TITLE PAGE

Title: Cross-linguistic variation in word-initial cluster production in adult and child language: evidence from English and Norwegian

Authors and affiliations:
Nina Gram GARMANN, Department of Early Childhood Education, OsloMet – Oslo Metropolitan University and MultiLing, University of Oslo
Hanne Gram SIMONSEN, MultiLing, University of Oslo
Pernille HANSEN, MultiLing, University of Oslo
Elisabeth HOLM, Department of Early Childhood Education, OsloMet – Oslo Metropolitan University
Brechtje POST, University of Cambridge
Elinor PAYNE, University of Oxford

Keywords: consonant clusters, epenthesis, vowel insertion, vowel intrusion, prosodic-phonetic biases

Acknowledgements:
This research was funded by a British Academy Small Research Grant SG122210 'The acquisition of consonant timing: a study in cross-linguistic micro-variation' (PI Elinor Payne), by funding from the University of Oslo's Centre for MultiLingualism in Society Across the Lifespan (MultiLing, The Research Council of Norway through its Centres of Excellence scheme, project number 223265), and from the Faculty of Education and International Studies at OsloMet – Oslo Metropolitan University.

We thank the participants in the study, parents and children, for providing invaluable data for this piece of research. We would also like to thank Nina Hagen Kaldhol, Holly Kennard, Ben Molineaux, Elaine Schmidt, Eirik Tengesdal, and Ane Theimann for their assistance in recording and segmenting the data, Henrik Torgersen for blind coding, and our late friend Inger Moen for invigorating discussions on clusters. We wish to thank the audiences at Cognitive Summer Seminars in Norway 2012, 2014, 2016, 2018, ICPC in York 2011, ICPLA-conferences 2010, 2016, 2018, ISMBS 2015, Nordic Child Language Symposium 2016, 2018, NorPhLex meetings 2012–13, as well as the reviewers of this article, for valuable comments and suggestions.

Abstract

Young children simplify word initial consonant clusters by omitting or substituting one (or both) of the elements. Vocalic insertion, coalescence and metathesis are said to be used more seldom (McLeod, van Doorn & Reed, 2001). Data from Norwegian children, however, have shown vocalic insertion to be more frequently used (Simonsen, 1990; Simonsen, Garmann & Kristoffersen, 2019). To investigate the extent to which children use this strategy to differing degrees depending on the ambient language, we analysed word initial cluster production acoustically in nine Norwegian and nine English speaking children aged 2;6–6 years, and eight adults, four from each language. The results showed that Norwegian-speaking children produce significantly more instances of vocalic insertions than English-speaking children do. The same pattern is found in Norwegian- versus English-speaking adults. We argue that this cross-linguistic difference is an example of the influence of prosodic-phonetic biases in language-specific developmental paths in the acquisition of speech.

Introduction

When infants learn to speak, they need to master a complex combination of knowledge and skills simultaneously. Their developmental path will be shaped by universal constraints on speech production and perception imposed by vocal tract dynamics, audition, and neurophysiology, all of which are still very much in flux in the developing child (Bernthal & Beukelman, 1978; Stathopoulos & Sapienza, 1993; Koenig, 2000; Imbrie, 2005). Thus, for instance, any infant will struggle to produce consonant clusters, since clusters require a precision of gestural coordination which will initially exceed the capabilities of the child's maturing vocal tract and motor control. In addition, the infant's developmental path will be determined by language-specific structure, when exposure to the particular distributional frequencies of the ambient language influences whether and when certain structures are acquired (e.g. Vihman & Velleman, 2000; Prieto Vives & Bosch-Baliarda, 2006) as the infant capitalizes on that language's statistical properties to learn its structures (Saffran, 2003).

A type of language-specific structural constraint which has been largely overlooked in previous research on acquisition is a specific language's patterns of gestural co-ordination, resulting from biases (or dominant tendencies) in how adult speakers phonetically implement the phonological contrasts and structures of that language (Payne, Post, Garmann, & Simonsen, 2015; Payne, 2016). Such 'prosodic-phonetic biases' are language-specific, subphonemic tendencies that appear repeatedly (but not categorically) in speech, constituting salient characteristics of a language without being contrastive. In this paper, we examine the relative presence, and impact, of one such prosodic-phonetic bias in Norwegian and English, languages that, though phonotactically similar, exhibit interesting micro-variation in consonant clusters' timing and coordination. Cluster acquisition is of particular interest because both reduction and vocalic insertion[1] are reported cross-linguistically as strategies for cluster production in early speech, albeit to different extents across languages.

