From perception to phonology: The emergence of perceptually motivated constraint rankings

Sverre Stausland Johnsen

Department of Linguistics, Boylston Hall, 3rd Floor, Harvard University, Cambridge, MA 02138, USA

A B S T R A C T

A phonological process in Norwegian takes underlying alveolar /t d n s/ to retroflex [ʈ ɖ ɳ ʂ] after /r/, but the likelihood of its application depends on the phonological context: The process is obligatory for /t d n/, but optional for /s/, and the likelihood of /s/ undergoing retroflexion to [ʂ] depends on the following segment. Perceptual experiments with Norwegian listeners show that the likelihood of retroflexion correlates with its perceptual properties. The greater the perceived distance is between an alveolar and a retroflex in a given context, the less likely it is that the alveolar will undergo retroflexion. In a constraint grammar like optimality theory, this suggests that the internal ranking of the constraints relevant for retroflexion is determined by perceptual properties. Rather than assuming that perceptual properties and constraint rankings are linked in the grammar by inherent design, I demonstrate in this paper how such constraint rankings can emerge from the phonological learning process as a result of how perceptual distances influence word categorization.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

The main object of phonological research is to study phonological processes. One essential observation of such processes is that they tend to apply more often in some linguistic contexts than others, and it is typically this very fact that phonological research attempts to account for. To name but a few examples, we know that obstruents tend to devoice in final position more often than elsewhere, that back vowels are more often fronted before front vowels than in other positions, that vowels are reduced in unstressed syllables more often than in stressed syllables, that voiceless stops are more likely to affricate than voiced stops, and so on, and phonologists aim to uncover why such processes follow exactly these patterns.

Norwegian retroflexion, the phonological process discussed in this paper, is no different in this regard. The process takes underlying alveolar /t d n s/ to retroflex [ʈ ɖ ɳ ʂ] in the position after /r/, yet its application is sensitive to the linguistic context. Retroflexion is applied less frequently to /s/ than to /t d n/, and retroflexion of /s/ is more common before /k/ than before /t/, and more common before consonants than before vowels. Why should retroflexion follow exactly this pattern? The account given in this paper is that the pattern reflects perceptual properties. In short, the finding is that the more perceptually similar the alveolar is to the retroflex in a given context, the more likely retroflexion is to apply in that context.

That perception plays a role in phonology has always been recognized, yet it was traditionally invoked as an explanation only when no other explanation seemed at all plausible, as in loanword phonology (Sievers, 1901:271) or with sound changes that incurred great articulatory leaps (Paul, 1920:63). With the advent of evolutionary phonology (Ohala, 1981; Blevins, 2004) and phonetically based phonology (Hayes et al., 2004), perception has been lifted from being an explanation in

E-mail address: johnsen@fas.harvard.edu.

case of emergency to occupying a pivotal role in explaining phonological patterns. Yet despite this paradigm shift, perceptually motivated accounts are typically given only to explain cross-linguistic typology, loanword phonology, and diachronic sound changes. Attempts at seeking perceptual explanations for patterns of synchronic alternation within the native phonology of a language have been relatively few (Huang, 2001; Seo, 2001; Mielke, 2003a). This paper demonstrates that the pattern of synchronic variation in Norwegian retroflexion is directly correlated with the perceptual properties of retroflexion, which therefore indicates that perception must play a crucial role in accounting for the existence of this pattern.

More specifically, the experimental data reported in this paper shows that the likelihood of retroflexion being applied to an alveolar in Norwegian correlates with the perceived similarity between the alveolar and the retroflex. As such, it provides strong support for the theory in Steriade (2001:222), who proposes that the likelihood of a phonological mapping between an underlying representation x and a modified output x' is a function of the perceived similarity between x and x'. If this view is correct, then it is possible to formalize this principle within optimality theory (Prince and Smolensky, 2004) in the following way: If the perceived distance between x and x' is greater than the perceived distance between y and y', then the correspondence constraints referring to the distinction between x and x' will outrank those referring to the distinction between y and y' (Steriade, 2001:239; Steriade, 2009:164), with the effect that x is protected from surfacing as x' more than y is protected from surfacing as y', as in FAITH $(x) \gg$ FAITH (y). The remaining question is why there is such a correlation between perceptual properties and grammatical constraint rankings in the first place. According to Steriade (2001:239, 2009:164) and followers, there is an inherent mechanism of the grammar that translates facts about perceived distances directly into constraint rankings. Yet this account does not explain where the correlation comes from, rather it assumes it is there from the outset. I propose in this paper that the link between perceived distances and phonological patterns is an emergent rather than inherent property of grammar. As a fact of perceptual categorization, the greater the perceived distance is between a perceived form x' and a category x, the less likely people are to categorize x' with x. Given similar findings in linguistic experiments, I suggest that the link between perceived distances and constraint rankings is a result of how word tokens were categorized during the learning process when constraint rankings were deduced. In short, the greater the perceived distance is between x and x', the less likely it is that the learner will categorize x' as a token of x, and the less likely it is that the learner will construct a grammar in which x surfaces as x'.

Section 2 outlines the patterns of retroflexion in Norwegian, and it is proposed in sections 3 and 4 that these patterns are best accounted for with reference to the perceptual properties of retroflexion. Experiments testing this hypothesis are reported in sections 5–7, and summarized in section 8. The idea that perceptual properties and constraint rankings are linked as a result of word categorization is outlined in sections 9–11, and this proposal is put to test in a learning simulation in section 12. A discussion of this simulation follows in section 13, before the paper is summarized in section 14.

2. Norwegian retroflexion

2.1. Norwegian

The Norwegian language is generally a cover term for the North Germanic dialects traditionally spoken within the kingdom of Norway, currently with about 4.5 million speakers. The only variant of Norwegian which will be treated in this paper is the spoken variety currently used by most speakers in urbanized areas of South-East Norway. The phonological properties of this spoken standard are extensively treated in Kristoffersen (2000).

2.2. Norwegian retroflexes

Norwegian distinguishes two sets of coronals in postvocalic position:

(1) | Name | Transcription | Articulation[1] |
|---|---|---|
| Alveolar | /t d n s/ | Laminal alveolar coronal |
| Retroflex | /ʈ ɖ ɳ ʂ/ | Apical postalveolar coronal |

The contrast between alveolars and retroflexes in postvocalic position is illustrated in (2):

(2) | /kɑt/ | 'cat' | /kɑʈ/ | 'unripe fruit' |
|---|---|---|---|
| /bɔːd/ | 'boring' | /bɔːɖ/ | 'a man's name' |
| /tʉːn/ | 'yard' | /tʉːɳ/ | 'gymnastics' |
| /kɔs/ | 'heap' | /kɔʂ/ | 'cross' |

---

[1] For articulatory studies of these segments, see Simonsen and Moen (2004) and Simonsen et al. (2008).

2.2.1. Phonological representation of Norwegian retroflexes

The retroflex /ʈ ɖ ɳ ʂ/ originate from clusters /ɾt ɾd ɾn ɾs/ in older stages of the language. To mention a couple of examples, Norwegian /kɑʈ/ 'unripe fruit' comes from Old Norwegian *kart-*, and Norwegian /kɔʂ/ 'cross' is an old loanword from Danish *kors*. The change from such clusters to retroflex coronals never took place in the closely related Danish language, whose orthographic system forms the basis for Norwegian spelling conventions. As a consequence, retroflex /ʈ ɖ ɳ ʂ/ are graphically represented in modern Norwegian written standards as consonant clusters ‹rt rd rn rs›. Norwegian /kɑʈ/ 'unripe fruit' and /kɔʂ/ 'cross' are therefore spelled ‹kart› and ‹kors› respectively.

Given the historical origin of Norwegian retroflexes, it has been proposed that these retroflex consonants are represented phonologically in speakers' grammar as consonant clusters /ɾt ɾd ɾn ɾs/, and that a transformational rule changes these clusters into surfacing retroflexes (Fretheim, 1969:89f.; Hovdhaugen, 1969:147; Endresen, 1974:75; Standwell, 1975:344ff.). The predominant view among Norwegian linguists has nevertheless been that retroflexes are represented as retroflexes at all levels of representation (Borgstrøm, 1938:255; Vogt, 1939; Rinnan, 1969; Vanvik, 1972:147f.; Kristoffersen, 2000:88f.; Simonsen et al., 2008:387f.). The main argument for the latter approach is that retroflex /ʈ ɖ ɳ ʂ/ occasionally contrast with /ɾt ɾd ɾn ɾs/ on the surface, as in /mɔɳ/ 'morning' – /nɔɾn/ 'Norn', /faʂə/ 'farce' – /fɑɾsi/ 'Farsi', and /fæɖi/ 'done' – /væɾdi/ 'worthy'. For the sake of consistency, I will follow the dominant view in this paper, and I will therefore take all morpheme internal retroflexes to be underlyingly retroflex. It is important to point out, however, that nothing hinges on this assumption.

