Perceptual bias in learning a vowel nasalization pattern

Kim Strütjen
Department of General Linguistics
Heinrich Heine University Düsseldorf, Germany
struetjen@phil.hhu.de

Dinah Baer-Henney
Department of General Linguistics
Heinrich Heine University Düsseldorf, Germany

Peter Indefrey
Department of General Linguistics
Heinrich Heine University Düsseldorf, Germany

Ruben van de Vijver
Department of General Linguistics
Heinrich Heine University Düsseldorf, Germany

Abstract Do learners use phonetic knowledge when learning phonological patterns? If so, what is the contribution of production and perception to this bias? We investigated these questions with a vowel nasalization pattern in an artificial language taught to Northern Germans. [ã] is easier to produce than [ɛ̃] or [ĩ], whereas [ĩ] and [ɛ̃] are easier to distinguish from [i] and [ɛ] than [ã] from [a]. We found that Northern Germans produce allophonic nasalization in non-high vowels and that they confuse oral and nasal non-low vowels less often with each other than oral and nasal low vowels. In an artificial language learning experiment we found a learning advantage for non-low vowel nasalization over low vowel nasalization. This shows that during phonological learning Northern Germans rely on perceptual cues rather than on articulatory cues. A maximum entropy grammar with constraint weights based on our perceptual confusion data predicts our learning results and thus underscores the role of a phonetic bias during learning.

Keywords: phonological learning, vowel nasalization, perception, production, phonetic bias, German, artificial language, MaxEnt grammar

1 Introduction

Phonetics and phonology have a surprisingly complicated relationship. Blevins (2004) argues that phonetics shapes phonology diachronically exclusively. Vagaries of perception cause changes in the phonology over time, but cannot be used synchronically by language learners. Kingston & Diehl (1994), on the other hand, argue that phonology exerts control over the phonetic implementation of a contrast. However, when it comes to learning phonological patterns, the usefulness of phonetic details is doubted (Moreton & Pater 2012a;b).

Wilson (2006) provided evidence that phonetic detail aids phonological learning by showing that a phonetically motivated pattern is learned better and generalized more often than one that is phonetically motivated to a lesser extent. He concludes that learners have a substantive bias: a learning bias which favors phonetically based patterns. A pattern is phonetically based if it is either easier to produce or if it makes a distinction easier to perceive (Hayes & Steriade 2004). During learning substance acts as a soft bias: phonetics guides learners through the learning process by favoring more phonetically grounded patterns. Obviously, phonetically arbitrary patterns are learnable as well, but they require a greater effort on part of the learner (Albright & Hayes 2011; Baer-Henney & van de Vijver 2012; van de Vijver & Baer-Henney 2014; White 2013).

Evidence for and critique of the substantive bias. Peperkamp, Skoruppa & Dupoux (2006) studied the learnability of phonetically motivated and phonetically unmotivated phonological rules with French native speakers in an artificial language. The phonetically motivated rules were intervocalic fricative voicing and intervocalic stop voicing. These rules are motivated because they make articulation easier. Voiced obstruents are easier to produce between vowels than voiceless obstruents (Westbury & Keating 1986). The changes in the phonetically unmotivated condition linked obstruents in an arbitrary way: [p g z] alternated with [ʒ f t] or [ʃ v d] alternated with [b k s]. The results show that participants were better in learning the phonetically motivated rules than the arbitrary rules and thus provide evidence for a substantive bias.

Wilson dives deeper into the matter by including the degree of phonetic motivation into his considerations. He investigated the learning of a palatalization pattern. In many languages velars palatalize before front vowels: [k] → [tʃ] / __[i]. Guion (1996; 1998) showed that in languages that have velar palatalization, the burst after a velar followed by a high, front vowel perceptually resembles the burst of a palatal stop. If the velar is followed by a mid vowel, there is less resemblance with a palatal stop, and if the vowel is low, there is no resemblance. Wilson created two artificial languages in which velars palatalized before either high, or mid vowels. He trained two groups of learners: one group with palatalization of velars before high vowels and one group with palatalization before mid vowels. In the test phase he found that participants generalized palatalization to novel items if they were trained with palatalization before mid vowels, but not when they were trained with high vowels. He interpreted his results by means of an implicational relationship: learning a phonetically less motivated pattern implies the application in more motivated patterns, but not vice versa.

Moreton & Pater (2012a;b) cast doubt on the interpretation of the results of Peperkamp, Skoruppa & Dupoux (2006) and Wilson (2006). They argue that the experiments have confounded phonetic motivation with structural simplicity. The arbitrary changes in the experiment of Peperkamp, Skoruppa & Dupoux involve changes affecting two features: voicing and continuancy. The phonetically motivated changes only involve one feature: voicing. Therefore, the results can be the consequence of a phonetic bias, but can be explained as the consequence of structural simplicity with just as much justification. The critique of Wilson (2006) involves typology: his results do not completely match the typological data. If a language palatalizes voiced velars, voiceless velars should also be palatalized, but not vice versa. However, what Wilson found was that palatalization of voiceless velars implies palatalization of voiced velars.

Addressing the criticism by means of vowel nasalization. We will address the criticism of Moreton & Pater by exploring the substantive bias and its phonetic motivation in more detail by investigating the learnability of a vowel nasalization rule for vowels of different heights. Nasal vowels are structurally more complex than oral vowels, but all nasal vowels are more complex to the same degree in comparison to their oral counterparts. As to the typology of nasal vowels, they are more marked than their oral counterparts, and it is not clear whether high nasal vowels are more marked than low ones, or the other way around (Hajek 1997). One possible explanation for the unpredictability is that articulatory ease and perceptual ease of vowel nasalization are different. This allows us to distinguish between any separate influence of production or perception on the substantive bias.

Low nasal vowels are easier to produce than high nasal vowels. The reason for this lies in the anatomic connection between the muscles used for lowering the velum and the muscles used for lowering the tongue body. The palatoglossus connects the tongue with the velum. A contraction of the palatoglossus causes a lowering of the velum which means that the velopharnygeal port is open and a nasal sound is articulated (Bell-Berti 1993). The lowering of the tongue body for the production of low vowels is achieved by the hyoglossus. Both muscles, palatoglossus and hyoglossus, are anatomically connected with each other, which means that a lowering of the tongue body by the hyoglossus to produce a low vowel pulls the palatoglossus down, which causes the lowering of the velum and thus the nasalization of the low vowel (Ohala 1975). This is the reason why low vowels are easier to nasalize than high vowels for which additional muscles would be required.

As to perception, the distinction between mid and high nasal and oral vowels is easier to perceive than the distinction between low nasal and oral vowels. The reason for the greater perceptual difference between oral and nasal non-low vowels compared to oral and nasal low vowels lies in the different degree of acoustic modification due to the nasalization. The most prominent acoustic consequences of nasalization are the reduction of the first formant's amplitude (Delattre 1954; Delvaux 2009; Fant 1960; House & Stevens 1956; Macmillan et al. 1999; Pruthi & Epsy-Wilson 2004; Schwartz 1968; Stevens 1998) as well the introduction of additional resonances and anti-resonances in the vicinity of the first formant due to the role of the nasal cavity as an additional resonator (Chen 1995; 1997; House & Stevens 1956; Kingston & Macmillan 1995; Mermelstein 1977; Schwartz 1968; Stevens 1998). Measuring nasality is possible by investigating spectral tilt, which is the difference of amplitude between particular adjacent frequencies. This spectral tilt is very similar in low oral and nasal vowels, but less so in mid or high oral and nasal vowels (Schwartz 1968; Styler 2015). As a consequence, low oral and nasal vowels are acoustically more similar than non-low oral and nasal vowels. This difference is illustrated in the figures below which were created with Praat (Boersma & Weenink 2017). Figure 1 shows the spectral differences between oral and nasal vowels of different heights. The spectra are taken from our material which was recorded by an adult simultaneous bilingual native speaker of European Portuguese and German who was able to pronounce nasal vowels (Azevedo 2005). It can be seen that the oral spectrum (left) and the nasal spectrum (right) are more different from each other in the non-low [i] (top) and [ɛ] (middle) than in the low [a] (bottom). The spectrum of the nasal [ã] is similar to that of the oral [a]. The prominent differences in the low frequency regions are highlighted with ellipses:

The highlighted parts of oral and nasal vowels in mid and high vowels differ from each other whereas those in low vowels are quite similar.

This acoustic similarity between low oral and nasal vowels and the concomitant perceptual similarity was attested in several studies (Ohala 1975; Schwartz 1968; Styler 2015). Native speakers of American English identified high and mid nasal vowels easier than low nasal vowels (Bond 1976). These results are supported by a study of House & Stevens (1956) who showed that the lower the vowel, the more velum lowering is required to judge a vowel as nasal by native speakers of American English even though they are familiar with nasal vowels (Hayes 2009a).

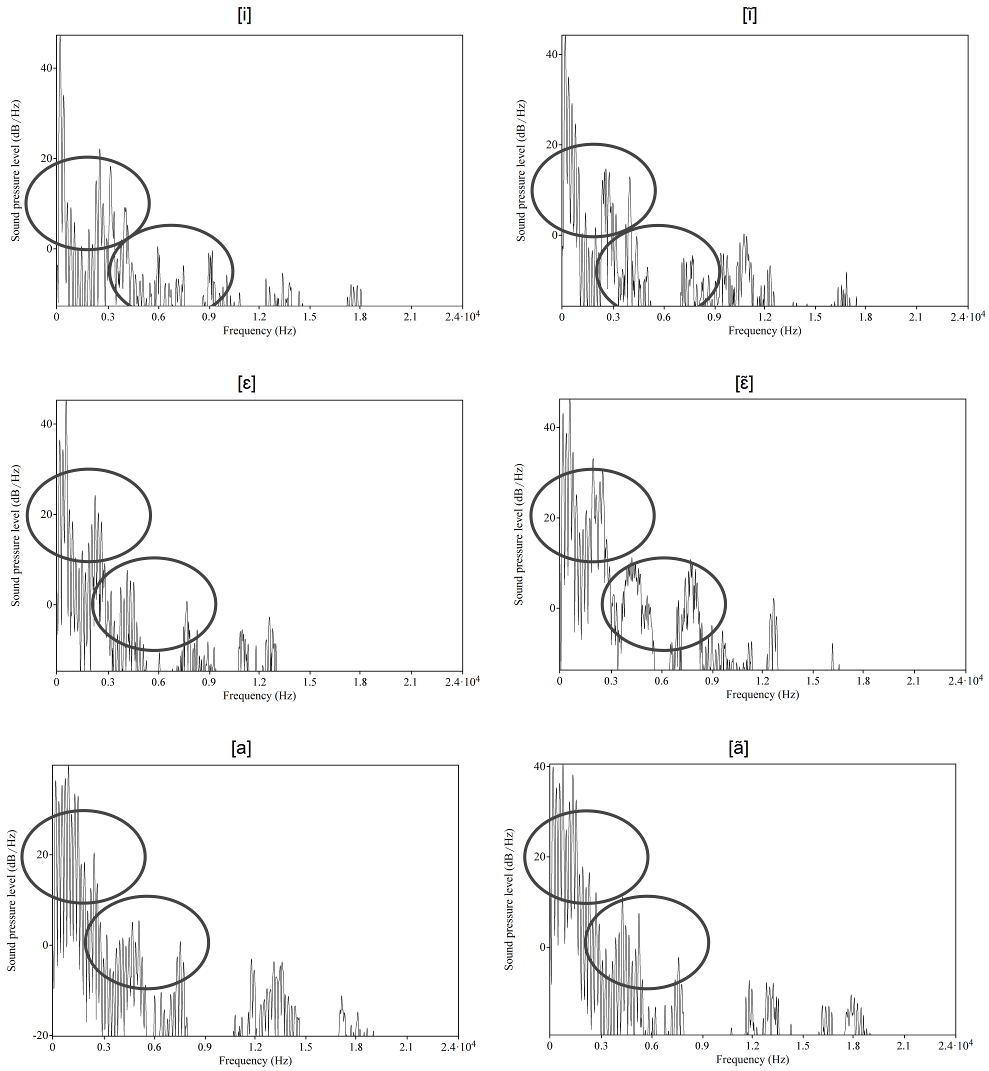

Figure 1: Narrowband spectra of oral vowels (left) and nasal vowels (right). From top to bottom: [i] vs. [ĩ], [ɛ] vs. [ɛ̃], [a] vs. [ã]. The ellipses highlight the frequency regions to compare.