Vocalic insertion in child and adult language

In a review article of children's acquisition of consonant clusters across languages, McLeod et al. (2001) looked at the kind of errors children make in the acquisition process. In spite of substantial variation, they found that across languages the most common error pattern was 'cluster reduction', which involves the omission of at least one of the consonants in the cluster. The second most common error pattern was 'cluster simplification', which refers to cases where all of the consonants in the target word are produced, but one or all of the elements are non-target-like in their phonetic implementation. Other processes such as 'coalescence', 'metathesis' and 'epenthesis' were found to be less frequent. McLeod et al. (2001) based their overview mainly on English, but also included other languages, for example Dutch, Danish, Italian, Telugu, German, Spanish, Cantonese, Portuguese, and Turkish.

Bernhardt and Stemberger and collaborators (Stemberger & Bernhardt, 2018; Bernhardt & Stemberger, 2018) compared the production of word initial liquid clusters in preschoolers aged 3-5 years across several languages: Icelandic (Másdóttir, 2018), Swedish (Lundeborg Hammarström, 2018), Portuguese (Ramalho & Freitas, 2018), Spanish (Perez, Vivar, Bernhardt, Mendoza, Ávila, Carballo, ... & Vergara, 2018), Bulgarian (Ignatova, Bernhardt, Marinova-Todd, & Stemberger, 2018), Slovenian (Ozbič, Kogovšek, Stemberger, Bernhardt, Muznik, & Novšak Brce, 2018), and Hungarian (Tár,

2018). For most of the languages studied, there was either no evidence of epenthesis as a strategy for cluster production (Swedish, Icelandic) or epenthesis was found to be a relatively infrequent mismatch pattern (Spanish, Slovenian, Hungarian). However, it is worth noting that for Hungarian, earlier studies of rhotic clusters in adult speech have found a high occurrence of vocalic insertion (Gósy, 2008). Tár (2018) took this pattern into account when identifying mismatch patterns in Hungarian child speech, so the actual occurrence of vocalic insertion in children may be somewhat higher. For Bulgarian, epenthesis was reported as a relatively prominent pattern for three-year-olds (Ignatova et al., 2018), and for Portuguese, epenthesis was moderately frequent (around 10%) across the ages 3-5 (Ramalho & Freitas, 2018). Overall for this project, the researchers concluded that "... additional research is needed, particularly with careful measurement of duration of vowel-like elements in both child and adult speech, and careful consideration of quantitative properties of the phenomenon at all ages" (Bernhardt & Stemberger, 2018, p. 570). In other words, judgment about the phenomenon in child speech presupposes an understanding of what is happening in adult speech from the relevant language, and an appreciation of the cross-linguistic variation in adult speech.

Several studies have studied cluster production in Norwegian child language (Vanvik, 1971; Simonsen, 1990; Kristoffersen & Simonsen, 2006; Yavaş, Ben-David, Gerrits, Kristoffersen & Simonsen, 2008; Simonsen, Garmann & Kristoffersen, 2019) and data from Simonsen (1990) and Simonsen, Garmann & Kristoffersen (2019) indicate that while both cluster reduction and cluster simplification are common error patterns, epenthesis is also a prevalent cluster simplification strategy. For example, the target word *trikk* 'tram' [ˈtɾɪcʰ(ɛn)] could be pronounced [ˈtʰɪcʰɛn](with omission of the *r*), [ˈtði:cʰ] (with ð substituting the r), and [tʰiˈlɪcʰ] (with an epenthesis as well as a substitution) (Tomas, 2;0 in Simonsen, 1990). In the three children from the Simonsen (1990) study, epenthesis constitutes between 16% and 32% of their cluster errors in a period around age 2 – while reduction varies between 25% and 65% and simplification varies between 13% and 35%. So while the studies mentioned in McLeod et al (2001) have vocalic epenthesis at a rate of 2.5%-7.2% across a range of languages, Norwegian children appear to use this strategy much more.