2.3. Norwegian retroflexion

2.3.1. Deletion of morpheme final /ɾ/

A morpheme final apical alveolar tap /ɾ/ deletes when the following morpheme begins with a consonant (Rykkvin, 1946; Haugen, 1948; Kristoffersen, 2000:311ff.)[2]:

(3)	/vintər-føːrə/	→	[vintəføːrə]	'winter condition'
	/vintər-jakə/	→	[vintəjakə]	'winter coat'
	/vintər-kʉlə/	→	[vintəkʉlə]	'winter cold'

2.3.2. Retroflexion of morpheme initial alveolars

When a morpheme beginning with an alveolar /t d n s/ follows a morpheme ending in the tap /ɾ/, the tap deletes (3), and the alveolar surfaces as a retroflex [ʈ ɖ ɳ ʂ][3]:

(4)	/vintər-tiː/	→	[vintəʈiː]	'winter time'
	/vintər-dɑː/	→	[vintəɖɑː]	'winter day'
	/vintər-nat/	→	[vintəɳat]	'winter night'
	/vintər-sœvn/	→	[vintəʂœvn]	'winter sleep'

2.4. Variation in Norwegian retroflexion

According to previous descriptions in the literature, the retroflexion process in (4) is obligatory, absent only when there is a significant intonational or pausal boundary between the morphemes (Eliasson, 1986:282; Kristoffersen, 2000:316f.; Torp, 2007:70). As claimed by Kristoffersen, the retroflexion process "seems to be beyond speakers' active control" (2000:317). This description is by and large correct, as retroflexion is indeed obligatory when a morpheme beginning in alveolar /t d n/ follows a morpheme ending in the tap /ɾ/:

(5)	/vintər-tiː/	→	[vintəʈiː]	* [vintətiː]	'winter time'
	/vintər-dɑː/	→	[vintəɖɑː]	* [vintədɑː]	'winter day'
	/vintər-nat/	→	[vintəɳat]	* [vintənat]	'winter night'

However, this is decidedly not the case when the morpheme begins with an alveolar sibilant /s/. In these cases, the retroflexion process is optional:

(6)	/vintər-sœvn/	→	[vintəʂœvn] ~ [vintəsœvn]	'winter sleep'

---

[2] For articulatory studies of the tap, see Foldvik (1977) and Moen et al. (2003).
[3] This retroflexion process was first described by Brekke (1881:18ff.), Storm (1884:96f.), and Western (1889:275). For a recent treatment, see Kristoffersen (2000:96f).
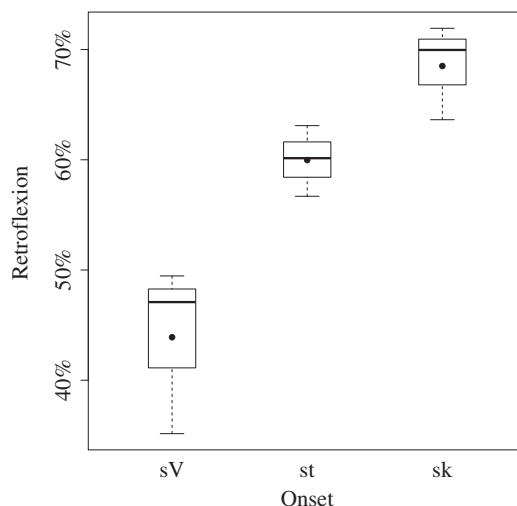
Fig. 1. Retroflexion for nonce words in /s-/ (from Stausland Johnsen, 2011).

Just how likely the retroflexion process is for a given word in /s-/ is in part determined by the segment following the /s/. When this following segment is a consonant, retroflexion to [ʂ] is generally preferred, but when the /s/ is followed by a vowel, retroflexion to [ʂ] is generally not preferred. In (7) below, the preferred output is marked with a smiley face ☺:

(7)   /ʋintər-skuː/   →   ☺ [ʋintəʂkuː] ∼ ☻ [ʋintəskuː]   'winter shoes'
         /ʋintər-suːɽ/   →   ☻ [ʋintəʂuːɽ] ∼ ☺ [ʋintəsuːɽ]   'winter sun'

The productivity of the tendency in (7) is put to test in Stausland Johnsen (2011), where ten Norwegian speakers produced nonce words in /sC-/ and /sV-/ in a retroflexing environment. The nonce words in /sC-/ had the onsets /st-/ and /sk-/, which are the two most frequent /sC-/ onsets in Norwegian. The results, seen in Fig. 1, show that retroflexion is applied more frequently to words in /sC-/ than to words in /sV-/, and that retroflexion of /s/ is more common before /k/ than before /t/. The differences between the three groups shown in Fig. 1 are all significant.

Based on the observations discussed in this section, we can set up a hierarchy for the probability of retroflexion of alveolar onsets:

(8)   /t/, /d/, /n/ > /sk/ > / st/ > /sV/

The hierarchy in (8) illustrates that the probability of retroflexion is the highest for /t/, /d/, /n/, followed by a lower probability of retroflexion for /sk/, and so on. What remains to be accounted for is why retroflexion should follow the pattern described in (8). Whereas the articulatory properties of retroflexion do not offer any clear motivation for this phonological pattern, the perceptual properties of retroflexion do, as will be shown in the following sections.

3. Retroflexion and articulation

The articulatory modification involved in the retroflexion process is a change in the point of contact between the tongue and the palate, as it shifts from a laminal alveolar contact for [t d n s] to an apical postalveolar contact for [ʈ ɖ ɳ ʂ]. Since this articulatory change is the same for all these alveolar consonants, it is not clear why this shift should be applied less frequently to /s/, nor is it clear why it should be applied to /s/ less often before a vowel than before a consonant.

A relatively common approach in explaining phonological patterns is to resort to the notion of 'markedness', by which some segments are universally disfavored by the grammatical system (Chomsky and Halle, 1968:402ff.). Under this approach, the less frequent retroflexion of /s/ to [ʂ] could indicate that the postalveolar sibilant [ʂ] is a 'marked' segment that the grammar strives to avoid. To what extent a segment is 'marked' is typically deduced from how commonly it is found cross-linguistically (Chomsky and Halle, 1968:413). Typological surveys of languages across the world indicate that postalveolar sibilants in fact are prevalent, whereas postalveolar stops are relatively uncommon (Maddieson, 1984). If the articulatory 'markedness' of retroflexes played any role in determining the likelihood of retroflexion in Norwegian, we would therefore expect the reverse pattern of (8), a pattern in which retroflexion of /s/ to [ʂ] should be more common than retroflexion of /t d n/ to [ʈ ɖ ɳ].

In sum, the articulatory shift from alveolar [s] to retroflex [ʂ] is the same as the shift from alveolar [t d n] to retroflex [ʈ ɖ ɳ], and to judge by typological evidence, [ʂ] is less 'marked' than [ʈ ɖ ɳ]. Neither of these two facts can in any obvious way explain the retroflexion pattern in (8). In conclusion, seeking an articulatory explanation for this pattern bears therefore little promise.

4. Retroflexion and perceived distance

4.1. Perceived distances and phonology

Although the articulatory modification from alveolars to retroflexes is the same for /t d n s/, the resulting perceptual shift need not be. In other words, the perceived distance between [t] and [ʈ] is not necessarily equivalent to the perceived distance between [s] and [ʂ], even though these relations are articulatorily equivalent. The idea which will be pursued here is that these perceived distances are indeed not the same, and that these differences are the ultimate cause of the Norwegian retroflexion pattern in (8).

As explained in section 2.4, alveolar stops and nasals consistently undergo retroflexion in Norwegian (/t d n/ → [ʈ ɖ ɳ]), whereas the alveolar fricative /s/ only optionally does (/s/ → [ʂ] ~ [s]). Kohler (1990:86ff.) finds similar data in German, where alveolar stops and nasals assimilate to a following labial (/tp/ → [pp], /nm/ → [mm]), but the alveolar fricative does not (/sf/ → *[ff]). Kohler points out that the main difference between these assimilations lies in their perceptual properties, in that the perceived distances in [tp] – [pp] and [nm] – [mm] are relatively small, whereas the perceived distinction between [sf] and [ff] is quite substantial.

Based on this and similar data, Steriade (2001:222) proposes as a general principle that the likelihood of an underlying representation x surfacing as a modified x' is a function of the perceived similarity between x and x'. The more distinct x and x' are perceptually, the less likely x is to surface as x'. As seen in (8), onsets with alveolar /s/ are less likely to surface with retroflex [ʂ] than onsets with alveolar /t d n/ are to surface with retroflex [ʈ ɖ ɳ]. Applying Steriade's principle to this pattern, it predicts that the perceived distance between alveolar [s] and retroflex [ʂ] is greater than the perceived distance between alveolar [t d n] and retroflex [ʈ ɖ ɳ].

4.2. Perceptual hypothesis of Norwegian retroflexion

When applying Steriade's principle to Norwegian retroflexion, the strong hypothesis would be that there is a direct correlation between the likelihood of retroflexion and the perceived distances between alveolars and retroflexes, as formulated in (9):

(9)   The greater the perceived distance between an alveolar and a retroflex, the less likely it is that the alveolar undergoes retroflexion.

The strong version of this hypothesis not only predicts that the perceived distance between [s] and [ʂ] is greater than between [t d n] and [ʈ ɖ ɳ], but it also predicts that the perceived distance between [s] and [ʂ] is greater before a vowel than before a consonant, and greater before the consonant /t/ than before the consonant /k/. This hypothesized direct correlation between retroflexion and perceived distances can be illustrated as in (10):

(10)

Probability of retroflexion		Perceived distance	
Increasing ↑	/t d n/		[t d n]–[ʈ ɖ ɳ]
	/sk/		[sk]–[ʂk]
	/st/	↓	[st]–[ʂʈ]
	/sV/	Increasing	[sV]–[ʂV]

In (10), the low probability of retroflexion for /s/ in the position before a vowel is predicted to correspond to a relatively large perceived distance between alveolar [s] and retroflex [ʂ] in that position. At the other end of the scale, the high probability of retroflexion for /t d n/ is predicted to correspond to a smaller perceived distance between alveolar [t d n] and retroflex [ʈ ɖ ɳ].

Perceived distances between segments are best measured from observing how well people can distinguish those segments from one another. A perceptual experiment was therefore designed to document the patterns of confusability between alveolar and retroflex coronals in Norwegian. This experiment and its results are reported in the following sections.

5. Experiment design

5.1. Task

The more similar two items are to each other, the harder they will be to distinguish. The standard procedure is therefore to measure the perceived similarity between two items as a function of their confusability (Luce, 1963:113; Macmillan and Creelman, 2005:15). Confusability of linguistic segments is typically measured with recognition experiments, where participants are asked to recognize a stimulus as *x* or *y* in a predefined stimulus set of *x* and *y*. In the perceptual recognition experiment performed in this study, the task was designed as an AX discrimination task, also called the same-different design. In such a task, the participant is presented with a stimulus pair, and is asked to decide whether the items in the pair belonged to the same or different types. The stimulus pairs are balanced between same and different types, and the perceived distance between two items is measured by how accurately participants determine whether their stimulus pairs were same or different (Macmillan and Creelman, 2005:213ff.). The AX discrimination design was chosen for this experiment because it is a relatively easy procedure for participants to understand and follow, and because discrimination tasks are claimed to allow a more direct measure of perceived similarity (Macmillan and Creelman, 2005:132ff.).