Our experiment. We tested whether a phonetic bias – in production or in perception – affects learning vowel nasalization by Northern German native speakers. First of all, we wanted to check whether the predictions based on the production and on the perception of nasal vowels found in the literature are true for our participants as nasalization in not contrastive in German (Wiese 1996). We therefore conducted an acoustic analysis of a production experiment (section 2) and ran a perception experiment (section 3). Next we did an artificial language learning experiment to investigate whether there is a learning advantage for a specific vowel height nasalization (section 4). In a last step we fed our perception results into a MaxEnt grammar (section 5). We found that a biased grammar can explain our learning results whereas as unbiased does not – proving further evidence for a substantive bias based on perception during synchronic phonological learning.

2 Production experiment and acoustic analysis

We conducted a production experiment in order to assess whether Northern German native speakers produce allophonic nasalization. The scant literature claims that there is no phonemic vowel nasalization (Laeufer 2010; Wiese 1996), and only allophonic nasalization in French loan words in the speech of Southern German speakers (Laeufer 2010). Nevertheless, the connection between the articulation of low vowels and nasalization (Bell-Berti 1993) suggests that this impressionistic claim underestimates the extent of allophonic nasalization. Therefore we recorded and measured the degree of nasality of vowels in different contexts. Should Northern German speakers produce allophonic nasalization, this could be a source of phonetic information that they use in learning a phonological nasalization pattern.

2.1 Method

2.1.1 Stimuli

We created a list of 75 phonetically legal pseudowords. The stimuli consisted of 25 $C_1V_1C_2V_2$-, 25 $C_1V_1C_2V_2$[m]-, and 25 C_1V_1[m]V_2-items. [d p k ʃ v] were used in C_1-position and [b t g z f] were used in C_2-position. V_1 was one of the vowels [o u] and V_2 was one of the vowels [a ɛ i o u]. V_2 was the vowel of which the nasality was measured: after a non-nasal consonant (context CV), before a nasal consonant (context VN), and after a nasal consonant (context NV). As our participants were not familiar with IPA transcriptions we gave all stimuli in German orthography. An example of a CVCV-item is *wusa* [vuza], an example of a CVCV[m]-item is *wusam* [vuzam], and an example of a C1V[m]V-item is *wuma* [vuma]. A complete list of all stimuli used in this experiment is given in the appendix (see table 6).

2.1.2 Procedure and analysis

The participants read aloud the stimuli which were embedded in one of two target sentences. Vowel-final items (CVCV and CV[m]V) were embedded in *Ich habe X gesagt.* 'I said X.' and nasal-final items (CVCV[m]) were embedded in *Ich habe X erblickt.* 'I saw X.'. The sentences were displayed separately on a computer screen in randomized order. After having read out one sentence the participants pressed a button and the next sentence was displayed. Recording took place in an anechoic booth in the phonetics laboratory of the Heinrich-Heine University Düsseldorf. The sampling rate was 48 KHz.

In the recordings we calculated the degree of nasality as reflected in the spectral tilt which is the difference in amplitude between the peaks of A1 and P0. This value was used in several other studies as well (Chen 1995; 1997; Styler 2015; 2017), and revealed congruent and robust results. A1 refers to the amplitude of the highest harmonic near the first formant whereas P0 refers to the amplitude of a nasal peak at about 250 Hz. As proposed by Chen (1997) A1 is either the first (H1) or the second harmonic (H2), which depends on whichever of the two harmonics has a higher amplitude. The smaller the differences, the higher the degree of nasalization. The reason behind this is that a larger velopharyngeal port opening causes a more prominent nasal peak (P0) as well as a more reduced amplitude of the first formant (A1) (Chen 1997). The measurement of A1-P0 may be unsuitable for high vowels as A1 and P0 can occur at the same place in high vowels (Chen 1997). We therefore also measured the A1-P1-difference – although this measurement is criticized as well (Styler 2015; 2017). P1 refers to the amplitude of a nasal peak harmonic at about 950 Hz closest to the first formant.

2.1.3 Participants

15 adult native speakers of Northern German (9 women, 6 men, mean age: 23.7, Range: 21-28) were recorded. No one reported knowledge of a language which uses nasalized vowels distinctively. All participants had normal or corrected vision, no reported hearing problems and did not suffer from hoarseness during the recordings. They participated voluntarily.

2.2 Results

Our original data set consisted of 1125 words (75 stimuli x 15 participants). Eleven words had to be excluded due to incorrect recordings. As nasality has similar acoustic features as creaky voice (Zhang 2015), we also had to exclude all stimuli (n = 36) produced with creaky voice. Creak has several acoustic properties, e.g. irregular F0 (measurable by harmonicity-to-noise-ratio (HNR)), low F0, glottal constriction (measurable with spectral tilt), and damped pulses (Keating, Garellek & Kreiman 2015). The literature provides no guidance as to the cut-off point between creaky voice and nasalization. We therefore conducted the following procedure: We decided to measure HNR with Praat (Boersma & Weenink 2017) to determine which of the recordings sound creaky. As sounds with lower HNR are more creaky, we ordered the stimuli according to their HNR-value and two authors listened independently from each other to the ones with the lowest HNR. In addition to that they checked the distance between the pulses visually because irregularly spaced glottal pulses are a further hint to creaky voice. By doing so both listeners labeled all sounds with a HNR lower than 8 dB as creaky, which was then chosen as cut-off point.

In the remaining recordings (n = 1078) we measured nasality by means of spectral tilt with the help of Praat (Boersma & Weenink 2017) and the Nasality Automeasure script (Styler & Scarborough 2017). This script measured A1-P0- and A1-P1-values at three different time points for each V_2-vowel: at the beginning of the vowel, in the middle of the vowel, and at the end of the vowel. Due to problems with the measurement indicated by the script (A1 = P0 (n = 930), others (n = 353)) only 1951 data points from the original 3234 data points (1078 stimuli x 3 time points) were analyzable. We measured the mean difference between the A1-P0-values or A1-P1-values in the oral context (CV) and the nasal contexts (VN and NV). The greater this difference between oral and nasal contexts is, the more nasal the vowels in the nasal context are. The data were analyzed separately for each participant, as those values are highly speaker-dependent. The results concerning the A1-P0-values can be seen in figures 2-3. The figures were created with ggplot2 (Auguié 2016; Hope 2013; R Core Team 2015; Sarkar 2008; Wickham 2009; 2011). The box plots show the differences in A1-P0-values between the oral context CV and the two nasal contexts VN and NV for each participant dependent on vowel height (high vowels vs. non-high vowels). As low and mid vowels patterned together in our production data we subsumed them under the term non-high vowels.

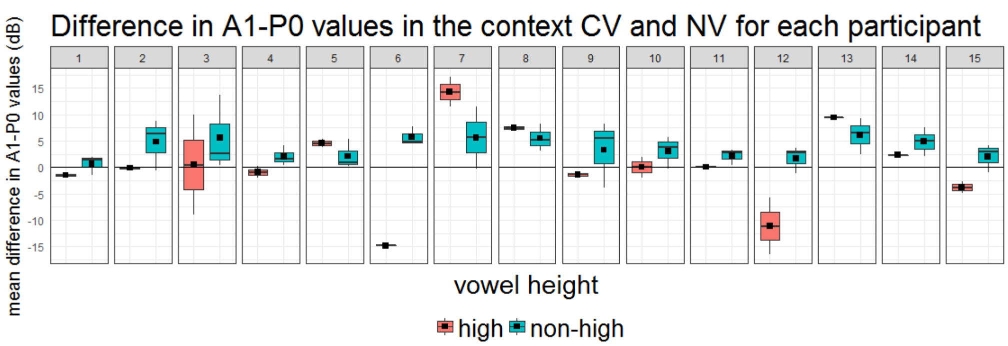

Figure 2: Boxplot showing the difference in A1-P0 values in the context CV (oral) and NV (nasal) for each participant depending on vowel height. Non-high vowels (blue) are almost always higher than high vowels (red), which indicates a greater degree of nasality in NV than high vowels (red). Mean differences are indicated by the black square.

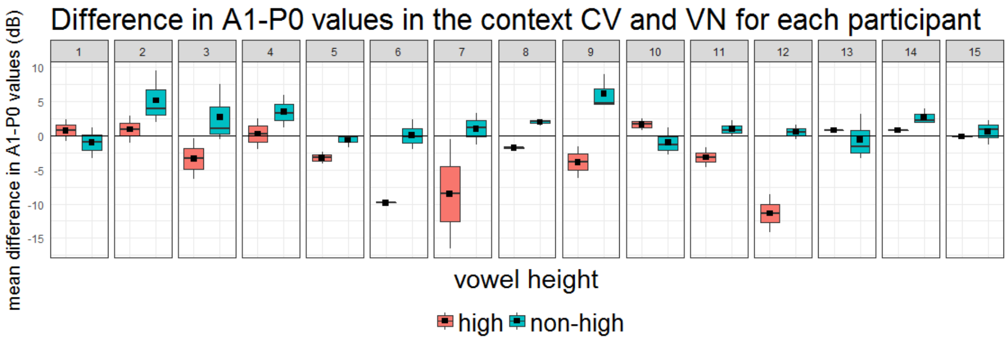

Figure 3: Boxplot showing the difference in A1-P0 values in the context CV (oral) and VN (nasal) for each participant depending on vowel height. Non-high vowels (blue) are almost always higher than high vowels (red), which indicates a greater degree of nasality in VN than high vowels (red). Mean differences are indicated by the black square.

Overall, the blue box plots for non-high vowels are higher than the red ones for high vowels in both figures, which means that there is more nasalization in non-high vowels than in high vowels. To establish whether nasality depends on vowel height and context we performed a linear mixed effects analysis using R (R Core Team 2015) and the corresponding packages lme4 (Bates et al. 2015) and lmerTest (Kuznetsova, Brockhoff & Christensen 2018). The participants and vowel heights were analyzed separately. The A1-P0-values served as dependent variable and context served as independent variable. Items were random intercepts.

The model showed significant differences between CV and NV in non-high vowels for all participants, but no significant differences in high vowels. The difference between CV and VN was marginally significant in non-high vowels for ten of our 15 participants, whereas five participants (1, 2, 5, 10, 11) did not show a significant difference in non-high vowels. There was no significant difference between CV and VN in high vowels.

As the measurement of A1-P0 may be unsuitable for high vowels, we also measured A1-P1. The results can be seen in the figures 4-5.