General descriptions of Norwegian (Endresen, 1991, p. 127) indicate that clusters – in particular heterorganic ones – have an open transition (i.e. there is a clear audible release of the first consonant before the closure of the second, see Catford, 1988) while clusters in English more commonly have a close transition, often characterised with gestural overlap and no audible release of the first segment in the cluster (Ladefoged & Maddieson, 1996, p. 329; Catford, 1977, p. 220-226; Gafos, 2002). Consonant overlap in English clusters has been extensively studied (e.g., Catford, 1977; Hardcastle & Roach, 1977; Hardcastle, 1985; Barry, 1991; Browman & Goldstein, 1990; Nolan, 1992; Zsiga, 1994; Byrd, 1996; Byrd & Tan, 1996) showing significant overlap in articulation for all consonant sequences within words and across word boundaries.

Rather than being characterised as a binary distinction of open or close, the transition between consonants in a cluster seems to be gradual. There is evidence from other languages as well that there is cross-linguistic variation in the way that the individual gestures of a cluster sequence are phased, and that consonant clusters are produced with different degrees of articulatory timing lags in different languages (Kwon &

Chitoran, 2016). As Davidson (2005) reports, these differences result from language specific detail concerning gestural coordination (see also Browman & Goldstein, 1992b). Gafos (2002) and Gafos & Goldstein (2012) develop a grammar of gestural coordination to account for and model these language-specific patterns, which is one possible way of framing such cross-linguistic differences.

We instead propose a different framework, within a usage-based approach, where such cross-linguistic differences are seen as the results of so-called prosodic-phonetic biases, a broader category of sub-phonemic phenomena. These biases neither arise from universal phonetic (physical or neural) constraints, nor do they reflect cross-linguistic differences in phonological contrasts. Instead, they can be explained as in Payne (2016, p. 191): "Through the iterative, constraining influence between existing structure and speech behaviour, apparent structural 'conspiracies' may arise, with a particular structure generating phonetic patterns that then reinforce that very same structure." As such, patterns, or biases, in gestural behaviour may take root and be reinforced sub-phonemically, resulting in a broader gestural and auditory coherence.

In the case of stop consonants, the open transition associated with Norwegian clusters is coherent with the greater extent of audible release in singleton consonants in coda position, whereas the closed transition characteristic of English clusters is coherent with unreleased coda stops. Another example of a prosodic-phonetic bias is the contrast between voiced and unvoiced stops, which can be realised with gradual differences in coordination and timing of gestures in different languages. During the process of acquisition, the child encounters the phonetic variation in these realisations Based on the input, she can abstract both what constitutes a meaningful contrast and the boundaries for the (non-meaningful, but systemic) variation within the given language. Prosodic-phonetic biases are compatible with, but not limited to a purely gestural model, invoking as they do considerations also of auditory coherence. Irrespective of the type of model, they are part of the body of knowledge that a child must acquire if she is to be perceived as a native speaker. Cluster production presents an interesting testing ground for our claim that prosodic-phonetic biases also play a crucial role in shaping cross-linguistically divergent paths of speech development in infants.

Vocalic insertion, intrusion and epenthesis

In Articulatory Phonology, consonantal articulations are seen as superimposed on the tongue body gesture of the vowel. Thus, what appears to be insertion of a vowel is actually a result of "exposure" of the (already existing) vowel gesture as a result of a lack of overlap during an open transition between adjacent consonant gestures (Browman & Goldstein, 1991, p. 371; Steriade, 1990, p. 390).

This interpretation is schematically illustrated in Figure 1 which compares the gestural scores of the Norwegian target *blå* 'blue': [^{1}bɭɔː] with the production [bə^{1}ɭɔː] in the child Nora (2;8) (Simonsen, 1990).