5.2. Participants

Strictly speaking, the only perceptibility scale that is relevant for Norwegian phonology is that of Norwegian listeners (cf. Mielke, 2003b:222). Additionally, it is found that speakers of languages without contrastive retroflexes sometimes perform at chance level when attempting to distinguish retroflex coronals from non-retroflex coronals (Polka, 1991:2966f.; Golestani and Zatorre, 2004:498). For these reasons, only Norwegian participants were used in the following experiments.

5.3. Stimuli

Boomershine et al. (2009) find that phonetic segments tend to be perceived as more similar to each other when they are allophones than they do when they are contrastive phonemes. As an example, Spanish listeners perceive [d] – [ð] to be more similar to each other than English listeners do, correlating with the fact that [d] and [ð] are allophones in Spanish, but contrastive phonemes in English.

It was shown in section 2.4 that morpheme initial /t d n/ systematically surface as [ʈ ɖ ɳ] when preceded by /ɾ/, whereas this is optional for morpheme initial /s/. In this environment, [t d n] – [ʈ ɖ ɳ] can therefore be analyzed as regular allophones, with [ʈ ɖ ɳ] appearing after /ɾ/, and [t d n] appearing everywhere else. Since retroflexion is optional for /s/, the pair [s] – [ʂ] could be considered as 'optional' allophones in the said environment. There is a risk in this case that Norwegian listeners will perceive [t d n] – [ʈ ɖ ɳ] in morpheme initial position as more similar than [s] – [ʂ] as a result of the former being regular allophones in this environment, while [s] – [ʂ] are 'pseudo-contrastive'. If we find [s] – [ʂ] to be perceptually more distinct in this environment than [t d n] – [ʈ ɖ ɳ], we can therefore not be entirely sure whether this *causes* [s] to alternate less with [ʂ], or if it is the *result* of [s] alternating less with [ʂ].

To avoid this potential issue, the perceived distances between [t d n s] and [ʈ ɖ ɳ ʂ] need to be measured in a position where all these segments enjoy the same status. As shown in section 2.2, Norwegian /t d n s/ and /ʈ ɖ ɳ ʂ/ are fully contrastive elements in postvocalic position within morphemes. The retroflexes /ʈ ɖ ɳ ʂ/ cannot under any circumstance surface as alveolars [t d n s] here, nor can the alveolars /t d n s/ surface as retroflexes [ʈ ɖ ɳ ʂ]. For this reason, all stimuli in this experiment have [t d n s] and [ʈ ɖ ɳ ʂ] placed between two [ɑ] vowels in morpheme internal position: [ɑCɑ].

6. Experiment A – discriminating alveolars and retroflexes

6.1. Participants

14 native speakers of Norwegian with a mean age of 25.1 participated in the experiment, nine male and five female. 12 were visiting students in the Boston area, and the remaining two participated in Norway.

6.2. Stimuli

A phonetically trained male Norwegian speaker was recorded reading multiple tokens of monomorphemic words with alveolar and retroflex coronals between two [ɑ] vowels. All words were produced with the same lexical tone. The recorded words are listed in (11)[4]:

(11)   Alveolar: [ɑtɑ], [ɑdɑ], [ɑnɑ], [ɑsɑ], [ɑskɑ], [ɑstɑ]
           Retroflex: [ɑʈɑ], [ɑɖɑ], [ɑɳɑ], [ɑʂɑ], [ɑʂkɑ], [ɑʂʈɑ]

---

[4] The retroflex variant of alveolar /st/ is a retroflex cluster /ʂʈ/, as */ʂt/ is a phonologically illicit sequence in Norwegian.

Two recorded tokens of each word were selected according to how uniform they were in speech rate, amplitude, and intonation compared with all other selected tokens. The amplitude of the selected tokens was then normalized.[5]

6.3. Procedure

The selected tokens were grouped together as pairs. The two words within each pair were either both alveolar, both retroflex, or differed in this respect. When the two words were matching, the pair consisted of two different tokens of that word. Using [ɑnɑ] – [ɑɳɑ] as an example, the participants would over the course of the experiment be exposed to the following eight combinations, with subscript 1 and 2 denoting the two recorded tokens for each word:

(12) $[ɑnɑ_1]$ – $[ɑnɑ_2]$ Same $[ɑnɑ_2]$ – $[ɑnɑ_1]$ Same
       $[ɑɳɑ_1]$ – $[ɑɳɑ_2]$ Same $[ɑɳɑ_2]$ – $[ɑɳɑ_1]$ Same
       $[ɑnɑ_1]$ – $[ɑɳɑ_2]$ Different $[ɑnɑ_2]$ – $[ɑɳɑ_1]$ Different
       $[ɑɳɑ_1]$ – $[ɑnɑ_2]$ Different $[ɑɳɑ_2]$ – $[ɑnɑ_1]$ Different

The stimulus set described in (12) will be referred to as 'Category n'. The corresponding categories for the other stimulus sets will be category 'sV', 'st', 'sk', 't', and 'd'. Each stimulus set was presented four times to each participant for every category. All stimulus pairs were randomized and masked with multi-talker babble noise.

The experiment was preceded by a brief training session without babble noise. To ensure that participants in the experiment both understood the AX task and were sensitive to the distinction between alveolar and retroflex coronals, only participants with 100% correct responses in the training session were allowed to participate. One person was excluded on this basis.[6]

6.4. Results

The experiment contained six stimulus categories ('sV', 'st', 'sk', 't', 'd', 'n'), each with eight stimulus pairs (4 same – 4 different), and each pair presented four times. With 14 participants, the total number of trials was $6 \times 8 \times 4 \times 14 = 2688$. The confusability of alveolar and retroflex consonants within a stimulus category is measured from the proportion of correct responses within that category. Since the perceived distance between two items is a function of their confusability, we can illustrate the relative differences of perceived distance between stimulus categories either directly with the proportions of correct responses (Fig. 2), or we can convert these proportions into the widely used distance parameter d' (Fig. 3).[7]

These results confirm the strong version of the hypothesis from section 4.2. Not only is the perceived distance between [s] and [ʂ] greater than between [t d n] and [ʈ ɖ ɳ], but the perceived distance between [s] and [ʂ] is greater before a vowel than before a consonant, and greater before the consonant /t/ than before the consonant /k/.

A mixed effects logistic regression model was fitted to the data in order to estimate the significance of the differences observed between these categories. The dependent variable in this model is the 'same' and 'different' responses given by the participants for each of the 2688 trials. The independent variables which will tell whether the differences in Figs. 2 and 3 are significant are **Stimulus** and **Category**.[8]

Stimulus codes whether the trial presented to a participant was a 'same' trial or a 'different' trial – see (12) for examples of such trials. The effect of **Stimulus** is therefore an estimation of how sensitive participants are to the distinction between 'same' and 'different' trials.

Category represents the different stimulus sets described in section 6.3. The stimulus set in (12), for example, represents category 'n'. The effect of **Category** is therefore an estimation of how sensitive participants are to the distinction between categories. However, since the number of 'same' and 'different' trials are the same for each category, an unbiased participant

---

[5] The tokens were recorded in mono using a KSM27 Shure microphone in a sound-attenuated booth. The signal was amplified through an M-AUDIO Firewire 410 amplifier and recorded digitally with Audacity software (Audacity Team, 2006) in .wav format at a sampling rate of 44.100Hz on an iMac7,1 Macintosh computer, then digitally converted to .aiff format using Praat software (Boersma and Weenink, 2009). The amplitude of the vowels in the selected tokens was then RMS-equalized using a modified version of the 'rms equalize' script (Beckers, 2002–2008) in Praat. Since the amplitude at consonant release is an important cue for place and manner features of consonants (Repp, 1984), no such normalization was performed for the consonants.

[6] The stimuli were played to participants over Audio-Technica ATH-A500 headphones at a fixed volume using the PsyScope X 1.2.5 B53 software (Cohen et al., 1993) on a MacBook Macintosh computer in quiet locations. The stimuli were masked with multi-talker babble taken from the Signal Processing Information Base (http://spib.ece.rice.edu) scaled at a signal-to-noise ratio of −6 dB. The stimuli were presented with an inter-stimulus interval of 2 seconds and an inter-trial interval of 1 second. Participants received instructions at the start of the training session on how to perform during the task. They were instructed to press one color marked key for stimuli belonging to the same type, and another color marked key for stimuli belonging to different types. The keys were adjacent keys on the left side of the keyboard. They were asked to use only one hand when pressing the keys, using the hand of their preference. The instructions clarified that the words in the experiment were not real words and did not need to be identified. The training session played 10 trials drawn from the pool of stimuli.

[7] The d' parameter is calculated according to the 'independent observation strategy' using the dprime.SD() function from the psyphy package (Knobluach, 2011) in R (R Development Core Team, 2011), see Macmillan and Creelman (2005:216) for details. The values for d' for each participant cannot be accurately estimated due to infinities (see section 6.5), which is why Fig. 3 has no error bars.