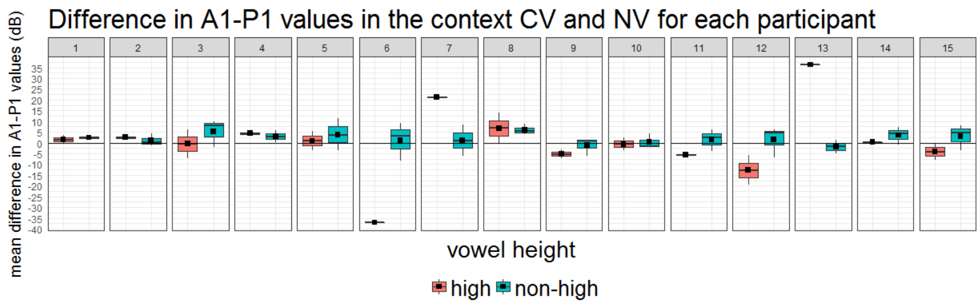

Figure 4: Boxplot showing the difference in A1-P1 values in the context CV (oral) and NV (nasal) for each participant depending on vowel height. Non-high vowels (blue) are almost always higher than high vowels (red), which indicates a greater degree of nasality in NV than high vowels (red). Mean differences are indicated by the black square.

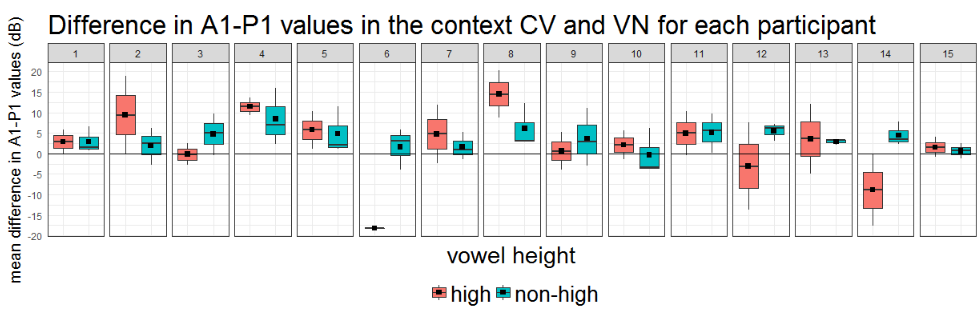

Figure 5: Boxplot showing the difference in A1-P1 values in the context CV (oral) and VN (nasal) for each participant depending on vowel height. Non-high vowels (blue) and high vowels (red) have almost always positive values, which indicated a greater degree of nasality in VN than in CV. Mean differences are indicated by the black square.

A linear mixed effects analysis revealed the following results: 13 of 15 participants showed significant differences in A1-P1-values between CV and NV in non-high vowels, two of 15 participants (12, 15) showed marginally significant differences. There was no significant difference in A1-P1-values between CV and NV in high vowels. For the VN context the results are the following: Despite of four of our 15 participants (2, 4, 5, 7), who showed only a marginally significant difference between CV and VN in non-high vowels, all A1-P1-differences between CV and VN were significant – independent of speaker and height. In comparison to other languages the nasalization effects in Northern German are small (see Chen 1997: for data on English and French).

2.3 Discussion

The results show that there is a small amount of allophonic nasalization in Northern German. Looking at the A1-P0-values non-high vowels are slightly allophonically nasalized in NV and in VN contexts by Northern German native speakers. High vowels are not nasalized in NV or VN contexts. Looking at the A1-P1-values non-high vowels are slightly allophonically nasalized in NV and VN contexts and high vowels are slightly allophonically nasalized in VN context. This means that there is allophonic nasalization in non-high vowels in Northern German. In high vowels we only find allophonic nasalization in the context VN, which indicates an allophonic regressive nasalization in Northern German (pace Laeufer (2010) and Wiese (1996)). The pattern produced by Northern Germans confirms the tight connection in production between low vowels and nasalization (Bell-Berti 1993; Ohala 1975).

As Northern German speakers produce a slight amount of nasalization in non-high vowels, they might benefit from this when learning a phonological pattern involving vowel nasalization.

3 Perception experiment

In the perception experiment the participants were asked to identify oral and nasal vowels in order to asses their perceptual distance. This task tests the perceptual confusability of oral and nasal vowels of different heights. On the basis of acoustics, and previous studies on the perception of oral and nasal vowels (Bond 1976; House & Stevens 1956), we expect that low oral and nasal vowels are more likely to be confused with each other than non-low oral and nasal vowels. In contrast to Bond (1976); House & Stevens (1956) we will investigate the perception of nasalization in speakers who are not familiar with strong nasalization.

3.1 Method

3.1.1 Stimuli

The stimuli were the three oral vowels [a ɛ i] and their three nasal counterparts [ã ɛ̃ ĩ]. The vowels were spliced out of CV(C)-syllables. The oral vowels were spliced out of CV-syllables and the nasal vowels were spliced out of CV[m]-syllables. In addition to the experimental vowels we used the four vowels [o õ u ũ] for a short practice phase at the beginning of the experiment. There was exactly one token of each vowel for all parts of the experiment.

The stimuli were recorded in an anechoic booth in the phonetics laboratory at Heinrich-Heine University Düsseldorf. The sampling rate of the recording was 48 kHz. The material was recorded by a fully bilingual German-Portuguese native speaker. The intensity was scaled to 70 dB and white noise was added to the experimental vowels at a signal-to-noise ratio of 15 dB using Praat (Boersma & Weenink 2017). On the basis of a discrimination task (see Appendix) we decided to add white noise to the material, in order to make the identification task more difficult. Stimuli without noise would probably induce few confusions, which means that the results would not offer us enough information about the vowels' perceptual similarity and perceptual distance (White 2017). The practice vowels were not masked with noise.

3.1.2 Procedure

The participants were tested with a forced-choice identification task which was scripted in Praat (Boersma & Weenink 2017). In a quiet room they listened to the stimuli via headphones. In a short introductory phase the participants became familiar with the experiment and its setting. During this introductory phase they learned how the sounds are represented orthographically on the screen; we used the four vowels [o õ u ũ]. At the same time they saw a transcription on the screen. The participants learned that an oral vowel was represented by an orthographic symbol of the vowel: [o] was transcribed as <o> and the nasal vowel was transcribed with a tilde above the corresponding oral vowel, e.g. <õ>. In the subsequent test phase [a] was transcribed as <a>, [ɛ] as <ä>, and [i] as <i>, whereas the nasal vowels were transcribed with a tilde above. Oral vowels were always written in red and nasal vowels in blue. Participants listened to vowels with masking noise and were forced to identify each vowel as one of these: [a], [ã], [ɛ], [ɛ̃], [i], [ĩ]. The participants responded by touching one of the vowels on a touch screen. After they had made their choice the next vowel was presented. Each vowel was presented ten times in random order. The experiment lasted about ten minutes.

3.1.3 Participants

We tested 30 adult native speakers of Northern German (15 women, 15 men, mean age: 34.5, Range: 19-67) who did not take part in the former experiment. The participants had no knowledge of a language with phonemic vowel nasalization. All participants had normal or correct vision and no hearing problems. All of them participated voluntarily.

3.2 Results

In the confusion matrix (see table 1) it is shown how often each stimulus was identified as one of the six vowels. For example, the auditory stimulus [ã] was identified as <ã> 112 times – 37% of all responses given to the auditory stimulus [ã] – and it was identified as <a> 168 times – 56% of all responses given to the stimulus [ã]. We are interested in how often each oral vowel was confused with its nasal counterpart, and vice versa. The relevant cells are shaded in table 1.

Table 1: Confusion matrix of oral and nasal vowels in Northern German. Each row corresponds to one of the stimuli (in IPA) and each column corresponds to one of the available responses (in German orthography). Proportions are given in brackets. Shaded cells highlight the important comparisons.

stimulus \ response	<ã>	<a>	<ä̃>	<ä>	<ĩ>	<i>
[ã]	112 (0.37)	168 (0.56)	3 (0.01)	6 (0.02)	7 (0.02)	4 (0.01)
[a]	20 (0.07)	276 (0.92)	2 (0.01)	1 (< .00)	1 (< .00)	0 (0.00)
[ɛ̃]	44 (0.15)	18 (0.06)	121 (0.40)	45 (0.15)	61 (0.20)	11 (0.04)
[ɛ]	10 (0.03)	2 (0.01)	1 (0.06)	260 (0.87)	4 (0.01)	6 (0.02)
[ĩ]	71 (0.24)	7 (0.02)	38 (0.13)	10 (0.03)	143 (0.48)	31 (0.10)
[i]	0 (0.00)	0 (0.00)	0 (0.00)	1 (< .00)	4 (0.01)	295 (0.98)

In order to assess the perceived distance on the basis of the confusion matrix in table 1 we calculated *d'*. The results can be seen in figure 6. A linear mixed effect model with *d'* as dependent variable, vowel as independent variable and participant as random intercept showed that the perceived distance based on *d'* between oral and nasal [i] is significantly greater than the perceived distance between oral and nasal [ɛ] (β = -1.86, SE = 0.67, df = 58.00, t = -2.78, $p < 0.01$) and between oral and nasal [a] (β = -4.01, SE = 0.67, df = 58.00, t = -5.99, $p < 0.001$). There is also a significant difference between the [ɛ]-[ɛ̃]-distance and the [a]-[ã]-distance (β = -2.15, SE = 0.67, df = 58.00, t = -3.21, $p < 0.01$) (Baayen 2013; Bates et al. 2015; Knoblauch 2014; Kuznetsova, Brockhoff & Christensen 2018; R Core Team 2015).

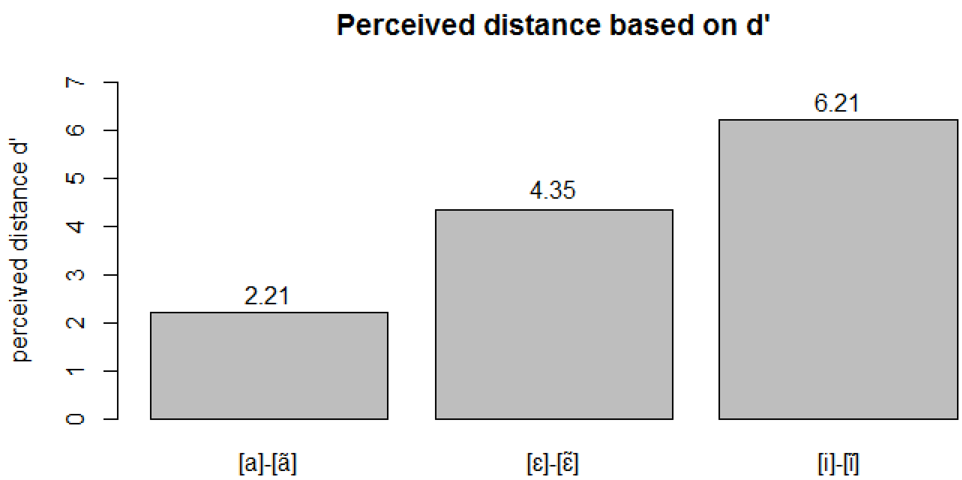

Figure 6: Perceived distance based on *d'* between oral and nasal vowels. The higher the vowel, the greater is the perceived distance between the oral and the nasal vowel.

3.3 *Discussion*

The results confirm our expectations that Northern German native speakers are more likely to confuse [a] with [ã] than [ɛ] with [ɛ̃] or [i] with [ĩ]: The lower the vowel, the more difficult is the perception of the oral-nasal-contrast.

This pattern confirms the perceptual similarity between low oral and nasal vowels, and the dissimilarity between non-low oral and nasal vowels as shown by Bond (1976); House & Stevens (1956) – even in a language which has only a small amount of nasalization (see section 2).