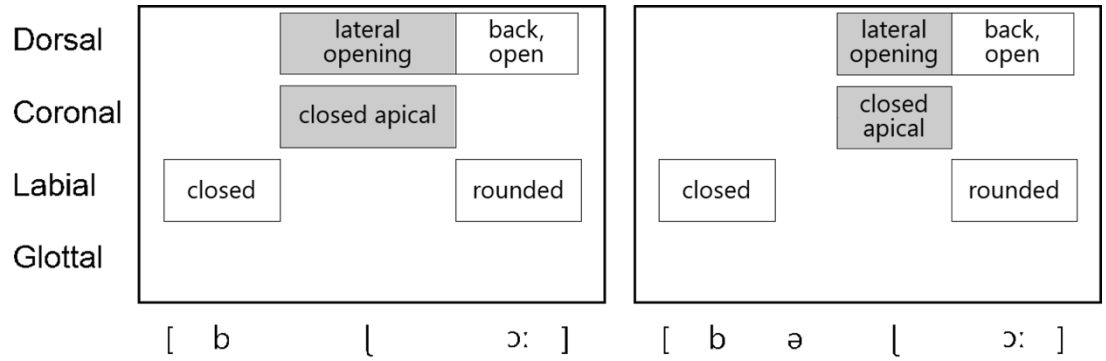

Figure 1: Illustration of how apparent vocalic insertions in Norwegian child speech may be conceived of and represented as «exposed» vocalic gestures within an Articulatory Phonology account. The affected gestures are marked in grey.

The illustration in figure 1 is inspired by Browman and Goldstein (1992a, p. 158). The inserted, or apparent, schwa-like vowel is modelled as the result of a neutral tongue and jaw position. Since the description is underspecified and voicing is seen as the default in speech production, the vowel is not specified on the glottal tier. The reason why a vocalic 'event' emerges *acoustically* is that there is no overlap between the closed lip gesture for /b/ and the apical closure for /ɭ/, and hence a default vocalic configuration of the vocal tract occurs, with associated vocalic acoustic output. The apical gesture is moved forward in time, and the vocalic gesture is 'exposed'.

In the relevant literature on this phenomenon, we find considerable terminological variation, and this variation sometimes, though not always, hints at different phonological statuses afforded to the vocalic event, or at least different interpretations of phonological status. In the child language literature, the term 'epenthesis' is generally used to refer to all vocalic insertions between consonants in a cluster. However, Stemberger and Bernhardt (2018) point to the difficulty of distinguishing between 'true' vocalic epenthesis and the insertion of transitional vowel-like elements in child speech, arguing for the necessity of measuring the duration of these elements and comparing them to the duration of short unstressed vowels.

The need to distinguish between two distinct types of vocalic insertions – epenthesis and vocalic intrusion – is also highlighted by Hall (2006). In her view, epenthesis is to be seen as a categorical phenomenon, whereas vocalic intrusion is more gradient and variable. Building on Articulatory Phonology (e.g. Browman & Goldstein, 1992a), she argues that intrusive vowels are the phonetic result of reduced overlap or retiming of adjacent consonant gestures. The retiming of gestures results in an acoustic release following the first consonant, which may be interpreted as a vowel. Depending on the presence and degree of voicing, the duration of the release and the position of the tongue during the vocalic interval, the "vowel" is perceived as a schwa or a copy of one of the surrounding vowels, and not as a lexical vowel in its own right. Intrusive vowels are gradient, likely to be optional, have a variable duration and may disappear at high speech rates, and they are typically found in heterorganic clusters and in homorganic clusters involving taps and flaps. Epenthesis, by contrast, is categorical and its presence is not dependent on timing or speech rate. It is also phonologically visible in the sense that it affects phonological patterns like stress assignment and syllabification.

Epenthetic vowels may sound like lexical vowels, or they may be schwas or similar to surrounding vowels. Native speakers are more likely to be aware of epenthetic than intrusive vowels in their own speech. Hall (2013) modifies this distinction somewhat concerning epenthetic vowels in Lebanese Arabic: Here, epenthetic vowels do not affect stress assignment, and their presence may vary within and among speakers. Thus, epenthetic vowels are less categorical than lexical vowels, but more categorical than vowel intrusions.