[8] The model was fitted with the glmer() function from the lme4 package (Bates et al., 2011) in R, with random intercepts for participants. Other independent variables in the model were **Trial** and **Order**. The latter variable codes for each AX trial whether the A and X stimuli were alveolar or retroflex, or in other words the order of alveolar and retroflex coronals within each trial.
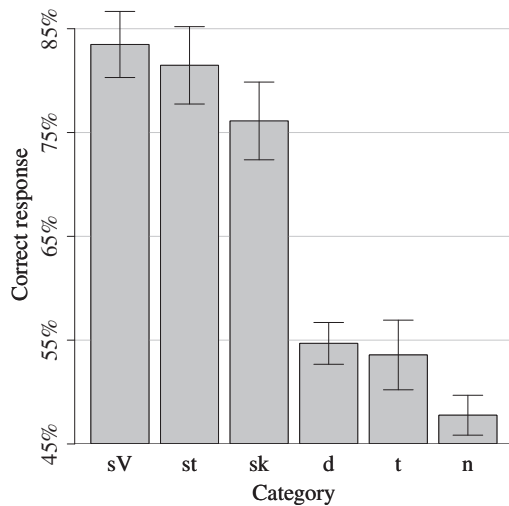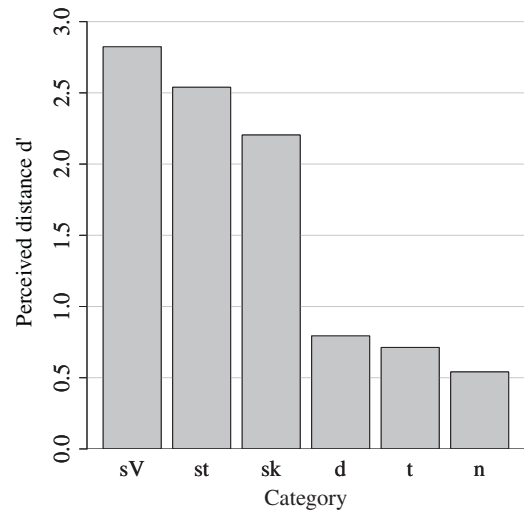
Fig. 2. Correct responses with standard error bars.

Fig. 3. Perceived distance d'.

should not be sensitive to this category distinction at all. If, for example, participants respond 'same' more often for category 'sk' than for category 'n', this will have nothing to do with the distinction between 'same' and 'different' trials, since these are evenly distributed for both categories. If participants then do respond 'same' more often for category 'sk', then this is caused by a bias participants have to respond 'same' whenever a trial from this category is presented to them. Including **Category** as an independent variable in the model allows us therefore to factor out such biases.

The interaction between **Stimulus** and **Category** gives an estimation of the sensitivity to 'same' and 'different' trials depending on the category the trials belong to. If participants are more likely to respond 'same' to 'same' trials and 'different' to 'different' trials in category 'sk' than in category 'n', then 'same' and 'different' trials must be more distinct in category 'sk' than in category 'n'. If 'same' and 'different' trials are less confusable in category 'sk', then the perceived distance between alveolar and retroflex coronals is greater for [sk] – [ʂk] than for [n] – [ɳ]. If the effect of the interaction **Stimulus + Category** is significant when comparing two categories, then the perceived distances between their alveolar and retroflex coronals are significantly different. These effects are reported in (13).[9]

(13)

	Category		Chi-square	p	
'sV'	vs.	'st'	$\chi^2(1) = 1.9$	=.17	
		'sk'	$\chi^2(1) = 12.21$	=.0005	***
		'd'	$\chi^2(1) = 104.95$	<.0001	***
		't'	$\chi^2(1) = 105.49$	<.0001	***
		'n'	$\chi^2(1) = 155.27$	<.0001	***
'st'	vs.	'sk'	$\chi^2(1) = 4.26$	=.04	*
		'd'	$\chi^2(1) = 78.74$	<.0001	***
		't'	$\chi^2(1) = 82.37$	<.0001	***
		'n'	$\chi^2(1) = 125.98$	<.0001	***
'sk'	vs.	'd'	$\chi^2(1) = 51.54$	<.0001	***
		't'	$\chi^2(1) = 53.13$	<.0001	***
		'n'	$\chi^2(1) = 86.07$	<.0001	***

The observed differences in Figs. 2 and 3 agree with the strong version of the hypothesis from section 4.2, and the results from the statistical model in (13) above confirm that these observed differences are significant: The perceived distance between

---

[9] The model is fitted using backward elimination (Draper and Smith, 1998:339ff.; Faraway, 2005:131f.). Under this approach, an initial model with all variables and all their interactions is built, and then the least significant term is removed from the model, as determined by its Wald z score. A model without this term is fitted, and this reduction continues until no insignificant terms remain. The significance of the interaction **Stimulus + Category** is determined by a likelihood ratio test (Hilbe, 2009:81f.). A separate model was built for each comparison reported in (13).

[s] and [ʂ] is greater than between [t d n] and [ʈ ɖ ɳ], and the perceived distance between [s] and [ʂ] is greater before a vowel than before a consonant, and greater before the consonant /t/ than before the consonant /k/. The only exception to this is that the perceived distance between [s] and [ʂ] is not significantly greater before a vowel than before the consonant /t/ (p = .17). As seen in Figs. 2 and 3, however, the difference between these two categories clearly trends in favor of the hypothesis.

In a post hoc analysis of this experiment, on the other hand, it is seen that the non-significance of the difference between 'sV' and 'st' is due to a ceiling effect. Furthermore, the difference between 'sV' and 'st' is significant in the second part of this perception experiment (experiment B). As a result, we cannot accept the null hypothesis that there is no difference between 'sV' and 'st' in experiment A, as this would be a type II error. The post hoc analysis and the second part of the perception experiment are outlined in the following sections.

6.5. Post hoc analysis

When the proportion of correctly identified stimuli in an experiment is 100%, or when the proportion of falsely identified stimuli is 0%, and at the same time the proportion of correctly identified stimuli is not equal to the proportion of falsely identified stimuli, then the perceived distance between the stimuli in question is infinite (Macmillan and Creelman, 2005:8). If the perceived distances within two stimulus categories reach infinity, the relative difference between these two categories cannot be estimated. Five of the fourteen participants in experiment A obtained infinite perceived distances in both category 'sV' and 'st'. When these five participants are removed from the data set, the difference in perceived distance between category 'sV' and 'st' is significant ($\chi^2(1) = 4.15$, $p = .04$).

In the second part of this experiment, participants were once again asked to recognize stimulus pairs of alveolar and retroflex coronals as same or different, using the same stimuli as in the first part of the experiment. This time, however, they had to respond quickly. The results from this part of the experiment will show that the perceived distances in category 'sV' and category 'st' are indeed significantly different.

7. Experiment B – discriminating alveolars and retroflexes quickly

7.1. Participants

Twelve native speakers of Norwegian with a mean age of 24.1 participated in the experiment, five male and seven female. They were temporary students or visitors to the Boston area. Two of them had also participated in experiment A.

7.2. Procedure

Experiment B employed the same stimuli as in experiment A, organized and presented in the same manner (see sections 6.2 and 6.3). Based on feedback reports from participants in experiment A, experiment B was shortened to avoid a similarly long experiment. Since the perceived distances for category 'sk' were seen to be significantly different from all other categories tested (see (13)), it was not included a second time in experiment B. Stimuli from category 't', 'd', and 'n' were included to distract participants from focusing only on category 'sV' and 'st', whose perceived distances were of primary interest in this experiment.

The allotted time for response in a trial was limited in experiment B. As addressed in section 6.5, infinite perceived distances occur only when participants do not make any errors in the identification of stimuli. Since participants tend to be less accurate when given less time to reach a decision (Pachella and Pew, 1968), the time constraint for responses in this part of the experiment is expected to remove the risk of such infinite perceived distances. Participants would receive visual and auditory feedback that their response was too slow if they had not responded within 900 ms from the onset of the second token in a trial. Responses were nevertheless recorded up to 1500 ms, at which point the trial would time out. All stimulus pairs were randomized and masked with multi-talker babble noise. The experiment was preceded by a brief training session intended to accustom the participants to responding quickly before the main experiment started. All participants were able to respond within 900 ms before the training session ended.[10]

7.3. Results

The experiment contained five stimulus categories ('sV', 'st', 't', 'd', 'n'), each with eight stimulus pairs (4 same – 4 different), and each pair presented six times. With 12 participants, the total number of trials in the experiment was 5 × 8 × 6 × 12 = 2880. Of these trials, 23 timed out because the participants did not respond within 1500 ms. The total number of trials submitted for analysis was therefore 2857. The relative differences of perceived distance between stimulus categories are illustrated in the figures below with proportions of correct responses (Fig. 4) and the distance parameter d' (Fig. 5).

---

[10] Experiment B employed the same multi-talker babble noise as in experiment A, but was scaled at a signal-to-noise ratio of −5 dB. The stimuli were presented with an inter-stimulus interval of 250 ms and an inter-trial interval of 500 ms. The experiment was preceded by a brief training session with the same instructions as in experiment A. The participants were also informed that it was important to answer quickly, and that they would receive instant feedback if they did not respond fast enough. The training session randomly selected 20 trials from the pool of stimuli, which were played to the participants with reduced overlaid babble.
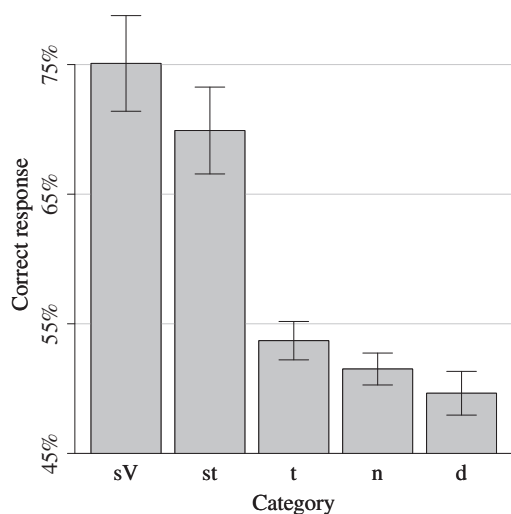
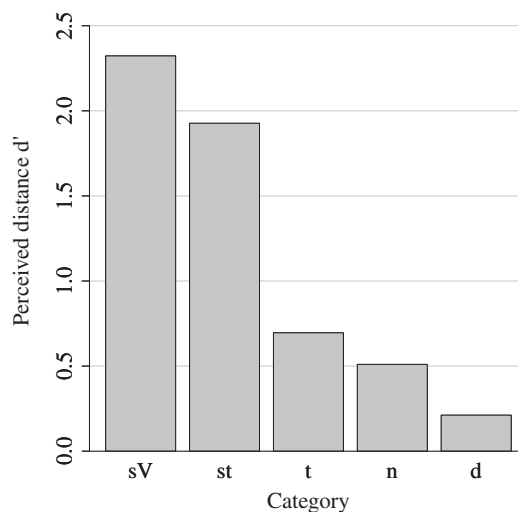**Fig. 4.** Correct responses with standard error bars.