We continue our investigation of the hypothesis that native speakers of Northern German are biased by phonetics in phonological learning. We are further interested in whether this phonological learning is affected to a larger degree by production or by perception.

4 Learning experiment

The results of our previous experiments show that the perceptibility and production of nasal vowels is affected by vowel height. The acoustic analysis showed that non-high vowels are more easily nasalized than high vowels. The reason for this is the articulatory connection between lowering of the tongue and opening the velar port (see section 2). Hence, production favors non-high vowel nasalization. In perception, on the other hand, non-low oral and nasal vowels are easier to distinguish than low oral and nasal vowels. The reasons for this are the similarities in spectral tilt for low oral and nasal vowels (see section 3). Hence, perception favors non-low vowel nasalization. In our learning experiment we study the learnability of vowel nasalization of different heights, and investigate whether production or perception is guiding the learning process or whether neither of these factors is guiding the learning process.

If phonetics in general acts as a bias, we expect a learning advantage dependent on vowel height: More specifically, if production acts as a bias, we expect a learning advantage for nasalization of non-high vowels over high vowels. If perception acts as a bias we expect a learning advantage for nasalization of non-low vowels over low vowels. If phonetics does not act as a bias, we expect no learning advantage for any of the three groups. See table 2 for an overview of the different predictions.

Table 2: Predictions for the learning experiment.

substantive bias based on production	**substantive bias based on perception**	**no substantive bias**
non-high > high	non-low > low	high = mid = low

Questions of biases in learning are usually investigated by means of poverty-of-the-stimulus experiments (Baer-Henney 2015; Baer-Henney, Kügler & van de Vijver 2015; Cristià & Seidl 2008; Finley 2008; 2012; White 2013; 2014; White & Sundara 2014; Wilson 2003; 2006; Zuraw 2007). In such experiments participants are exposed to one pattern, but are tested on new patterns that they had not been exposed to.

We created an artificial language by means of which we compared three groups of Northern German learners who all learned a nasalization pattern but each with nasal vowels of a different height. Our participants learned that vowels are nasalized before [m] (see 1). In this context we found a small degree of vowel nasalization in non-high and high vowels. VN will reveal a phonetic bias, if any, and will allow us to observe it.

(1) $$/\mathrm{V}/ \rightarrow [\tilde{\mathrm{V}}] \ / \ \_[\mathrm{m}]$$

4.1 *Method*

4.1.1 *Stimuli*

The artificial language consisted of singular, plural, and diminutive forms. The stimuli were constructed from a subset of the German and Portuguese phoneme inventories. All items are in agreement with the phonotactics of German (Wiese

1996) – except for the nasalized vowels. The structure of the stimuli and examples of each grammatical form are illustrated in table 3. The singulars were pseudowords of the form $C_1V_1C_2V_2$ with a high, mid or low vowel as V_2. The plural was expressed by a final [m], which caused nasalization of the preceding vowel, and the diminutive was expressed by a final [l] without any other phonological change. [p d k ʃ v] were used in C_1-position and [b t g f z] were used in C_2-position. V_1 was one of the back vowels [o u] whereas V_2 was one of the front vowels [i ɛ a]. For the training phase we used 48 items for each group of participants (16 singulars, 16 plurals, and 16 diminutives). For the test phase we used 48 stimulus pairs, each consisting of a form which conformed to the nasalization rule (e.g. the correct plural [kogãm] and the correct diminutive [dufil]) and one which does not (e.g. the incorrect plural [kogam] and the incorrect diminutive [dufĩl]). Half of the pairs (n = 24) tested the plural, and half of the pairs (n = 24) tested the diminutive. In both stimulus groups there was an equal number for each of the three vowel heights (n = 8). Half of these pairs were part of the training items, and half of them were not. The complete set of all stimuli is listed in the appendix (see table 7-10).

Table 3: Structure of the pseudowords.

	C_1	V_1	C_2	V_2	suffix	example
singular	[p d k ʃ v]	[o u]	[b t g f z]	[a ɛ i]	∅	[koga]
plural	[p d k ʃ v]	[o u]	[b t g f z]	[ã ɛ̃ ĩ]	[m]	[kogãm]
diminutive	[p d k ʃ v]	[o u]	[b t g f z]	[a ɛ i]	[l]	[kogal]

The stimuli were recorded by the same bilingual native speaker of Portuguese and German who did the recording for the other experiment. Recording took place in an anechoic booth in the phonetics laboratory of the Heinrich-Heine University Düsseldorf. The sampling rate was 48 KHz. Stimuli were recorded as the answer to the Portuguese question *O que é que eu disse?* 'What did I say?' to focus on the stimulus item and to ensure a uniform language environment that allows the reader to naturally produce nasalized vowels. The target sentence itself was read silently. After recording the intensity of all stimuli was adjusted to 70 dB using Praat (Boersma & Weenink 2017).

4.1.2 Procedure

The experiment was divided into a perceptual training phase and a perceptual forced-choice test. It was scripted in the software PsychoPy (Peirce 2007) and it ran on a Windows laptop. Participants listened to the auditory stimuli via headphones. The experiment lasted 15 minutes and took place in a quiet room in the phonetics laboratory of the Heinrich-Heine University Düsseldorf.

In the experiment the poverty-of-the-stimulus method (Wilson 2006) was used. Our participants were trained on a subset of the stimuli but tested on all stimuli – including stimuli known from training and stimuli they had not yet heard. There were three training groups: During training each group heard the nasalization of only one vowel; either [a], or [ɛ], or [i]. For example, a member of the [a]-group never received any training for [ɛ] or [i]. The test was identical for all groups and included all vowels.

In a short introductory phase the participants were familiarized with the set up of the experiment. They listened to three German animal names in the singular, plural, and diminutive, e.g. *Hase* 'rabbit', *Hasen* 'rabbits', *Häschen* 'little rabbit'. A picture of a rabbit was shown simultaneously with the auditory stimulus to illustrate their meaning. The pictures are part of the Snodgrass & Vanderwart (1980) collection.

The introductory phase was followed by the training and test phase. The participants were told that the experiment consisted of two parts, a 'first phase' and a 'second phase' (Wilson 2006). During training our participants heard two repetitions of 48 stimuli (16 singulars, 16 plurals, and 16 diminutives) in randomized order. An auditory stimulus was played while a picture was shown for 1000 ms. The pictures were fantasy animals (van de Vijver & Baer-Henney 2014). A singular form was accompanied by a single fantasy animal, a plural form by two fantasy animals, and a diminutive form by a small fantasy animal. There was an inter-stimulus-interval of 500 ms (see figure 7). During training participants received positive input only, which means that they never listened to an incorrect plural, e.g. [kogam], or to an incorrect diminutive, e.g. [dufĩl].

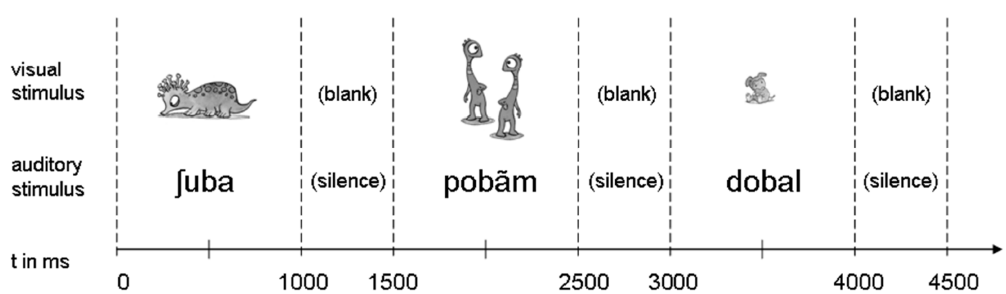

Figure 7: Time course of training phase.

The test after the training was a forced-choice task and was identical for all groups. There were 48 stimulus pairs which consisted of a correct and an incorrect form which only differed in the nasalization of V_2. Half of the pairs ($n = 24$) tested plurals and half of the pairs ($n = 24$) tested diminutives. There were 16 pairs with high, mid, and low vowels. A trial consisted of the presentation of the first form, followed by the presentation of the second form, each lasted one second (see figure 8). The inter-stimulus-interval was 200 ms. During the auditory presentation the corresponding visual support was displayed. After that the participants had 3000 ms to decide which of the two forms was correct by pressing either the right or the left arrow key. After an inter-trial-interval of 500 ms the next stimulus pair was presented. The stimulus pairs were presented in randomized order.

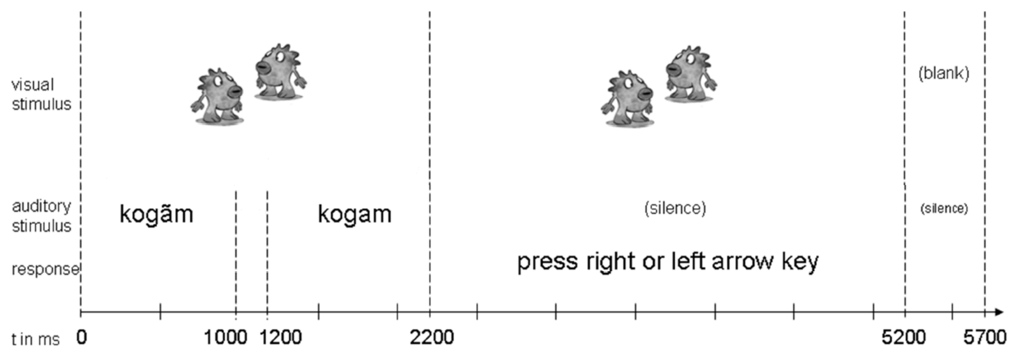

Figure 8: Time course of one trial in the test phase.

4.1.3 *Participants*

61 native speakers of Northern German took part in the experiment (39 women, 22 men, mean age: 28.0, Range: 18-74). They were randomly assigned to one of the three experimental groups. 20 participants were trained with nasalization of the high vowel [i], 20 participants were trained with nasalization of the mid vowel [ɛ], and 21 participants were trained with nasalization of the low vowel [a]. All groups were tested with all vowels [i ɛ a]. Participants were given a small expense allowance for their participation. None of them had participated in the previous experiments, nor did they have any knowledge of a language with phonemic nasal vowel contrasts.

4.2 *Results*

From 2928 data points (61 participants x 48 trials) 51 were not analyzable because the participants did not respond within 3000 ms. We analyzed the proportion of correct answers from the remaining 2877 data points. Overall [ɛ]-learners yielded the best performance. 74% of their responses to all items were correct. The percentage of correct responses of [i]-learners to all items was 70%, and that of [a]-learners was 61%.

We calculated a generalized logistic mixed effects model with accuracy as dependent variable and vowel height of the trained vowel as independent variable with an interaction with training and form (plural or diminutive); participants, items, and order were random intercepts (Baayen 2013; Bates et al. 2015; R Core Team 2015). A comparison of the three learning groups showed that the [i]-learners' performance did not differ significantly from that of the [ɛ]-learners ($\beta = -0.47$, $SE = 0.37$, $z = -1.28$, $p = 0.20$). The [ɛ]-learners ($\beta = -0.86$, $SE = 0.33$, $z = -2.59$, $p < 0.01$) and the [i]-learners ($\beta = -1.33$, $SE = 0.35$, $z = -3.78$, $p < 0.001$) performed significantly better than the [a]-learners. An overview of the results can be seen in figure 9 and table 4.