Bradley (2007, p. 964), looking at word medial or final /r+C/clusters in Norwegian also appeals to an Articulatory Phonology framework to explain why an open transition between consonants allows for a vowel to appear, while a close transition prevents apparent vocalic insertion. He also notes (2007, p. 956) that according to Hall (2003, p. 28) vocalic insertion cross-linguistically occurs more frequently with liquids (in the position of C2 in the cluster) than with other sonorants, and more with rhotics than with laterals (except when the rhotic is an alveolar trill). The rhotic tap [ɾ], being one of the shortest segments cross-linguistically, would also appear to be particularly prone to triggering vocalic insertion in clusters. Walsh (1997) argues that this may be to enhance perceptibility and to maintain sonority, however there are also good articulatory and aerodynamic reasons why a vocalic interval appears before the articulation of an alveolar tap, as the latter requires a ballistic motion of the tongue tip, which is facilitated by a 'run-up' vocal tract configuration which is open, and thus vocalic-like.

There is, indeed, wider evidence that the articulatory and phonatory details of the consonantal elements of the cluster in question may play a decisive role in the rate and properties of vocalic insertion. Studies suggest that vocalic insertion occurs more frequently in voiced clusters than in voiceless clusters (e.g. Davidson, 2000). In studies on repetition of non-native clusters, Davidson (2010), Wilson and Davidson (2013), and Davidson and Wilson (2016) also found a higher incidence of vocalic insertion after stops than after fricatives, and a higher incidence after voiced stops than after voiceless stops.

In this study, we investigate firstly the extent to which systematic differences in transitions in onset clusters can be identified between Norwegian and English adult speech. While cross-linguistic differences in cluster coordination are already attested for a variety of languages, the comparison between English and Norwegian is of particular interest because, in formal phonological terms at least, the two languages show very similar phonotactic constraints. Examining the phonetic implementation of putatively "identical" clusters in the two languages can provide greater insight into the nature of phonotactic constraints more generally.

Although the clusters chosen are comparable between the languages at the phonological and phonotactic level, there are some phonetic differences. The most important difference is that occurring between the realisation of rhotics in the two languages: while the Southern British English production of the rhotic is a postalveolar approximant [ɹ], the East Norwegian production of the rhotic is an apical tap [ɾ] – which is very short, and for both perceptual and articulatory reasons facilitates a vocalic insertion. A further difference is to be found in the non-liquid approximant: the Norwegian equivalent of English [w] is the (unrounded) approximant [ʊ]. We do not, however, expect this slight difference in articulation to cause a substantive difference in

the gestural coordination of clusters. As for the laterals in these clusters, the Southern British English realisation is an apicolaminal [l], while the East Norwegian one is an apical [l̻]. As a consequence of coarticulation, the Eastern Norwegian initial *sl*-clusters are apical in both segments [ʂl̻], in contrast with the English [sl]. Finally, in Norwegian, for some speakers, the clusters /pl, bl, kl, gl, fl/ may be pronounced with an apical flap [ɾ̻] instead of [l̻], thereby shortening the duration of the second consonant and thus the likelihood of vocalic insertion. (For a detailed description of Norwegian liquids and approximants see Moen, Simonsen Lindstad & Cowen, 2003; Simonsen, Moen & Cowen, 2008; Kristoffersen, 2000:25.).

If there are differences in cluster transition between English and Norwegian, as suggested in the literature, we would expect children to attempt to replicate these different cluster transitions. Infants also speak more slowly and produce different patterns of overlap than adults, at least when they are 2 or 4 years old (Payne, Post, Garmann & Simonsen, 2017). So, provided that the transition in Norwegian is more open than in English, we would expect Norwegian children to have *longer* vocalic insertions than English children, and for children to have more vocalic insertion than adults overall. Furthermore, during acquisition children may overgeneralise patterns that are salient in their input, increasing the probability of a high incidence of vocalic insertions, both in environments where it is found in adult speech, and possibly even in other environments. Thus, we predict vocalic insertions to have a higher incidence in children's speech in Norwegian than in English, and to be longer occur more frequently and appear in more contexts, than in Norwegian adults' speech.

Our overarching research question is the following: How do the differences in transition between English and Norwegian materialise in initial consonant clusters in adult speech, and what are the possible consequences for children's speech in the two languages?

As a first step in answering this question, we investigate comparable consonant cluster productions in English and Norwegian adult speakers, to see to what extent the alleged differences in transition between the two languages are found across a range of cluster types. We focus on initial clusters, and investigate the possible variation in terms of place and manner of articulation.