Fig. 5. Perceived distance d′.

These results once again confirm the strong version of the hypothesis from section 4.2, and agree with the results from experiment A (see section 6.4): The perceived distance between [s] and [ʂ] is greater than between [t d n] and [ʈ ɖ ɳ], and the perceived distance between [s] and [ʂ] is greater before a vowel than before the consonant /t/. A mixed effects logistic regression model was fitted to this data as in experiment A (see section 6.4), with the addition of a variable for reaction time. The difference in perceived distances between category 'sV' and 'st' is highly significant ($\chi^2(1) = 7.13$, $p = .008$).

7.4. Post hoc analysis

In experiment A in section 6.4, the difference in perceived distances between category 'sV' and 'st' was not significant ($p = .17$). However, a post hoc analysis in section 6.5 revealed that this was due to a ceiling effect. Some participants had infinite perceived distances for both categories, and when these participants were removed, the difference between category 'sV' and 'st' was significant ($p = .04$). In experiment B, the time restriction on responses was predicted to prevent a similar ceiling effect. A post hoc analysis of this experiment reveals that no participants achieved infinite perceived distances for both category 'sV' and 'st'. The highly significant difference between these two categories ($p = .008$) is therefore a reliable indication of their true difference, so we can safely reject the null hypothesis that the perceived distances between alveolar and retroflex coronals in these two categories are the same.

One other possibility will nevertheless be considered. It is often found that participants are more accurate when they have more time at their disposal to reach a decision (Pachella and Pew, 1968). If the stimulus tokens in the 'sV' category are significantly shorter than the stimulus tokens in the 'st' category, then participants could possibly be more accurate in their responses for 'sV' tokens than for 'st' tokens, since they would have more time to reach a decision before the 900 ms limit. If this is indeed the case, we expect at least one of the following three effects to be present: (1) The stimulus tokens in the 'sV' category should be significantly shorter than the stimulus tokens in the 'st' category. (2) Participants should answer significantly faster in the 'sV' category than in the 'st' category. (3) There should be a significant interaction between reaction time and category on response accuracy (i.e. the effect of the category distinction between 'sV' and 'st' should be significantly greater at shorter reaction times than at longer reaction times). A post hoc analysis reveals that none of these three effects are found in the data, so this possibility can safely be discarded: (1) Student's paired *t*-test shows that there is no significant difference between the length of the stimulus tokens in the 'sV' and 'st' categories ($t(3) = -.55$, $p = .62$). (2) A Wilcoxon rank sum test shows that there is no significant difference in the reaction time for 'sV' and 'st' ($W = 164330$, $p = .55$). (3) A mixed effects logistic regression model shows that there is no significant effect of the interaction between reaction time and category on response accuracy ($\chi^2(2) = 1.06$, $p = .59$).

8. Summary and discussion of experiments A and B

Based on the production data of Norwegian retroflexion outlined in section 2.4, a hierarchy for the probability of retroflexion of alveolar onsets was established, in which the segments higher in the hierarchy undergo retroflexion more often than the segments lower in the hierarchy:

(14)   /t/, /d/, /n/ > /sk/ > /st/ > /sV/

In section 4.2, the hypothesis was formulated that the likelihood of retroflexion is directly correlated with the perceived distances between alveolars and retroflexes:

(15)   The greater the perceived distance between an alveolar and a retroflex, the less likely it is that the alveolar undergoes retroflexion.

(16)   <u>Probability of retroflexion</u>   <u>Perceived distance</u>

Increasing   /t d n/   [t d n]–[ʈ ɖ ɳ]
             /sk/   [sk]–[ʂk]
             /st/   [st]–[ʂʈ]
             /sV/   Increasing   [sV]–[ʂV]

The results from experiments A and B verify this hypothesis on all accounts. As predicted, the perceived distances within each alveolar-retroflex category constitute the exact mirror image of the likelihood hierarchy for retroflexion of alveolar onsets in (14):

(17)   sV > st > sk > t, d, n

As explained in section 2.4, retroflexion applies across the board for onsets in /t-/, /d-/, and /n-/. The results from experiments A and B show that the perceived distances between alveolars and retroflexes for these three categories are considerably smaller than for the other categories. This indicates that when the similarity between alveolars and retroflexes reaches a certain threshold, retroflexion always applies. Since retroflexion always applies for /t-/, /d-/, and /n-/, the perceived distances for these categories have clearly reached this threshold. Possible differences in the perceived distances for these categories are therefore not investigated further. Even if there were differences among them, they cannot correspond to an increase in retroflexion rate, since retroflexion cannot apply more than 100% of the time.

Experiments A and B show that the perceived distance between alveolars and retroflexes is quite substantial for categories 'sV', 'st', and 'sk', and it was necessary to employ both background noise and reduced response time in order to prevent participants from correctly distinguishing the tokens in these categories in every trial. A consequence of this method is that it becomes quite difficult to distinguish between alveolars and retroflexes in categories 't', 'd', and 'n' under these conditions. However, this does not mean that the perceived distance between alveolars and retroflexes in categories 't', 'd', and 'n' is so small that this distinction risks being neutralized in the Norwegian language. The difficult conditions were imposed in order to estimate the relative difference in perceived distances between the categories, with no implication of what the absolute perceived distances between alveolars and retroflexes are. Since all participants in experiment A completed a training session in which they had to correctly distinguish alveolars from retroflexes in every category 100% of the time, it should be clear that their difficulty in doing the same during the main experiment is primarily an artifact of the strict conditions imposed on the task. There are also no indications from spoken or written Norwegian that there is any confusion or risk of neutralization of alveolar and retroflex coronals within categories 't', 'd', and 'n'.

Now that a clear correlation between the likelihood of retroflexion and perceptual properties has been identified, the remaining big question is why this correlation exists. In the following sections, it is proposed that the more perceptually distinct a retroflex token is from the alveolar base form, the greater the risk that language learners will not categorize that retroflex token as a variant of that word. As a result, language learners will construct a grammar in which perceptually distinct retroflex tokens of this kind are less likely to be produced.

9. From perception to phonology

When alveolar onsets undergo retroflexion after /r/, the phonetic features distinguishing alveolar coronals from retroflex coronals necessarily change. Within the framework of optimality theory, there is an unfaithful mapping of these features from the input, which is alveolar, to the output, which is retroflex (Prince and Smolensky, 2004:2ff.). As an example, when an underlying alveolar /t/ surfaces as a retroflex [ʈ], faithfulness constraints referring to the phonetic features distinguishing the two are violated. These faithfulness constraints will simply be called 'Faith (t)' here. The faithfulness constraints which are violated when an alveolar /d/ surfaces as a retroflex [ɖ] will be called 'Faith (d)', and correspondingly for /sV/, /st/, /sk/, and /n/. When alveolar onsets in /sV-/ undergo retroflexion less often than alveolar onsets in /st-/, it implies that the retroflex candidates violating Faith (sV) are less well-formed than the retroflex candidates violating Faith (st), by which we deduce that Faith (sV) must be ranked above Faith (st) in the hierarchy of faithfulness constraints: Faith (sV) ≫ Faith (st). Transposing the likelihood hierarchy of retroflexion in (14) into a hierarchy of faithfulness constraints will give us the ranking in (18):

(18)   FAITH (sV) ≫ FAITH (st) ≫ FAITH (sk) ≫ FAITH (t), FAITH (d), FAITH (n)

The ranking in (18) corresponds to the ranking in (17) of perceived distances within each alveolar-retroflex category, repeated in (19) below:

(19)   sV > st > sk > t, d, n

The suggestion which will be made here is that the ranking of perceptual properties in (19) leads to the corresponding ranking of phonological constraints in (18). There are, however, two distinct ways of accounting for the fact that perceptual properties and constraint rankings are linked in the grammar.

According to the 'perceptibility-map' hypothesis, an inherent mechanism of the grammar translates rankings of relative perceived distances directly into constraint rankings (Steriade, 2001:239; Steriade, 2009:164; Wilson, 2006:958f.). Although this approach provides an easy way of implementing the observation that the two are linked, it does so at considerable cost, since it stipulates that the link is simply there by design. In other words, a property of the grammar is explained by assuming that the grammar already comes equipped with this property, an approach which therefore increases the number of assumptions made about the inherent state of the grammatical system.

If, on the other hand, it is possible to derive this property from other principles, it will allow us to dispense with this assumption. For this reason, the idea which will be pursued here is that the connection between perceived distances and constraint rankings is not inherent, but rather an emergent property of grammar, one that arises from mechanisms of grammar learning. The remainder of this paper is devoted to this endeavor.

10. Categorization of perceptual stimuli

10.1. General perception

A very basic property of human perception is given in (20) (cf. e.g. Shepard, 1957; Luce, 1963; Nosofsky, 1986):

(20)   The greater the perceived distance between category x and stimulus x', the less likely x' is to be labeled as an instance of x.

This property can be illustrated with color perception. From our experience with colors, we have established certain color categories, such as 'blue'. When a new object is perceived, the likelihood that we will label the color of that object as an instance of our color category 'blue' will depend on its perceived similarity to 'blue'. For example, we are more likely to label an instance of navy blue as 'blue' than we are for an instance of turquoise, since navy blue is perceived as more similar to stereotypical 'blue' than is an instance of turquoise (Fig. 6).

It is assumed here that categorization of linguistic stimuli operates in the same manner as categorization of perceptual stimuli in general. In other words, the likelihood that a token x' is categorized as word x will be a function of its perceived similarity to x. Experimental evidence for this proposal is reviewed in the next section.