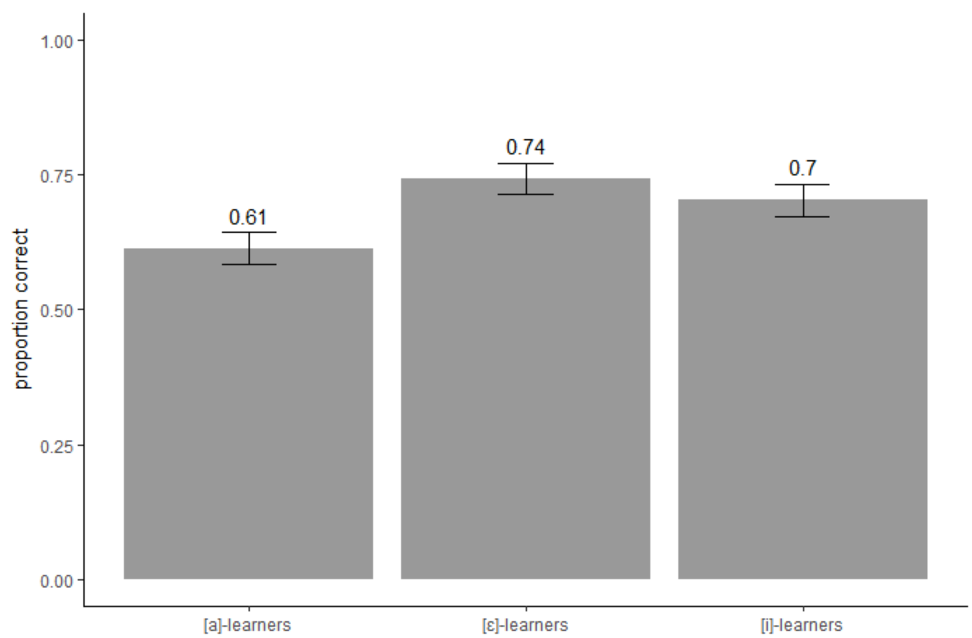

Figure 9: Overall results in the three experimental conditions: Proportion of correct responses across all vowel heights ± 1.96 *SE*.

Table 4: Results of the generalized logistic mixed effects analysis: Fixed effects for overall results.

	Est.	*SE*	*z*-value	*p*-value
(Intercept)	1.67	0.28	5.89	$< .001$
training [a]	-1.33	0.35	-3.78	$< .001$
training [ɛ]	-0.47	0.37	-1.28	0.20
training untrained	-1.51	0.29	-5.25	$< .001$
form PL	0.37	0.38	0.98	0.33
training [a]*training untrained	1.19	0.40	3.01	< 0.01
training [ɛ]*training untrained	0.66	0.41	1.61	0.11
training [a]*form PL	0.42	0.47	0.90	0.37
training [ɛ]*form PL	0.67	0.52	1.30	0.20
training untrained *form PL	0.52	0.42	1.24	0.22
training [a]*training untrained*form PL	-0.56	0.59	-0.95	0.35
training [ɛ]*training untrained*form PL	-0.15	0.63	-0.23	0.82

We were also interested in the learning and generalization behavior across experimental groups. As we found a significant interaction with training in the overall results, we calculated separately the proportion of correct responses for trained vowels and for untrained vowels. We investigated how participants judged stimuli with vowels of different height. We analyzed the data by means of generalized logistic mixed effects analysis. In the model which fitted our data best (as assessed by backward stepwise elimination (Baayen 2008)) accuracy served as dependent variable and vowel height of the test items and form (plural or diminutive) as independent variable. Participants, items and order were random intercepts. The results are illustrated in figure 10.

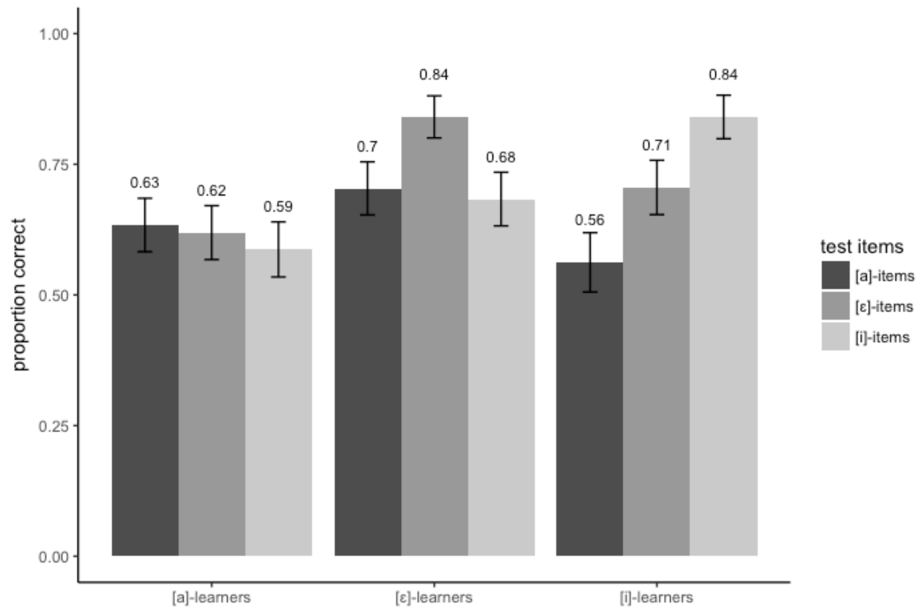

Figure 10: Results in the three experimental conditions for all vowel heights: Proportion of correct responses ± 1.96 *SE*.

[a]-Learners responded correctly to items with [a] in 63% of the cases, to items with [ɛ] in 62%, and to items with [i] in 59%. The percentage correct to [a]-items did not differ significantly from [ɛ]-items (β = -0.08, SE = 0.20, z = -0.41, p = 0.68), nor from [i]-items (β = -0.24, SE = 0.20, z = -1.24, p = 0.22). There was also no significant difference between correct responses to [ɛ]-items or [i]-items (β = -0.16, SE = 0.19, z = -0.83, p = 0.41).

[ɛ]-Learners responded correctly to items with [ɛ] in 84% of the cases, to items with [i] in 68%, and to items with [a] in 70%. The percentage correct responses to [ɛ]-items differed significantly from [a]-items (β = -0.86, SE = 0.27, z = -3.24, $p < 0.01$) and [i]-items (β = -1.01, SE = 0.26, z = -3.83, $p < 0.001$). The correct responses to [a]-items and [i]-items did not differ significantly from each other (β = -0.15, SE = 0.25, z = -0.60, p = 0.55).

[i]-Learners responded correctly to items with [i] in 84% of the cases, to items with [ɛ] in 71%, and to items with [a] in 56%. The percentage correct responses to [i]-items differed significantly from [a]-items (β = -1.53, SE = 0.26, z = -5.87, $p <$ 0.001) and [ɛ]-items (β = -0.84, SE = 0.26, z = -3.19, $p < 0.01$). The percentages correct responses to [a]-items and [ɛ]-items differed significantly as well (β = 0.70, SE = 0.24, z = 2.90, $p < 0.01$).

We also compared the results to the trained vowel in each group by means of a generalized logistic mixed effects analysis. Accuracy was the dependent variable and vowel height and form were independent variables. Participants, items, and order were random intercepts. The model with an interaction between vowel height and form fitted our data best (as assessed by backward stepwise elimination (Baayen 2008)). In the trained condition the percentage of correct responses of [ɛ]-learners was 84%, of [i]-learners was 84%, and of [a]-learners was 63%. The percentage correct of [ɛ]-learners and [i]-learners did not differ from each other (β = -0.39, *SE* = 0.36, z = -1.09, p = 0.28). Both groups gave significantly more correct responses than [a]-learners ([ɛ]-learners: β = -1.20, *SE* = 0.32, z = -3.72, $p < 0.001$; [i]-learners: β = -1.60, *SE* = 0.34, z = -4.64, $p < 0.001$).

As shown in figure 10 [i]- and [ɛ]-learner were more successful in their trained vowel than [a]-learners. The [a]-learners generalized vowel nasalization to vowels of different heights to the least degree. Their proportion of correct responses to different heights was almost identical. The [ɛ]-learners distinguished between the trained vowel [ɛ] and the untrained vowels [a] and [i], which seemed to be treated identically, whereas the [i]-learners generalized differently to all three vowel heights.

4.3 *Discussion*

The artificial language learning experiment shows that native speakers of Northern German are able to learn a nasalization pattern. Their performance differed regarding vowel height. Participants who were trained with non-low vowels did significantly better than participants who were trained with low vowels. This is true for both the results in the trained and in the untrained conditions.

The results in the trained conditions indicate that the pattern with nasal high and mid vowels is learned with greater ease than the pattern with nasal low vowels. Participants responded best to vowels they were trained with, but [a]-learners responded correctly to less items than [ɛ]- and [i]-learners. These results are similar to the results of the perception experiment, in which we found that [a] and [ã] are most likely to be confused.

The [ɛ]-learners generalized most, both to [a]-items and to [i]-items. [i]-learners generalized to [ɛ]-items and to a lesser extent to [a]-items. [a]-learners generalized least, but, to the extent that they did, in equal measure to [ɛ]- and [i]-items. These results, too, are similar to the results of the perception experiment. In general, learning nasalization of low vowels is difficult. [a]-learners make most mistakes – both in the trained condition and in the untrained conditions – and [i]-learners make more mistakes in items containing [a] than in items containing the other vowels.

These findings are in line with the account that a bias driven by perception – not production – influences phonological learning (see table 2). On the basis of these results it is conceivable that the phonetic confusability of nasal and oral vowels is the basis for the phonological grammar that our participants used to generalize the pattern in our artificial language to novel items. We created a maximum entropy grammar with which we aim to predict the results of our learning experiment.

5 Modeling the data with a maximum entropy grammar

We tested whether perceptual confusability can account for the results of our learning experiment – and thus whether there is evidence for a phonetic bias in phonology – by providing a maximum entropy learning algorithm (Goldwater & Johnson 2003) with a substantive bias based on perception. We used this algorithm as it has been applied in former experimental phonological studies and has been proven to converge (Goldwater & Johnson 2003; Jäger 2004). For an overview of the application of maximum entropy grammars in experimental phonology see Goldwater & Johnson (2003); Hayes & Wilson (2008); White (2017); Wilson (2006).

We implemented the bias by means of our confusion data (see section 3). These data provided us with information about the relative difficulty in the perception of oral and nasal vowels by Northern German speakers. With these data we were thus able to establish a perceptually based grammar with which we tested our hypothesis that learning a vowel nasalization pattern is facilitated by different degrees of perceptual difficulty. We provided the grammar with the same input that our participants received in the training phase of the learning experiment (see section 4), and we then tested it with the same items that our participants were confronted with during the test phase. Then we compared the grammar's predictions with the results of our learning experiment. If the model predicts the same results as we obtained in our learning experiment, there is a relationship between perceptual similarity and the learning performance. However, if the results differ, the learning results can not be attributed to perceptual similarity, and there is no evidence for a phonetic bias.

5.1 *Grammar*

We constructed a maximum entropy (MaxEnt) grammar consisting of weighted constraints that represent a perceptual bias. With these weighted constraints MaxEnt grammars calculate probabilities of possible output forms. The aim is to find those weights that maximize the probability of the learning data. Learning stops when the maximum is achieved. For a detailed description of MaxEnt grammars and their mathematical bases see Goldwater & Johnson (2003); Hayes & Wilson (2008).

5.1.1 *Constructing the grammar*

To implement the grammar we used the MaxEnt Grammar Tool developed by Wilson & George (2009) (see manual for MaxEnt Grammar Tool (Hayes 2009b) for detailed information).