Our hypotheses for adult speech are:

1. Norwegian clusters exhibit a clearer release in the first consonant and a higher incidence of vocalic insertion between the first and the second consonant than do English clusters.
2. Clusters are longer in Norwegian than in English (because of additional duration of the intervening vocalic insertion).
3. If the release/vocalic insertion is counted as part of the C1, C1 is longer in Norwegian than in English.
4. There is a higher incidence of vocalic insertion if (in order of importance)
    a) C2 is a liquid, and greatest incidence if C2 is a rhotic tap or flap
    b) C1 is a stop
    c) C1 is voiced
5. Vocalic insertion in Norwegian is optional and of varying duration.

As a second step, we investigate the production of the same initial consonant clusters in the speech of children acquiring English or Norwegian. We expect all children to produce some clusters that deviate in some respects from the adult target, but predict that they will adopt different types of strategies depending on the ambient language.

Our hypotheses for children are the following:

1. All children will exhibit strategies like cluster reduction (omission of one element in the cluster) and cluster simplification (distortion of one of the elements in the cluster), but Norwegian-speaking children will have more vocalic insertion than English-speaking children.
2. Concerning vocalic insertion, the children will mirror the adult patterns observed for their ambient language.
3. Due to articulatory restrictions, the vocalic insertions will be of longer duration in children's than in adults' speech.

Method

Adult study

Four Southern British English speaking and four East Norwegian speaking female speakers (in the age range of 25-45 years) were given a list of approximately 30 sentences designed to contain words with a variety of consonant clusters. The clusters were chosen so as to be comparable between the two languages. The participants read through the list first, and were then asked to read the sentences at a normal speech rate. In the Norwegian setting, the participants were recorded in their own homes with a portable digital recorder (Zoom H2 Handy Recorder). In Britain, the participants were recorded either in their own home or in the home of the experimenter, using a portable Marantz PMD660 recorder and Shure PGb1 microphones. In both datasets, the adults were mothers of the children studied (see below).

The list contained words with a selection of initial consonant clusters with stops and fricatives as the first consonant and stops, approximants and liquids as the second consonant. Based on the findings in Norwegian child data (Simonsen, 1990), our main interest was stop + liquid clusters, but we also looked at stop plus non-liquid approximant clusters, and fricative clusters.

Table 1

Context		Clusters	No. of words in NOR	No. of words in ENG
Stop + liquid	Stop + /r/	/pr, br, tr, dr, kr, gr/	56	56
	Stop + /l/	/pl, bl, kl, gl/	32	36
Stop + non-liquid approximant		/kw, tw/	12	12
Fricative (non-s) + liquid		/fl, fr/	16	16
S-clusters		/sp, st, sk, sn, sm, sw, sl/	72	58
Total			188	178

Table 1. Number of word tokens with consonant clusters according to cluster type in the adult data.

For the adults, we analysed the productions of 188 Norwegian and 178 English word tokens distributed on the clusters as illustrated in Table 1. In the Appendix, a more detailed distribution is given in Table I, and a full list of the stimuli is found in Table III.

To control for the possibility that speech rate could influence the number or duration of vocalic insertions, we estimated the speech rate through sampling 4 sentences of 13–23 syllables each from each of the adult participants. For each of these sentences, we measured the average syllable duration in milliseconds.

In order to be able to evaluate the impression of the vocalic insertion in adult speech, the length of the vocalic insertions found in consonant clusters was compared to the same speakers' production of unstressed vowels. A set of sentences containing one or more unstressed vowels between two consonants (some within and some across word boundaries) were chosen from the list of already recorded sentences. As we only found vocalic insertions in the Norwegian adult data, these measurements have only been done in the Norwegian sample. For two of the Norwegian adults, the recorded production of each of 12 selected vowels was measured, whereas for the other two, we measured 11 unstressed vowels. A list of the words included can be found in Table IV in the appendix.

Child study

Nine English-speaking children and nine Norwegian-speaking children took part in the study. There were three children in each age group at 2.5 years, 4 years, and 6 years for each language. (Age range for the Norwegian-speaking children: 2;5-2;7, 4;1-4;4, 5;6-6;5, age range for the English-speaking children: 2;6-2;10, 4;4-4;11, 6;1-6;6). The children were recruited through personal networks, all were typically developing children. The child was shown a picture story on a screen, while the mother read the story to the child. The instructions were that the mothers were free to tell the story as they wished, as long as they used the target words in the text. Following this, pictures from the story were shown again and the mother asked the child to name the pictures.

