation and Change*, *18*(3), 285–316.

Jesperson, O. 1949. *A Modern English Grammar on Historical Principles. Part 1: Sounds and Spellings.* London: George Allen & Unwin.

Kang, Y. 2016. A Corpus-Based Study of Positional Variation in Seoul Korean Vowels. *Japanese/Korean Linguistics, 23,* 1-20.

Kochetov, A. (2006). The role of social factors in the dynamics of sound change: A case study of a Russian dialect. *Language Variation and Change*, *18*(1), 99–119.

Kollmeier, T., Brand, and Meyer, B. (2008). Perception of speech and sound. In J. Benesty, M. M Sondhi, & Y. Huang (Eds.) *Springer handbook of speech processing* (pp. 61–82). Berlin: Springer.

Kreuz, R. J. (1987). The subjective familiarity of English homophones. *Memory & Cognition*, *15*(2), 154–168.

Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & psychophysics*, *50*(2), 93–107.

Labov, W. (1994). *Principles of Linguistic Change, vol.1, Internal Factors.* Cambridge, MA and Oxford, UK: Blackwell.

Labov, W. (2001). *Principles of linguistic change Volume 2: Social factors*. Oxford: Blackwell.

Labov, W., Yaeger, M., & Steiner, R. (1972). *A quantitative study of sound change in progress (Vol. 1).* Philadelphia: US Regional Survey.

Macmillan, N. A., Goldberg, R. F., & Braida, L. D. (1988). Resolution for speech sounds: Basic sensitivity and context memory on vowel and consonant continua. *The Journal of the Acoustical Society of America*, *84*(4), 1262–1280.

Maye, J. (2000). *The acquisition of speech sound categories on the basis of distributional information*. (Unpublished doctoral dissertation). University of Arizona.

Maye, J., & Gerken, L. (2000). Learning phonemes without minimal pairs. In S. C. Howell, S. A. Fish, & T. Keith-Lucas (Eds.) *Proceedings of the 24th annual Boston university conference on language development* (pp. 522–533). Somerville, MA: Cascadilla Press.

McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition*, *86*(2), B33–B42.

Miller, J. L., & Volaitis, L. E. (1989). Effect of speaking rate on the perceptual structure of a phonetic category. *Perception & Psychophysics*, *46*(6), 505–512.

Miller, J. L. (1997). Internal structure of phonetic categories. *Language and cognitive processes*, *12*(5–6), 865–870.

Olejarczuk, P. & Kapatsinski, V. (2016). Attention allocation in phonetic category learning. *Proceedings of Cognitive Modeling in Linguistics 14*, 148–156.

Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In J. L. Bybee & P. J. Hopper (Eds.) *Frequency and the emergence of linguistic structure* (pp .137–157). Philadelphia: John Benjamins.

Pisoni, D. B. (1975). Auditory short-term memory and vowel perception. *Memory & Cognition*, *3*(1), 7–18.

Pisoni, D. B., & Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics*, *15*(2), 285–290.

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Repp, B. H., & Crowder, R. G. (1990). Stimulus order effects in vowel discrimination. *The Journal of the Acoustical Society of America*, *88*(5), 2080–2090.

Repp, B. H., & Williams, D. R. (1987). Categorical tendencies in imitating self-produced isolated vowels. *Speech Communication*, *6*(1), 1–14.

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*(1), 27–52.

Sankoff, G. (2006). Age: apparent time and real time. In G. Sankoff and K. Brown (Eds.) *Elsevier Encyclopedia of Language and Linguistics, Second Edition* (pp. 110–15). Burlington, MA: Elsevier.

Singleton, J. L., & Newport, E. L. (2004). When learners surpass their models: The acquisition of American Sign Language from inconsistent input. *Cognitive Psychology*, *49*(4), 370–407.

Soskuthy, M. (2013). *Phonetic biases and systemic effects in the actuation of sound change.* (Doctoral dissertation). University of Edinburgh.

Stadler, M., Richter, P. H., Pfaff, S., & Kruse, P. (1991). Attractors and perceptual field dynamics of homogeneous stimulus areas. *Psychological Research*, *53*(2), 102–112.

Takeuchi, A. H., & Hulse, S. H. (1993). Absolute pitch. *Psychological bulletin*, *113*(2), 345–361.

Tagliamonte, S. A., & D'Arcy, A. (2009). Peaks beyond phonology: Adolescence, incrementation, and language change. *Language*, *85*(1), 58–108.

Trehub, S.E. & Trainor, L.J. (1993). Listening strategies in infancy: the roots of language and musical development. In S. McAdams & E. Bigand (Eds.) *Thinking in Sound: The Cognitive Psychology of Human Audition* (pp. 278–327). London: Oxford University Press.

Vallabha, G. K., & Tuller, B. (2004). Perceptuomotor bias in the imitation of steady-state vowels. *The Journal of the Acoustical Society of America*, *116*(2), 1184–1197.

Van der Ham, S., & de Boer, B. (2015). Cognitive Bias for Learning Speech Sounds From a Continuous Signal Space Seems Nonlinguistic. *i-Perception*, *6*(5), 2041669515593019.

Walker, Robert. (2008). Tune Smithy. Computer program. Available at http://www.tunesmithy.co.uk

Wedel, A. (2004). Category competition drives contrast maintenance within an exemplar-based production/perception loop. *Proceedings of the 7th meeting of the acl special interest group in computational phonology: Current themes in computational phonology and morphology* (pp. 1–10). Stroudsburg, PA: Association for Computational Linguistics.

Wedel, A. (2012). Lexical contrast maintenance and the organization of sublexical contrast systems. *Language and Cognition, 4*(4), 319–355.

Westfall, P. (2014). Kurtosis as peakedness, 1905 – 2014, R.I.P. *The American Statistician, 68*(3), 191–195. http://doi.org/10.1080/00031305.2014.917055

Winter, B. (2013). Linear models and linear mixed effects models in R with linguistic applications. arXiv:1308.5499. [http://arxiv.org/pdf/1308.5499.pdf]

Yaeger-Dror, M. (1994). Phonetic evidence for sound change in Quebec French. In P. A. Keating (Ed.) *Phonological structure and phonetic form (Vol. 3) (pp. 267–293).* Cambridge: Cambridge University Press.