We created a file which contained input forms – the stimuli from our perception experiment – and candidates for each input form – the responses given in the perception experiment. The frequency of these candidates was also included – the number of times a vowel was confused with another vowel in the perception experiment. Additionally, the file contained our constraints and their violations. Our constraints were from the family of *MAP(x,y) constraints (White 2017; Zuraw 2007), which belong to the class of correspondence constraints (McCarthy & Prince 1995). We followed the procedure described by White (2017), who created a MaxEnt grammar using *MAP constraints and formulated a substantive bias based on the p-map (Steriade 2001; 2009). The p-map is a perceptual map of contrasts that are confused with each other. According to White's (2017) symmetric interpretation of Zuraw's (2007) *MAP constraints, *MAP(x,y) is violated when x mapped onto y, and vice versa.

In addition to the input-candidate-file we also created a second file to calculate the weights for the *MAP constraints with which the confusion probabilities are predicted best. The weights in MaxEnt grammars are given a value for their mean and their standard deviation (White 2017). We initially set the mean to zero and the standard deviation to 10,000. The weights that we obtained in this way are provided in table 5.

Table 5: Constraints and their weights used in our learning simulation. The weights of the *MAP(x,y) constraints were calculated by the MaxEnt grammar based on data from our perception experiment.

constraints	weights
*MAP(a,i)	12.13
*MAP(a,ɛ)	4.95
*MAP(ã,i)	4.65
*MAP(ɛ,i)	4.28
*MAP(a,ĩ)	3.82
*MAP(ɛ̃,i)	3.71
*MAP(ɛ,ĩ)	3.41
*MAP(ã,ɛ)	3.20
*MAP(a,ɛ̃)	2.90
***MAP(ĩ,i)**	**2.55**
*MAP(ã,ɛ̃)	1.96
***MAP(ɛ̃,ɛ)**	**1.91**
*MAP(ã,ĩ)	1.57
*MAP(ɛ̃,ĩ)	1.31
***MAP(ã,a)**	**0.57**

Highly confusable mappings are the ones which are perceived as similar to each other by the participants in the perception experiment (see section 3). According to White (2017) we expect lower weights for more confusable mappings and higher weights for less confusable mappings. For example, *MAP(a,i) has a weight of 12.13 because these two sounds had been perceived as different by our participants and thus had not been confused with each other (0 times out of 571 total mappings of these vowels). The constraint *MAP(ã,a) has a weight of 0.57 because these two sounds had been perceived as similar by our participants and thus had been confused with each other often (168 + 20 = 188 times out of 576 total mappings of these vowels).

5.1.2 Testing the grammar

We compared the output of our grammar with the results of the learning experiment (see section 4). To test the grammar we created another input-candidate-file. The input forms were the types of test items of our learning experiment: [a]-plural, [a]-diminutive, [ɛ]-plural, [ɛ]-diminutive, [i]-plural, and [i]-diminutive. The candidates were the available responses for each type of item. For type [a]-plurals these were the [a]-plural with nasalization (correct) and without nasalization (incorrect), and for type [a]-diminutives these were [a]-diminutive with nasalization (incorrect) and without nasalization (correct). The frequency of the candidates corresponded to the number of times each option was available for the learner during training, e.g. for the [a]-learners [a]-plural with nasalization was presented 32 times and [a]-diminutive without nasalization was presented 32 times, no other options were available for an [a]-learner during training. We created three such files to simulate our learning experiment: one represented the [a]-learners, one represented the [ɛ]-learners, and one represented the [i]-learners. The files also included the *MAP constraints and their violations.

In a second file we specified our constraints, their weights and their standard deviation. To test which standard deviation leads to the best fit with the results of our learning experiment we provided the program with various values for the standard deviation. They varied between 0.1 and 10.0.

We provided the MaxEnt Grammar Tool with each combination of our three input-candidate-files together with our constraint-files. The program used these files to predict the probabilities for each input-candidate pair. We received the predicted probabilities for each response possibility. The predicted probabilities for the two available responses of one input form summed to 1.0. This allowed us to compare the predicted probabilities for each available response with the real proportions of answers given in the learning experiment (see figure 11). This comparison showed a match between the proportions of correct answers by the participants and the proportions of predicted probabilities of the grammar. For example, the proportions of correct answers for [i]-learners is highest, unsurprisingly, for [i]-items and lowest for [a]-items. This is the case for both the answers of the participants (orange bars) and the predicted probabilities of the grammar (gray bars). This is true for all proportions and predicted probabilities.[1]

[1] The grammar predicts a higher proportion of correct answers than the participants, which might indicate that participants do not solely rely on perception but are, to a degree greater than the computational grammar, affected by noise.

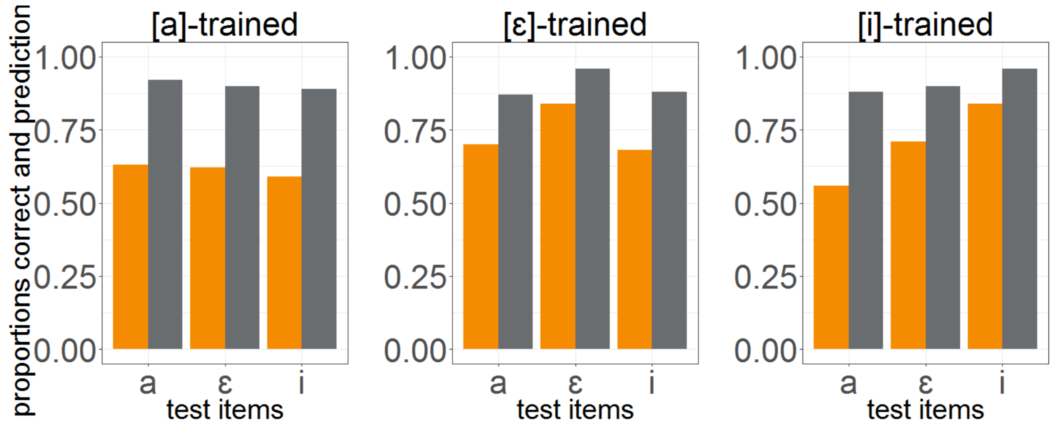

Figure 11: Comparison of the results of the learning experiment (orange bars) and the predictions of a biased MaxEnt grammar (gray bars).

We also created an unbiased grammar. In the unbiased grammar we set all weights to zero, and calculated the predicted probabilities of this grammar. The results are shown in figure 12. An unbiased grammar results in predicted probabilities that do not match the performance of our participants. In the trained conditions the grammar predicts almost always the correct answer, but, crucially, in the untrained conditions the grammar is at chance level. In other words, an unbiased grammar does not allow a participant to generalize to novel forms.

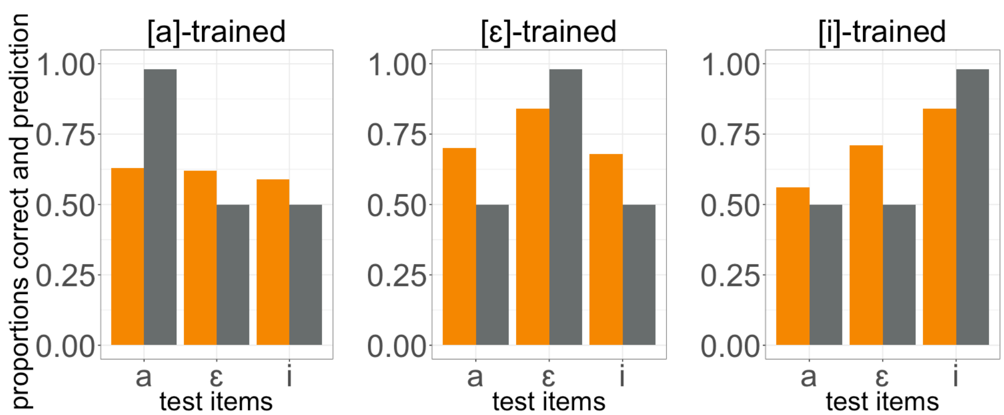

Figure 12: Comparison of the results of the learning experiment (orange bars) and the predictions of an unbiased MaxEnt grammar (gray bars).

5.2 *Discussion*

We constructed a biased grammar by means of a MaxEnt grammar model so as to investigate whether the grammar, which our participants used in the learning experiment, was biased by perception. We found that the biased grammar predicts the proportions of choices by our participants in the learning experiment, and we therefore conclude that their learning is shaped by a perception-based bias. This bias, which is based on the perceptual difficulties in vowels of different heights, explains the generalizations of the participants in the learning experiment. A grammar without bias does not predict the results of our learning experiment.

6 General discussion

The aim of this study was to find evidence for a substantive bias in phonology, and to address methodological issues raised by Moreton & Pater (2012a;b). To this end we investigated the production, perception, and learnability of vowel nasalization of high, mid, and low vowels by native speakers of Northern German. These vowels are of equal structural complexity, but their production and perception is of different difficulty. The production experiment indicated that Northern Germans slightly nasalize non-high vowels in the context of nasal consonants, but they do not nasalize high vowels in this context. This asymmetrical pattern seems to be language-independent: the anatomy of the vocal tract. The perception experiment showed that Northern Germans identify oral and nasal vowels differently for different heights – with a better identification in non-low vowels. This asymmetrical pattern seems to be language-independent: the acoustics of nasal and oral vowels. As oral and nasal non-low vowels are acoustically more different from each other than oral and nasal low vowels the former are less often confused with each other. The learning experiment showed that this different degree of confusability influences learning a nasalization pattern; the different difficulty in production does not influence learning. Participants performed better in the test phase when they had been trained with non-low vowels than when they had been trained with low vowels. This finding shows that perception acts as a prior in phonological learning: There are more generalizations to more perceptually grounded patterns than to less perceptually grounded patterns.

In our study of vowel nasalization we measured perceptual ease as confusability. Non-low oral and nasal vowels were less often confused with each other than low oral and nasal vowels. We therefore concluded that nasalization is easier to perceive in non-low vowels, and this leads to a learning advantage for a nasalization pattern in non-low vowels. The learners of non-low vowel nasalization have an easier task figuring out what the pattern is. The results of the learning experiment were then compared with the predictions of a MaxEnt grammar. We used the confusion data from the perceptual identification task to construct a grammar with a perceptual bias. More specifically, we constructed three different biased grammars – one representing the [a]-learners' grammar, one representing the [ɛ]-learners' grammar, and one representing the [i]-learners' grammar – and confronted these with the same input that our participants had been exposed to during training in the learning experiment. Then we tested the grammars with the same items that our participants had been offered. The grammars' predictions of the test items match the results of the learning experiment. We interpret this as an effect of a perceptual bias. A grammar without a bias did not predict our learning results. We conclude that perception actively supports phonological learning: it affects a generalization pattern. This is in agreement with the assumptions of phonetically based phonology and the hypothesis that phonetic patterns which are either easier to produce or easier to perceive than other patterns are reflected in phonology as grammatical constraints (Hayes & Steriade 2004).

The role of phonetics in synchronic phonology is controversial. While the role of phonetics is argued to be purely diachronic by some (Blevins 2004; Ohala 1993; Yu 2004), others explain their synchronic experimental results as a consequence of a phonetic bias (Baer-Henney & van de Vijver 2012; Finley 2008; 2012; van de Vijver & Baer-Henney 2014; White 2017; White & Sundara 2014; Wilson 2006) and argue that there is a phonetic influence on synchronic phonology (Hayes & Steriade 2004). However, these studies confound a phonetic bias with structural simplicity (Moreton & Pater 2012a;b; Pater & Moreton 2012). We avoided this confound by using vowel nasalization of vowels of different heights. These vowels are structurally akin, but differ in their production and perception. This allowed us to provide new and more solid evidence for a role of phonetics in learning synchronic phonology.