10.2. Perceived distance in word categorization

In a priming experiment by Marslen-Wilson et al. (1996), the assumption is made that when a prime word x' facilitates the recognition of a target word semantically related to word x, then x' has been categorized as a token of word x (1996:1379). In their experiment, they manipulate the acoustic properties of the initial consonant of prime word x', and they find that the more perceptually distinct x' is from x, the smaller the priming effect of x' on the target word (1996:1386f.). This finding indicates that the larger the perceived distance is between token x' and word x, the less likely it is that x' is categorized as a token of word x.
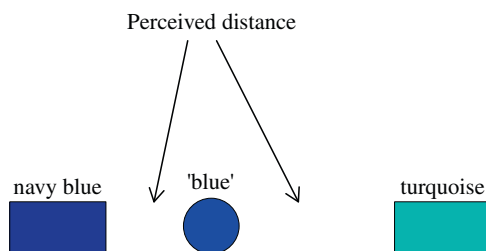

Fig. 6. Color perception.

Skoruppa et al. (in press) conduct a phonological learning experiment where they manipulate the feature distinctions of the initial consonant of phonologically alternating noun forms. In the first language, the alternating noun forms differ in the place features of the initial consonant, such that [pamu] alternates with [tamu]. In the second language, the feature distinctions between the alternating forms are increased, in that both place and manner features change in the initial consonant, such that [pamu] alternates with [samu]. When participants are tested on the same nouns as they were trained on, Skoruppa et al. find that learners of the first language reproduce the correct forms 78% of the time, whereas learners of the second language do so only 23% of the time. One interpretation of these results is that learners of the second language failed to reproduce the same alternating forms as they were trained on because they did not categorize perceptually distant forms such as [pamu] ~ [samu] as tokens of the same word, whereas this was not the case for learners of the first language, where the alternating forms were perceptually more similar to each other. This finding therefore also indicates that the larger the perceived distance is between token x' and word x, the less likely it is that x' is categorized as a token of word x.

10.3. Perceived distance in the categorization of retroflex tokens

The effect of perceived distance on the categorization of retroflex tokens should therefore be clear: The larger the perceived distance is between the base form of a word with an alveolar onset and a token of that word with a retroflex onset, the greater the possibility that listeners will not categorize the retroflex token as a token of the alveolar word. As an example, since the perceived distance between alveolar [sV] and retroflex [ʂV] is greater than between alveolar [n] and retroflex [ɳ], this means that a retroflex token in [ʂV-] is more likely to not be categorized as a token of the word in /sV-/ than would be the case for a retroflex token in [ɳ-] as a token of the word in /n-/.

11. Consequences for grammar learning

Language learners construct a grammar based on the distribution of forms in the learning data. At the same time, the exact manner in which the learner has perceived, identified, and categorized these forms plays a major role in how the grammar is constructed during the learning process. Specifically, if retroflex tokens that are perceptually distant from their base forms are less likely to be categorized as tokens of their alveolar words, then these alveolar words will have fewer retroflex tokens associated with them in the categorized input for the learner. The assumption here is that language learners aim to construct a grammar that replicates the distribution of forms in this categorized input. As a result, learners will then construct a grammar in which these alveolar words are less likely to surface with retroflex tokens. In short, learners of Norwegian construct a grammar in which perceptually distant retroflex tokens are less frequently produced because these perceptually distant retroflex tokens were less frequent in the categorized input for these learners.

When learners are less likely to produce retroflex tokens from underlying alveolar words in some contexts, we can phrase this pattern with the terminology from optimality theory and say that learners are more faithful to the underlying form in these cases. Modeling these faithfulness patterns with constraints, it would mean that the faithfulness constraint preventing retroflexion of an underlying alveolar /sV/ ranks above the corresponding constraint for an underlying /n/, given the effect perceived distances have on the categorization of the retroflex tokens of these onsets, as discussed in section 10.3. Since there is such a link between perceived distance, categorization, and ranking of faithfulness constraints, this predicts that we can derive the faithfulness ranking in (21) below from the observed perceived distances between alveolar and retroflex coronals.

(21)   FAITH (sV) ≫ FAITH (st) ≫ FAITH (sk) ≫ FAITH (t), FAITH (d), FAITH (n)

The predicted ranking in (21) is identical to the ranking in (18). In connection with (18), it was stated that it should be possible to derive this ranking of faithfulness constraints from perceived distances without simply assuming that an inherent property of grammar translates one into the other. As highlighted in this section, this is indeed possible when we consider the effect perceived distances have on the categorization of retroflex tokens during grammar learning. A learning simulation presented in the next section will illustrate how the connection between word categorization and grammar learning is able to result in a pattern that mimics the retroflexion pattern of Norwegian.

12. Learning simulation

For the simulation in this section, the phonological grammar will be modeled using constraints. As discussed in section 9, the constraints that militate against retroflexion of underlying alveolar onsets will be called FAITH (sV), FAITH (st), and so on. The constraints promoting retroflexion will simply be called APPLY RETROFLEXION AFTER /ɾ/, or just RETRO for short. As the name implies, this constraint is violated when a word with an underlying alveolar onset surfaces with an alveolar rather than a retroflex after a word ending in /ɾ/.

The constraint model used here is Harmonic Grammar (Pater, 2009). Unlike classic optimality theory (Prince and Smolensky, 2004), constraints are weighted rather than ranked. An output candidate violating a high-weighted constraint

receives a higher penalty than an output candidate violating a low-weighted constraint, and the candidate with the lowest penalty is the most 'harmonic' candidate selected as the winner. The likelihood of retroflexion of an alveolar onset is therefore determined by the weight assigned to the faithfulness constraints in (21). The higher the weight, the less likely a retroflex candidate is to surface for that onset, since the retroflex candidate would violate this faithfulness constraint. The ranking hierarchy in (21) is therefore equivalent to a weight hierarchy of the same constraints.

For the learning simulation, I have adopted the maximum entropy algorithm as implemented by Wilson (2006:956ff.) and Wilson and George (2009), where a function is applied to find the appropriate constraint weights needed to maximize the probability of the forms encountered in the learning data. Constraint biases can be added to this function by specifying the target weight value μ and deviation value σ for any given constraint, where a lower σ value yields greater penalties for deviating from μ. In the 'perceptibility-map' hypothesis, the connection between perceived distances and constraint rankings is captured by translating perceived distances directly into σ values (Wilson, 2006:958f.). The central point in this simulation, however, is that no assumptions about constraint biases based on perceived distances are needed in order to capture this link. Instead, the prediction is that the different weights assigned to the family of FAITH constraints will emerge from the learning process instead of being assumed at the outset (see section 9). For this reason, all constraints in this simulation are given the same default μ and σ values ($\mu = 0$, $\sigma^2 = 100,000$).

The learning algorithm takes a batch of mappings between underlying forms and surface forms as its input data (Wilson, 2006:956). As shown in section 2.4, the retroflexion rates are different for the various alveolar onsets in Norwegian. If the learning algorithm is fed such a distribution in the input data, we would only be testing its ability to replicate this distribution. Our goal, however, is to show how these differences in the distribution can arise from the learning process itself. As a consequence, the input data in this simulation assumes that all alveolar onsets behave uniformly. Specifically, it is assumed that retroflexion applies across the board for all onsets.

The number of tokens in the input data in this simulation is informed by the Norwegian lexicon. Based on the token frequencies in the LBK corpus,[11] an estimation was made of the probability that words would begin with the alveolar onsets /sV-/, /st-/, /sk-/, /t-/, /d-/, and /n-/, as well as the probability that a word would end in /-r/, which serves as the necessary trigger of retroflexion (see section 2.3). With the assumption that retroflexion applies consistently to all onsets, the product of these two probabilities will therefore give the probability of retroflex tokens for these alveolar onsets in the lexicon. In order to arrive at an absolute number of tokens for the input data, these probabilities were multiplied by 6,000,000, which is the average number of words in child-directed speech within a one-year period (Hart and Risley, 1995:132). The number of tokens in the input data for words in alveolar onsets in the position after /-r/ is given in (22):

(22)

Onset	Tokens
/sV-/	61179
/st-/	15995
/sk-/	9185
/t-/	46659
/d-/	65331
/n-/	21359

As mentioned above, it will be assumed that all of these tokens are produced with retroflexion. Note in this case that the numerical distribution of retroflex tokens in (22) does not correspond to the ranking of faithfulness constraints in (21). This is a welcoming result, since we want to be sure that the learning process does not derive the ranking in (21) from absolute token frequencies.

The hypothesis from sections 10.3 and 11 is that retroflex tokens perceptually close to the alveolar base form are more likely to be categorized with the underlying alveolar word, and that the learner constructs the constraint ranking in (21) as a result. The input data in this simulation will therefore need to distinguish between categorized and non-categorized forms. One possibility would be to remove non-categorized tokens from the data altogether, but this would only change the absolute frequencies of tokens in (22), and it would make the unintuitive assumption that retroflex tokens not categorized by the listener are equivalent to those tokens never having been produced by the speaker. Instead, I follow the suggestion in Marslen-Wilson et al. (1996:1388) that categorization is different from identification. As they point out, listeners are normally able to identify words that have been mispronounced, but they are less likely to categorize these tokens as eligible variants of the identified words. The same will hold true for words produced with abnormal or idiosyncratic features, words spoken in a foreign accent, and – crucially – for output forms perceptually distant from their underlying base forms.

In this simulation, the non-categorized retroflex tokens will be treated as 'NULL' output forms of their underlying representations. Since they are not categorized with the underlying forms, they vacuously satisfy correspondence constraints such as FAITH, but violate a constraint called 'PARSE', which requires all identified forms to be categorized with an underlying

---

[11] Lexicographic corpus for Norwegian Bokmål: http://www.hf.uio.no/iln/tjenester/kunnskap/sprak/korpus/skriftsprakskorpus/lbk/index.html.

representation. An illustration of the mappings between underlying forms and surface forms in this simulation is provided in (23) below.