We further argue that phonetics needs to be more tightly integrated into phonology as our results support the hypothesis that phonological representations must include phonetic details (Flemming 2001). According to Flemming (2001) allophonic nasalization is often described as phonetic because it is automatically achieved by lowering the velum during the production of a vowel in the context of a nasal consonant. However, the same process is found in phonology when the contrast between oral and nasal vowels is neutralized in the context of nasal consonants in many languages. Thus, phonetics and phonology should not be regarded as two independent components of our grammar. In our grammar phonetic details are stored which are used during learning phonological alternations. This enabled the learners to use phonetic details in generalizing a pattern to new items.

The question arises why our learning is more affected by perceptual ease than by articulatory ease. This might be explained with the perception-before-production hypothesis (Flege 1991). Perceptual skills are acquired first during the process of language learning, while articulatory skills are acquired later. Clarifying this question will be a matter of future research.

7 Conclusion

The present study investigated the role of phonetics in phonological learning. There are two different phonetic precursors (Moreton 2008) in the case of vowel nasalization, both depending on vowel height. Whereas the production of low nasal vowels is easiest, the perception of non-low nasal vowels is easiest. We tested Northern German native speakers who are not familiar with contrastive vowel nasalization to investigate which of the two precursors influences learning. We showed that Northern Germans produce allophonic nasalization only in non-high vowels (see section 2). In a perception experiment we found that the acoustic differences in vowels varying in height are responsible for an asymmetrical perception pattern. Whereas non-low oral and nasal vowels are easily identified, low oral and nasal vowels are difficult to identify (see section 3). The perceptual difference – not the articulatory one – plays a role when learning a nasalization pattern and generalizing it to novel vowels. In a learning experiment we found that the pattern involving non-low vowels is easily generalized, but not the pattern involving low vowels (see section 4). This mirrors the differences in the acoustics and perception. A biased MaxEnt grammar based on these perceptual differences predicts the behavior of our participants, whereas an unbiased MaxEnt grammar does not (see section 5). We thus provide support for an active role of phonetics in synchronic phonology.

References

Albright, Adam & Bruce Hayes. 2011. Learning and learnability in phonology. In John Goldsmith, Jason Riggle & Alan C. L. Yu (eds.), *The handbook of phonological theory, second edition*, 661–690. Malden, MA: Blackwell Publishing.

Auguié, Baptiste. 2016. *gridExtra: miscellaneous functions for "Grid" graphics. R package version 2.2.1*. https://CRAN.R-project.org/package=gridExtra.

Azevedo, Milton M. 2005. *Portuguese: a linguistic introduction*. Cambridge, UK: Cambridge University Press.

Baayen, R. Harald. 2008. *Analysing linguistic data (vol. 505)*. Cambridge, UK: Cambridge University Press.

Baayen, R. Harald. 2013. LanguageR: data sets and functions with "Analyzing linguistic data: a practical introduction to statistics". R package version 1.4.1. http://CRAN.R-project.org/package=languageR.

Baer-Henney, Dinah. 2015. *Learners' little helper*. Universität Potsdam (Doctoral dissertation).

Baer-Henney, Dinah, Frank Kügler & Ruben van de Vijver. 2015. The interaction of language-specific and universal factors during the acquisition of morphophonemic alternations with exceptions. *Cognitive Science* 39(7). 1537–1569. https://doi.org/10.1111/cogs.12209.

Baer-Henney, Dinah & Ruben van de Vijver. 2012. On the role of substance, locality, and amount of exposure in the acquisition of morphophonemic alternations. *Laboratory Phonology* 3(2). 221–249. https://doi.org/10.1515/lp-2012-0013.

Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. *lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-8*. http://CRAN.Rproject.org/package=lme4..

Bell-Berti, Fredericka. 1993. Understanding velic motor control: studies of segmental context. In Marie K. Huffman & Rena A. Krakow (eds.), *Nasals, nasalization, and the velum. phonetics and phonology, volume 5*, 63–85. San Diego: Academic Press.

Blevins, Juliette. 2004. *Evolutionary phonology: the emergence of sound patterns*. Cambridge, UK: Cambridge University Press.

Boersma, Paul & David Weenink. 2017. *Praat: doing phonetics by computer (version 6.0.36) [computer program]*. http://www.praat.org/.

Bond, Zinny S. 1976. Identification of vowels excerpted from neutral and nasal contexts. *Journal of the Acoustical Society of America* 59(5). 1229–1232. https://doi.org/10.1121/1.380988.

Chen, Marylin Y. 1995. Acoustic parameters of nasalized vowels in hearing-impaired and normal hearing speakers. *Journal of the Acoustical Society of America* 98(5). 2443–2453. https://doi.org/10.1121/1.414399.

Chen, Marylin Y. 1997. Acoustic correlates of english and french nasalized vowels. *Journal of the Acoustical Society of America* 102(4). 2360–2370. https://doi.org/10.1121/1.419620.

Cristià, Alejandrina & Amanda Seidl. 2008. Is infants' learning of sound patterns constrained by phonological features? *Language Learning and Development* 4(3). 203–227. https://doi.org/10.1080/15475440802143109.

Delattre, Pierre. 1954. Les attributs acoustiques de la nasalité vocalique et consonantique. *Studia Linguistica* 8(2). 103–109. https://doi.org/10.1111/j.1467-9582.1954.tb00507.x.

Delvaux, Véronique. 2009. Perception du contraste de nasalité vocalique en français. *Journal of French Language Studies* 19(1). 25–59. https://doi.org/10.1017/S0959269508003566.

Fant, Gunnar. 1960. *The acoustic theory of speech production*. Den Haag: Mouton.

Finley, Sara. 2008. *Formal and cognitive restrictions on vowel harmony*. Johns Hopkins University, Baltimore, Maryland (Doctoral dissertation).

Finley, Sara. 2012. Typological asymmetries in round vowel harmony: support from artificial grammar learning. *Language and cognitive processes* 27(10). 1550–1562. https://doi.org/10.1080/01690965.2012.660168.

Flege, James Emil. 1991. Perception and production: the relevance of phonetic input to l2 phonological learning. In Thom Hübner & Charles A. Ferguson (eds.), *Cross currents in second language acquisition and linguistic theory*, 249–290. Amsterdam: John Benjamins.

Flemming, Edward. 2001. Scalar and categorical phenomena in a unified model of phonetics and phonology. *Phonology* 18(1). 7–44.

Goldwater, Sharon & Mark Johnson. 2003. Learning ot constraint rankings using a maximum entropy model. In *Proceedings of the stockholm workshop on variation within optimality theory*, 111–120.

Guion, Susan. 1996. *Velar palatalization: coarticulation, perception, and sound change*. University of Texas at Austin (Doctoral dissertation).

Guion, Susan. 1998. The role of perception in the sound change of velar palatalization. *Phonetica* 55(1-2). 18–52.

Hajek, John. 1997. *Universals of sound change in nasalization*. Malden, MA: Blackwell Publishing.

Hayes, Bruce. 2009a. *Introductory phonology*. Malden, MA: Wiley-Blackwell.

Hayes, Bruce. 2009b. *Manual for maxent grammar tool*. Los Angeles, CA: UCLA.

Hayes, Bruce & Donca Steriade. 2004. Introduction: the phonetic bases of phonological markedness. In Bruce Hayes, Donca Steriade & Robert M. Kirchner (eds.), *Phonetically based phonology*, 1–33. Cambridge, UK: Cambridge University Press.

Hayes, Bruce & Colin Wilson. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39(3). 379–440. https://doi.org/10.1162/ling.2008.39.3.379.

Hope, Ryan M. 2013. *Rmisc: Ryan Miscellaneous. R package version 1.5*. https://CRAN.R-project.org/package=Rmisc.

House, Arthur S. & Kenneth N. Stevens. 1956. Analog studies of the nasalization of vowels. *Journal of Speech and Hearing Disorders* 21(2). 218–232. https://doi.org/10.1044/jshd.2102.218.

Jäger, Gerhard. 2004. *Maximum entropy models and stochastic Optimality Theory. Rutgers Optimality Archive 625*.

Keating, Patricia, Marc Garellek & Jody Kreiman. 2015. Acoustic properties of different kinds of creaky voice. In The Scottish Consortium for ICPhS 2015 (ed.), *Proceedings of the 18th international congress of phonetic sciences*, 821. Glasgow, UK: the University of Glasgow.

Kingston, John & Randy L. Diehl. 1994. Phonetic knowledge. *Language* 70. 419–454. https://doi.org/10.2307/416481.

Kingston, John & Neil A. Macmillan. 1995. Integrality of nasalization and F1 in vowels in isolation and before oral and nasal consonants: a detection-theoretic application of the Garner paradigm. *Journal of the Acoustical Society of America* 97(2). 1261–1285. https://doi.org/10.1121/1.412169.

Knoblauch, Kenneth. 2014. psyphy: Functions for analyzing psychophysical data in R. R package version 0.1-9. https://cran.r-project.org/package=psyphy.

Kuznetsova, Alexandra, Per Bruun Brockhoff & Rune Haube Bojesen Christensen. 2018. lmertest: tests in linear mixed effects models. R package version 3.1-1. https://cran.r-project.org/package=lmerTest.

Laeufer, Christiane. 2010. Nasal vowels in french loanwords in german: the effect of linguistic environment. *Folia Linguistica* 44(1). 53–101. https://doi.org/10.1515/flin.2010.003.

Macmillan, Neil A., John Kingston, Rachel Thorburn, Laura Walsh Dickley & Christine Bartels. 1999. Integrality of nasalization and F1. II. basic sensitivity and phonetic labeling measure distant sensory and decision-rule interactions. *Journal of the Acoustical Society of America* 106(5). 2913–2932. https://doi.org/10.1121/1.428113.

McCarthy, John J. & Alan S. Prince. 1995. Faithfulness and reduplicative identity. In Jill Beckman, Laura Walsh Dickey & Suzanne Urbancyck (eds.), *University of Massachusetts Occasional Papers in Linguistics [UMOP] 18: Papers in Optimality Theory*, 249–384. University of Massachusetts at Amherst: GSLA.

Mermelstein, Paul. 1977. On detecting nasals in continuous speech. *Journal of the Acoustical Society of America* 61(2). 581–587. https://doi.org/10.1121/1.381301.

Moreton, Elliott. 2008. Analytic bias and phonological typology. *Phonology* 25(1). 83–127.

Moreton, Elliott & Joe Pater. 2012a. Structure and substance in artificial phonology learning, part I: structure. *Language and Linguistics Compass* 6(11). 686–701. https://doi.org/10.1002/lnc3.363.

Moreton, Elliott & Joe Pater. 2012b. Structure and substance in artificial phonology learning, part II: substance. *Language and Linguistics Compass* 6(11). 702–718. https://doi.org/10.1002/lnc3.366.

Ohala, John J. 1975. Phonetic explanations for nasal sound patterns. In Charles A. Ferguson, Larry M. Hyman & John J. Ohala (eds.), *Nasálfest: papers from a symposium on nasals and nasalization,* 289–316. Stanford: Stanford University, Linguistics Department.

Ohala, John J. 1993. Sound change as nature's speech perception experiment. *Speech Communication* 13(1). 155–161. https://doi.org/10.1016/0167-6393(93)90067-U.

Pater, Joe & Elliott Moreton. 2012. Structurally biased phonology:complexity in learning and typology. *Journal of the English and Foreign Languages University, Hyderabad* 3(2). 1–44.

Peirce, Jonathan W. 2007. PsychoPy - psychophysics software in Python. *Journal of neuroscience methods* 162(1). 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017.

Peperkamp, Sharon, Katrin Skoruppa & Emmanuel Dupoux. 2006. The role of phonetic naturalness in phonological rule acquisition. In David Bamman, Tatiana Magnitskaia & Colleen Zaller (eds.), *Proceedings of the 30th annual boston university conference on language development*, vol. 2, 464–475. Somerville, MA: Cascadilla Press.

Pruthi, Tarun & Carol Epsy-Wilson. 2004. Acoustic parameters for automatic detection of nasal manner. *Speech Communication* 43(3). 225–239. https://doi.org/10.1016/j.specom.2004.06.001.

R Core Team. 2015. *R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria*. http://www.R-project.org.

Sarkar, Deepayan. 2008. *Lattice: multivariate data visualization with R. New York: Springer*. http://lmdvr.r-forge.r-project.org.

Schwartz, Martin F. 1968. The acoustics of normal and nasal vowel production. *Cleft Palate Journal* 5. 125–140. https://cleftpalatejournal.pitt.edu/ojs/cleftpalate/article/view/187.

Snodgrass, Joan G. & Mary Vanderwart. 1980. A standardised set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of experimental psychology: Human learning and memory* 6(2). 174–215. https://doi.org/10.1037/0278-7393.6.2.174.

Steriade, Donca. 2001. The phonology of perceptibility effects: the p-map and its consequences for constraint organization. *Ms., UCLA*.

Steriade, Donca. 2009. The phonology of perceptibility effects: the p-map and its consequences for constraint organization. In Kristin Hanson & Sharon Inkelas (eds.), *The nature of the word*, 150–178. Cambridge, MA: The MIT Press. https://doi.org/10.7551/mitpress/9780262083799.003.0007.

Stevens, Kenneth N. 1998. *Acoustic phonetics*. Cambridge, MA: MIT Press.

Styler, Will. 2015. *On the acoustical and perceptual features of vowel nasality*. University of Colorado at Boulder (Doctoral dissertation).

Styler, Will. 2017. On the acoustical features of vowel nasality in english and french. *The Journal of the Acoustical Society of America* 142(4). 2469–2482. https://doi.org/10.1121/1.5008854.

Styler, Will & Rebecca Scarborough. 2017. *Nasality automeasure script package*. https://github.com/stylerw/styler%7B%5C_%7Dpraat%7B%5C_%7Dscripts.

van de Vijver, Ruben & Dinah Baer-Henney. 2014. Developing biases. *Frontiers in Psychology* 5(634). https://doi.org/10.3389/fpsyg.2014.00634.

Westbury, John R. & Patricia A. Keating. 1986. On the naturalness of stop consonant voicing. *JL* 22(1). 145–166. https://doi.org/10.1017/S0022226700010598.

White, James. 2013. *Bias in phonological learning: evidence from saltation*. University of California, Los Angeles (Doctoral dissertation).

White, James. 2014. Evidence for a learning bias against saltatory phonological alternations. *Cognition* 130(1). 96–115. https://doi.org/10.1016/j.cognition.2013.09.008.

White, James. 2017. Accounting for the learnability of saltation in phonological theory: a maximum entropy model with a p-map bias. *Language* 1(93). 1–36. https://doi.org/10.1353/lan.2017.0001.

White, James & Megha Sundara. 2014. Biased generalization of newly learned phonological alternations by 12-month-old infants. *Cognition* 133(1). 85–90. https://doi.org/10.1016/j.cognition.2014.05.020.

Wickham, Hadley. 2009. *ggplot2: elegant graphics for data analysis. New York: Springer*. New York. http://ggplot2.org.

Wickham, Hadley. 2011. The split-apply-combine strategy for data analysis. *Journal of Statistical Software* 40(1). 1–29. http://www.jstatsoft.org/v40/i01/.

Wiese, Richard. 1996. *The phonology of german*. Oxford: Oxford University Press.

Wilson, Colin. 2003. Experimental investigation of phonological naturalness. In Gina Garding & Mimu Tsujimura (eds.), *Wccfl 22*, vol. 22, 533–546. Somerville, MA: Cascadilla Press.

Wilson, Colin. 2006. Learning phonology with substantive bias: an experimental and computational study of velar palatalization. *Cognitive Science: A Multidisciplinary Journal* 30(5). 945–982. https://doi.org/10.1207/s15516709cog0000_89.

Wilson, Colin & Ben George. 2009. *Maxent grammar tool [software]*. http://www.linguistics.ucla.edu/people/hayes/MaxentGrammarTool.

Yu, Alan C. L. 2004. Explaining final obstruent voicing in Lezgian: phonetics and history. *Language* 80(1). 73–97. https://doi.org/10.1353/lan.2004.0049.

Zhang, Hong. 2015. *Production and acoustics of creaky nasal vowels*. University of Colorado at Boulder (Doctoral dissertation).

Zuraw, Kie. 2007. The role of phonetic knowledge in phonological patterning: corpus and survey evidence from Tagalog infixation. *Language* 83(2). 277–316. https://doi.org/10.1353/lan.2007.0105.

8 Appendix

Table 6: Stimuli for the production experiment: IPA transcription and German orthography.

IPA transcription			German orthography		
CV[m]V	**CVCV**	**CVCV[m]**	**CV[m]V**	**CVCV**	**CVCV[m]**
duma	duba	dubam	duma	duba	dubam
dumɛ	dubɛ	dubɛm	dumä	dubä	dubäm
dumi	dubi	dubim	dumi	dubi	dubim
dumo	dubo	dubom	dumo	dubo	dubom
dumu	dubu	dubum	dumu	dubu	dubum
kuma	kufa	kufam	kuma	kufa	kufam
kumɛ	kufɛ	kufɛm	kumä	kufä	kufäm
kumi	kufi	kufim	kumi	kufi	kufim
kumo	kufo	kufom	kumo	kufo	kufom
kumu	kufu	kufum	kumu	kufu	kufum
poma	poga	pogam	poma	poga	pogam
pomɛ	pogɛ	pogɛm	pomä	pogä	pogäm
pomi	pogi	pogim	pomi	pogi	pogim
pomo	pogo	pogom	pomo	pogo	pogom
pomu	pogu	pogum	pomu	pogu	pogum
ʃoma	ʃota	ʃotam	schoma	schota	schotam
ʃomɛ	ʃotɛ	ʃotɛm	schomä	schotä	schotäm
ʃomi	ʃoti	ʃotim	schomi	schoti	schotim
ʃomo	ʃoto	ʃotom	schomo	schoto	schotom
ʃomu	ʃotu	ʃotum	schomu	schotu	schotum
vuma	vuza	vuzam	wuma	wusa	wusam
vumɛ	vuzɛ	vuzɛm	wumä	wusä	wusäm
vumi	vuzi	vuzim	wumi	wusi	wusim
vumo	vuzo	vuzom	wumo	wuso	wusom
vumu	vuzu	vuzum	wumu	wusu	wusum

Table 7: Stimuli in the training phase of the learning experiment experiment (three different groups): singular forms.

group [a]	group [ɛ]	group [i]
doba	dobɛ	dobi
poba	pobɛ	pobi
ʃuba	ʃubɛ	ʃubi
dofa	dofɛ	dofi
kufa	kufɛ	kufi
ʃofa	ʃofɛ	ʃofi
doga	dogɛ	dogi
kuga	kugɛ	kugi
voga	vogɛ	vogi
duta	dutɛ	duti
pota	potɛ	poti
ʃuta	ʃutɛ	ʃuti
doza	dozɛ	dozi
kuza	kuzɛ	kuzi
ʃoza	ʃozɛ	ʃozi
vuza	vuzɛ	vuzi

Table 8: Stimuli in the training phase of the learning experiment (three different groups): plural forms.

group [a]	group [ɛ]	group [i]
dobãm	dobɛ̃m	dobĩm
pobãm	pobɛ̃m	pobĩm
ʃubãm	ʃubɛ̃m	ʃubĩm
dofãm	dofɛ̃m	dofĩm
kufãm	kufɛ̃m	kufĩm
ʃofãm	ʃofɛ̃m	ʃofĩm
dogãm	dogɛ̃m	dogĩm
kugãm	kugɛ̃m	kugĩm
vogãm	vogɛ̃m	vogĩm
dutãm	dutɛ̃m	dutĩm
potãm	potɛ̃m	potĩm
ʃutãm	ʃutɛ̃m	ʃutĩm
dozãm	dozɛ̃m	dozĩm
kuzãm	kuzɛ̃m	kuzĩm
ʃozãm	ʃozɛ̃m	ʃozĩm
vuzãm	vuzɛ̃m	vuzĩm

Table 9: Stimuli in the training phase of the learning experiment (three different groups): diminutive forms.

group [a]	**group [ɛ]**	**group [i]**
dobal	dobɛl	dobil
pobal	pobɛl	pobil
ʃubal	ʃubɛl	ʃubil
dofal	dofɛl	dofil
kufal	kufɛl	kufil
ʃofal	ʃofɛl	ʃofil
dogal	dogɛl	dogil
kugal	kugɛl	kugil
vogal	vogɛl	vogil
dutal	dutɛl	dutil
potal	potɛl	potil
ʃutal	ʃutɛl	ʃutil
dozal	dozɛl	dozil
kuzal	kuzɛl	kuzil
ʃozal	ʃozɛl	ʃozil
vuzal	vuzɛl	vuzil

Table 10: Stimuli in the test phase of the learning experiment (identical for every group): correct form ∼ incorrect form.

V_2 = [a]	V_2 = [ɛ]	V_2 = [i]
potãm ∼ potam	dobɛ̃m ∼ dobɛm	ʃofĩm ∼ ʃofim
dofãm ∼ dofam	dozɛ̃m ∼ dozɛm	dogĩm ∼ dogim
kugãm ∼ kugam	kufɛ̃m ∼ kufɛm	dutĩm ∼ dutim
ʃubãm ∼ ʃubam	ʃutɛ̃m ∼ ʃutɛm	kuzĩm ∼ kuzim
kogãm ∼ kogam	vofɛ̃m ∼ vofɛm	vobĩm ∼ vobim
ʃotãm ∼ ʃotam	ʃobɛ̃m ∼ ʃobɛm	dotĩm ∼ dotim
dubãm ∼ dubam	vubɛ̃m ∼ vubɛm	vufĩm ∼ vufim
pufãm ∼ pufam	kutɛ̃m ∼ kutɛm	puzĩm ∼ puzim
dobal ∼ dobãl	pobɛl ∼ pobɛ̃l	dofil ∼ dofĩl
dogal ∼ dogãl	ʃozɛl ∼ ʃozɛ̃l	vogil ∼ vogĩl
kufal ∼ kufãl	dutɛl ∼ dutɛ̃l	ʃubil ∼ ʃubĩl
ʃutal ∼ ʃutãl	kuzɛl ∼ kuzɛ̃l	kugil ∼ kugĩl
vozal ∼ vozãl	kobɛl ∼ kobɛ̃l	ʃogil ∼ ʃogĩl
kotal ∼ kotãl	pofɛl ∼ pofɛ̃l	kozil ∼ kozĩl
pubal ∼ pubãl	dugɛl ∼ dugɛ̃l	dufil ∼ dufĩl
vutal ∼ vutãl	ʃuzɛl ∼ ʃuzɛ̃l	putil ∼ putĩl