(23)

/-ɾ sV/		RETRO	FAITH (sV)	PARSE
a.	sV	*		
b.	ʂV		*	
c.	NULL			*

In the tableau in (23), the underlying string /sV/ is in a retroflexing environment, i.e. after the trigger /ɾ/. Phonetically speaking, there are two possible surface forms: The alveolar form [sV] (candidate a) and the retroflex form [ʂV] (candidate b and c). The retroflex surface form is sufficiently perceptually distinct from its underlying alveolar representation that it risks not being categorized with it. A non-categorized retroflex surface form is treated as NULL in (23). As illustrated in this tableau, the alveolar form [sV] violates the constraint RETRO (= APPLY RETROFLEXION AFTER /ɾ/), the categorized retroflex form [ʂV] violates the faithfulness constraint FAITH (sV), and the non-categorized retroflex form NULL violates the constraint PARSE.

For each alveolar onset, the probability P of their retroflex tokens being categorized with the underlying alveolar form is a function of the perceived similarity between alveolars and retroflexes. In this simulation, P is calculated from the proportions of correct responses in experiment A with the function in (24), where μ represents the mean proportion of correct responses (see section 6.4). Since probabilities cannot be greater than 1, $P > 1 = 1$. Probability P for each onset is given in (25).

(24) $P = 1 - \frac{\mu - 0.7}{2}$

(25)

Onset	P
/sV-/	.933
/st-/	.943
/sk-/	.969
/t-/	1
/d-/	1
/n-/	1

In accordance with the token frequencies in (22) and the probabilities in (25), the number of categorized retroflex [ʂV-] tokens is $61{,}179 \times 0.933$, and the remaining non-categorized retroflex [ʂV-] tokens are treated as 'NULL' forms, as illustrated in (26) below.

(26)

/-ɾ sV/		Distribution	RETRO	FAITH (sV)	PARSE
a.	sV	0	*		
b.	ʂV	57054.45		*	
c.	NULL	4124.09			*

The same procedure applies to the other alveolar onsets. The learning algorithm takes these token distributions as its learning data and assigns constraint weights to maximize the probability of this distribution. The weights assigned to the faithfulness constraints by this algorithm are seen in (27), translated into a constraint hierarchy.

(27)

Constraint	Weight
FAITH (sV)	15.42
FAITH (st)	15.25
FAITH (sk)	14.59
FAITH (t)	0
FAITH (d)	0
FAITH (n)	0

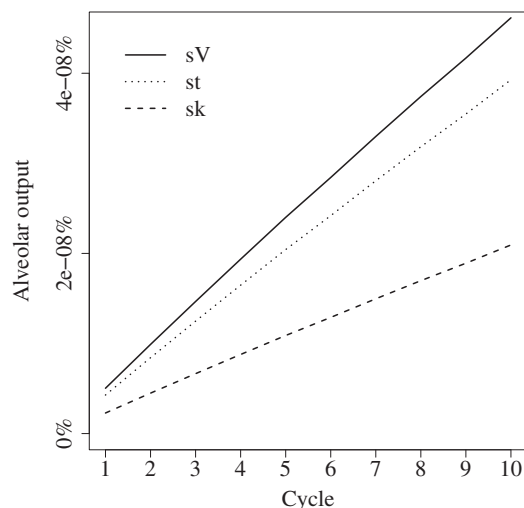⇒ FAITH (sV) ≫ FAITH (st) ≫ FAITH (sk) ≫ FAITH (t), FAITH (d), FAITH (n)

Fig. 7. Predicted alveolar output forms.

As predicted, the weights assigned by this learning algorithm (27) correspond to the hypothesized constraint ranking in (21).

In the position after /ɾ/, words with the alveolar onset /s-/ can surface either with an alveolar [s-] or a retroflex [ʂ-]. For the other alveolar onsets in /t- d- n-/, a retroflex surface form is the only option (see section 2.4). With the starting point assumed in this learning simulation that retroflexion applies consistently also to onsets in /s-/, the proportion of alveolar surface forms predicted to be produced from underlying forms in /s-/ is close to zero. However, the amount of predicted alveolar output forms is proportional to the categorization rate P of retroflex tokens, listed in (25). Specifically, the lower P is, the higher the proportion of alveolar output forms. When taking the output forms predicted by the learning algorithm as the input data to a consequential learning process, this effect is perpetuated such that the proportion of alveolar output forms increases for every learning cycle, as illustrated in Fig. 7.[12]

Even with a starting point where retroflexion applies to all output forms, the different probabilities of categorizing the retroflex tokens can therefore in theory be enough to start a phonological tendency by which alveolar onsets with a large perceived distance between their alveolar and retroflex tokens are more likely to surface with alveolar tokens than other alveolar onsets are.

13. Discussion of learning simulation

Without assuming any constraint biases based on perceived distances, the learning simulation in section 12 generates a grammar that ranks faithfulness constraints in correspondence with such perceived distances (27). It is necessary to point out, though, that this outcome is trivially predictable. Just as this ranking of constraints would straightforwardly follow from the assumption that there is a function that assigns prior values based on perceived distances, as in the 'perceptibility-map' hypothesis (Wilson, 2006:958f.), the same ranking also follows from the assumption that there is a function that generates probabilities of non-categorized tokens which do not violate FAITH. The important distinction between these two approaches is that there is no independent evidence for the function assumed under the 'perceptibility-map' hypothesis, whereas there is good experimental evidence that the likelihood for a token to be assigned to a category is a function of the perceived similarity between them.

The link between the proportions of categorized retroflex tokens and the probability that such tokens are produced can be implemented in any model that is centered on the connection between the perceived input in the learning data and the produced output by the learner. As an example, the hypothesis from section 11 invites a similar implementation in exemplar theory, where the proportion of categorized token variants x' to the base form x plays a direct role in predicting the proportion of x' tokens in the output (Goldinger, 1998). The learning simulation performed in section 12 was conducted using a maximum entropy algorithm in Harmonic Grammar primarily because it is more precise in its formulation and implementation of phonological alternations than current formulations of exemplar theory. Yet the fact that the general hypothesis from section 11 can be implemented in more than one model means that the predicted result in (27) does not, and should not, rely exclusively on the specifics of the model used in section 12, although it is naturally reassuring that the hypothesis does give the predicted results when implemented in this widely used model of phonological learning.

---

[12] The predicted proportions of alveolar output forms for /t- d- n-/ remain too close to zero to allow the learning algorithm to converge in ten consecutive cycles. These are therefore left out of this part of the simulation.

14. Summary

Based on the production data of Norwegian retroflexion outlined in section 2.4, a hierarchy for the probability of retroflexion of alveolar onsets was established, in which the segments higher in the hierarchy undergo retroflexion more often than the segments lower in the hierarchy:

(28)   /t/, /d/, /n/ > /sk/ > /st/ > /sV/

In section 4.2, the hypothesis was formulated that the likelihood of retroflexion is directly correlated with the perceived distances between alveolars and retroflexes:

(29)   The greater the perceived distance between an alveolar and a retroflex, the less likely it is that the alveolar undergoes retroflexion.

(30)   Probability of retroflexion   Perceived distance

        Increasing   /t d n/   [t d n]–[ʈ ɖ ɳ]

               ↑   /sk/   [sk]–[ʂk]

               | /st/   [st]–[ʂʈ]

               | /sV/   Increasing   [sV]–[ʂV]

The results from experiments A and B verify this hypothesis on all accounts. As predicted, the perceived distances within each alveolar-retroflex category constitute the exact mirror image of the likelihood hierarchy for retroflexion of alveolar onsets in (28):

(31)   sV > st > sk > t, d, n

In a constraint model of grammar where faithfulness constraints militate against phonological alternations, the tendency to apply retroflexion more to alveolar onsets higher up in the hierarchy in (28) implies the inverse ranking of their faithfulness constraints (32):

(32)   Faith (sV) ≫ Faith (st) ≫ Faith (sk) ≫ Faith (t), Faith (d), Faith (n)

The correlation between perceived distances in (31) and the constraint ranking in (32) suggests that the differences in perceived distances (31) are the cause of the grammatical patterns (32). Building on properties of human perception and psycholinguistic experiments, the hypothesis was made that the perceived distances between alveolars and retroflexes influence how retroflex tokens are categorized:

(33)   The greater the perceived distance between a retroflex token and the alveolar base form, the less likely the retroflex token is to be categorized as a token of the alveolar word.

The greater the perceived distance between the alveolar and the retroflex (31), the lower the proportion of retroflex tokens associated with words with alveolar onsets (33). When learners construct a grammar based on the forms they have categorized in the input data, the ranking in (32) is therefore predicted to emerge. A learning simulation is conducted to test this prediction in section 12, and this simulation generates a constraint ranking identical to the ranking in (32):

(34)   Faith (sV) ≫ Faith (st) ≫ Faith (sk) ≫ Faith (t), Faith (d), Faith (n)

The experiments conducted in this paper show that perceived distances are directly correlated with the probability of phonological alternations, and a learning simulation demonstrates that it is possible to derive this link from properties of stimulus categorization without the need to assume that grammar inherently favors phonological alternations with small perceptual modifications.

anonymous reviewer for their comments on earlier versions of this paper. An early version of this work was presented at the MIT Phonology Circle and NELS 40, and I wish to thank the audience at these events for their feedback.

References

Audacity Team, 2006. Audacity 1.2.6 (computer software). Retrieved from http://audacity.sourceforge.net/.

Bates, D., Maechler, M., Bolker, B., 2011. lme4: Linear Mixed-effects Models Using S4 Classes. R Package Version 0.999375-40. http://CRAN.R-project.org/package=lme4.

Beckers, G., 2002–2008. rms Equalize. Retrieved from http://www.gbeckers.nl/pages/praatscripts.html.

Blevins, J., 2004. Evolutionary Phonology. The Emergence of Sound Patterns. Cambridge University Press, Cambridge.

Boersma, P., Weenink, D., 2009. Praat: Doing Phonetics by Computer 5.1.07 (computer software). Retrieved from http://www.praat.org/.

Boomershine, A., et al., 2009. The impact of allophony versus contrast on speech perception. In: Avery, P., Dresher, B.E., Rice, K. (Eds.), Contrast in Phonology. Theory, Perception, Acquisition. Mouton de Gruyter, Berlin, pp. 145–171.

Borgstrøm, C., 1938. Zur Phonologie der norwegischen Schriftsprache (nach der ostnorwegischen Aussprache). Norsk tidsskrift for sprogvidenskap 9, 250–273.

Brekke, K., 1881. Bidrag til dansk-norskens lydlære. In: Indbydelsesskrift til den offentlige examen i juni og juli 1881 ved Aars og Voss's latin- og realskole. W. C. Fabritius, Kristiania, pp. 1–66.

Chomsky, N., Halle, M., 1968. The Sound Pattern of English. Harper & Row Publishers, New York, NY.

Cohen, J., et al., 1993. PsyScope: a new graphic interactive environment for designing psychology experiments. Behavior Research Methods, Instruments, and Computers 25 (2), 257–271.

Draper, N.R., Smith, H., 1998. Applied Regression Analysis. John Wiley & Sons, New York, NY.

Eliasson, S., 1986. Sandhi in Peninsular Scandinavian. In: Andersen, H. (Ed.), Sandhi Phenomena in the Languages of Europe. Mouton de Gruyter, Berlin, pp. 271–300.

Endresen, R.T., 1974. On retroflex segments in Norwegian. Norwegian Journal of Linguistics 28, 73–78.

Faraway, J.J., 2005. Linear Models with R. Chapman & Hall/CRC, Boca Raton, FL.

Foldvik, A.K., 1977. Realisasjonen av r i norsk. Svenska landsmål och svenskt folkliv 100, 110–118.

Fretheim, T., 1969. Norwegian stress and quantity reconsidered. Norsk tidsskrift for sprogvidenskap 23, 76–96.

Goldinger, S.D., 1998. Echoes of echoes? An episodic theory of lexical access. Psychological Review 105 (2), 251–279.

Golestani, N., Zatorre, R.J., 2004. Learning new sounds of speech: reallocation of neural substrates. NeuroImage 21, 494–506.

Hart, B., Risley, T.R., 1995. Meaningful Differences in the Everyday Experience of Young American Children. Paul H. Brookes Publishing Co., Baltimore, MD.

Haugen, E., 1948. Mere om r-bortfall i sørøstlandsk. Maal og minne 117–122.

Hayes, B., Kirchner, R., Steriade, D. (Eds.), 2004. Phonetically Based Phonology. Cambridge University Press, Cambridge.

Hilbe, J.M., 2009. Logistic Regression Models. CRC Press, Taylor & Francis Group, Boca Raton, FL.

Hovdhaugen, E., 1969. Transformasjonell generativ grammatikk. Annen utgave med øvelser. Universitetsforlaget, Oslo.

Huang, T., 2001. The interplay of perception and phonology in tone 3 sandhi in Chinese Putonghua. In: Hume, E., Johnson, K. (Eds.), The Ohio State University Working Papers in Linguistics No. 55. Studies on the Interplay of Speech Perception and Phonology. The Ohio State University, Department of Linguistics, Columbus, OH, pp. 23–42.

Knobluach, K., 2011. psyphy: Functions for Analyzing Psychophysical Data in R. R Package Version 0.1-5. http://CRAN.R-project.org/package=psyphy.

Kohler, K.J., 1990. Segmental reduction in connected speech in German: phonological facts and phonetic explanations. In: Hardcastle, W.J., Marchal, A. (Eds.), Speech Production and Speech Modeling. Kluwer Academic Publishers, Dordrecht, pp. 69–92.

Kristoffersen, G., 2000. The Phonology of Norwegian. Oxford University Press, Oxford.

Luce, R.D., 1963. Detection and recognition. In: Luce, R.D., Bush, R.R., Galanter, E. (Eds.), Handbook of Mathematical Psychology, vol. I. John Wiley and Sons, Inc., New York, NY, (Chapters 1–8), pp. 103–189.

Macmillan, N.A., Creelman, C.D., 2005. Detection Theory: A User's Guide. Lawrence Erlbaum Associates, Inc., Mahwah, NJ.

Maddieson, I., 1984. Patterns of Sounds. Cambridge University Press, Cambridge.

Marslen-Wilson, W., Moss, H.E., van Helen, S., 1996. Perceptual distance and competition in lexical access. Journal of Experimental Psychology: Human Perception and Performance 22 (6), 1376–1392.

Mielke, J., 2003a. The diachronic influence of perception: experimental evidence from Turkish. In: Nowak, P.M., Yoquelet, C., Mortensen, D. (Eds.), Proceedings of the Twenty-Ninth Annual Meeting of the Berkeley Linguistics Society. General Session and Parasession on Phonetic Sources of Phonological Patterns: Synchronic and Diachronic Explanations, February 14–17, 2003. Berkeley Linguistics Society, Berkeley, CA, pp. 557–567.

Mielke, J., 2003b. The interplay of speech perception and phonology: experimental evidence from Turkish. Phonetica 60, 208–229.

Moen, I., et al., 2003. The articulation of the East Norwegian apical liquids /l ɾ ʈ/. In: Solé, M.-J., Recasens, D., Romero, J. (Eds.), Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona, 2003. Causal Productions, Adelaide, pp. 1755–1758.

Nosofsky, R.M., 1986. Attention, similarity, and the identification-categorization relationship. Journal of Experimental Psychology: General 115 (1), 39–57.

Ohala, J.J., 1981. The listener as a source of sound change. In: Masek, C.S., Hendrick, R.A., Miller, M.F. (Eds.), Papers from the Parasession on Language and Behavior. Chicago Linguistic Society, Chicago, IL, pp. 178–203.

Pachella, R.G., Pew, R.W., 1968. Speed-accuracy tradeoff in reaction time: effect of discrete criterion times. Journal of Experimental Psychology 76 (1), 19–24.

Pater, J., 2009. Weighted constraints in generative linguistics. Cognitive Science 33, 999–1035.

Paul, H., 1920. Prinzipien der Sprachgeschichte. Fünfte Auflage. Verlag von Max Niemeyer, Halle an der Saale.

Polka, L., 1991. Cross-language speech perception in adults: phonemic, phonetic, and acoustic contributions. The Journal of the Acoustical Society of America 89 (6), 2961–2977.

Prince, A., Smolensky, P., 2004. Optimality Theory. Constraint Interaction in Generative Grammar. Blackwell Publishing, Malden, MA.

R Development Core Team, 2011. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN: 3-900051-07-0, http://www.R-project.org/.

Repp, B.H., 1984. Closure duration and release burst amplitude cues to stop consonant manner and place of articulation. Language and Speech 27 (3), 245–254.

Rinnan, G.D., 1969. Nok en gang om alveolarene. In: Tilegnet Carl Hj. Borgstrøm. Et festskrift på 60-årsdagen 12.10.1969 fra hans elever. Universitetsforlaget, Oslo, pp. 128–133.

Rykkvin, O., 1946. Om r-bortfall i søraustlandsk. Maal og minne 144–150.

Seo, M., 2001. A perception-based study of sonorant assimilation in Korean. In: Hume, E., Johnson, K. (Eds.), The Ohio State University Working Papers in Linguistics No. 55. Studies on the Interplay of Speech Perception and Phonology. The Ohio State University Department of Linguistics, Columbus, OH, pp. 43–69.

Shepard, R.N., 1957. Stimulus and response generalization: a stochastic model relating generalization to distance in psychological space. Psychometrika 22 (4), 325–345.

Sievers, E., 1901. Grundzüge der Phonetik zur Einführung in das Studium der Lautlehre der indogermanischen Sprachen. Fünfte verbesserte Auflage. Druck und Verlag von Breitkopf & Härtel, Leipzig.

Simonsen, H.G., Moen, I., 2004. On the distinction between Norwegian /ʃ/ and /ç/ from a phonetic perspective. Clinical Linguistics & Phonetics 18 (6–8), 605–620.

Simonsen, H.G., Moen, I., Cowen, S., 2008. Norwegian retroflex stops in a cross linguistic perspective. Journal of Phonetics 36, 385–405.

Skoruppa, K., Lambrechts, A., Peperkamp, S. The role of phonetic distance in the acquisition of phonological alternations. Proceedings of NELS 39, in press.

Standwell, G.J., 1975. Norwegian phonology. Acta Linguistica Academiae Scientiarum Hungaricae 25 (3–4), 339–377.

Stausland Johnsen, S., 2011. The origin of variation in Norwegian retroflexion. Ph.D. Dissertation. Harvard University.

Steriade, D., 2001. Directional asymmetries in place assimilation. A perceptual account. In: Hume, E., Johnson, K. (Eds.), The Role of Speech Perception in Phonology. Academic Press, San Diego, CA, pp. 219–250.

Steriade, D., 2009. The phonology of perceptibility effects: the P-map and its consequence for constraint interaction. In: Hanson, K., Inkelas, S. (Eds.), The Nature of the Word. Studies in Honor of Paul Kiparsky. The MIT Press, Cambridge, MA, pp. 151–179.

Storm, J., 1884. Norsk Lydskrift med Omrids af Fonetiken. Første Afsnit. Norvegia. Tidsskrift for det norske Folks Maal og Minder 1, 19–132.

Torp, A., 2007. R – ei urokråke i språket. Det norske samlaget, Oslo.

Vanvik, A., 1972. A phonetic-phonemic analysis of Standard Eastern Norwegian. Part I. Norwegian Journal of Linguistics 26 (2), 119–164.

Vogt, H., 1939. Some remarks on Norwegian phonemics. Norsk tidsskrift for sprogvidenskap 11, 136–144.

Western, A., 1889. Kurze darstellung des norwegischen lautsystems. Phonetische studien. Zeitschrift für wissenschaftliche und praktische phonetik mit besonderer rücksicht auf die phonetische reform des sprachunterrichts 2, 259–282.

Wilson, C., 2006. Learning phonology with substantive bias: an experimental and computational study of velar palatalization. Cognitive Science 30, 945–982.

Wilson, C., George, B., 2009. MaxEnt Grammar Tool (computer software). Retrieved from http://www.linguistics.ucla.edu/people/hayes/MaxentGrammarTool/.