The pictures in the task referred to words with more or less the same clusters as the sentence list for the adults, although some clusters were missing, namely /pr, dr, kw, tw, sl/ due to difficulties in finding imageable words known to children for both languages. The remaining clusters included in the child study are listed in Table 2.

Table 2

Context		**Clusters**	**No. of words in NOR**	**No. of words in ENG**
Stop + liquid	Stop + /r/	/br, tr, kr/	40	27
	Stop + /l/	/pl, kl, gl/	24	36
Fricative (non-s) + liquid		/fl, fr/	25	18
S-clusters		/sp, st, sk, sn, sm, sw/	67	63
Total			156	144

Table 2. Number of word tokens with consonant clusters according to cluster type in the children's data.

Table 2 shows that for the children, we analysed a total of 156 Norwegian and 144 English word tokens. Table II in the Appendix shows in more detail the distribution of word tokens by age group, language and consonant cluster, and the actual stimuli are found in Table III.

Analyses

An open transition or, in Articulatory Phonology terms, a low level of articulatory overlap between consonants, may produce acoustic effects that may or may not sound vowel-like. As mentioned by Hall (2006, p. 413): "Often, languages that have vowel intrusion in some consonant clusters have effects described as aspiration or consonant syllabification in other consonant clusters. All of these phenomena may be attributed to low gestural overlap. Aspiration between consonants can be seen as a kind of voiceless intrusive vowel." Due to these factors, the substance of the transition may be difficult to interpret.

Furthermore, the analysis of speech for evidence of vocalic insertions may not be straightforward. Productions are sometimes difficult to segment phonetically: articulatory gestures overlap in varying ways that do not always result in unambiguous discontinuities in the acoustic signal. The identification of what might be classified as vocalic insertion needs to take into account multiple parameters in the acoustic domain, and in particular discontinuities in amplitude, presence of voicing, patterns of formant structure and the duration of any of these properties. Any of these may be present to a greater or lesser extent, and in different combinations. At one end of the scale, there may be clear cases of vowel insertion, with unambiguous changes in amplitude, presence of voicing and a well-defined vocalic formant structure for a sustained duration; at the other end, there will be productions that contain no evidence at all. In between, there will be relative degrees of evidence, including cases where there may be the appropriate

articulatory conditions for vocalic insertion, i.e. a gestural lag in the oral articulation of C1 and C2, but the acoustic evidence is obscured by aperiodicity, and it is impossible to tell from acoustic evidence alone. This is particularly likely where C1 is a voiceless plosive, and there may be a positive VOT for a following voiced C2. With a gradient phenomenon, any imposition of realisation types may seem in some respects contrived. However, because vocalic insertion is not identifiable along a single articulatory parameter, a combination of auditory and visual judgment is necessary, and this presupposes a degree of categorisation. Thus, for the purposes of our study, we have divided the productions into four realisation types: a) clear vocalic insertion, b) relatively clear vocalic insertion, c) possible "masked" vocalic insertion and d) definitely no vocalic insertion (see below for the criteria that were used to distinguish between the realisation types).

The target words were segmented and analysed in spectrograms and waveforms created in Praat (Boersma & Weenink, 2016). To identify vocalic insertion four different types of realisation emerged from the data (see Figure 2):

a. **Clear vocalic insertion**: evident from a clearly definable period of high amplitude, voicing, formant structure, and of an easily measurable duration.
b. **Relatively clear vocalic insertion**: evidence of a post-consonantal period with lower amplitude than for (a), which may be either not fully voiced or with weak formant structure, and shorter in duration (or more difficult to measure) than for (a).
c. **Possible 'masked' vocalic insertion**: segment boundaries are hard to ascertain, e.g. because of a period of post-release aspiration and/or devoicing in an approximant may overlay a vocalic interval.
d. **Definitely no vocalic insertion**: there is no intervening acoustic material or discontinuity between C1 and C2.

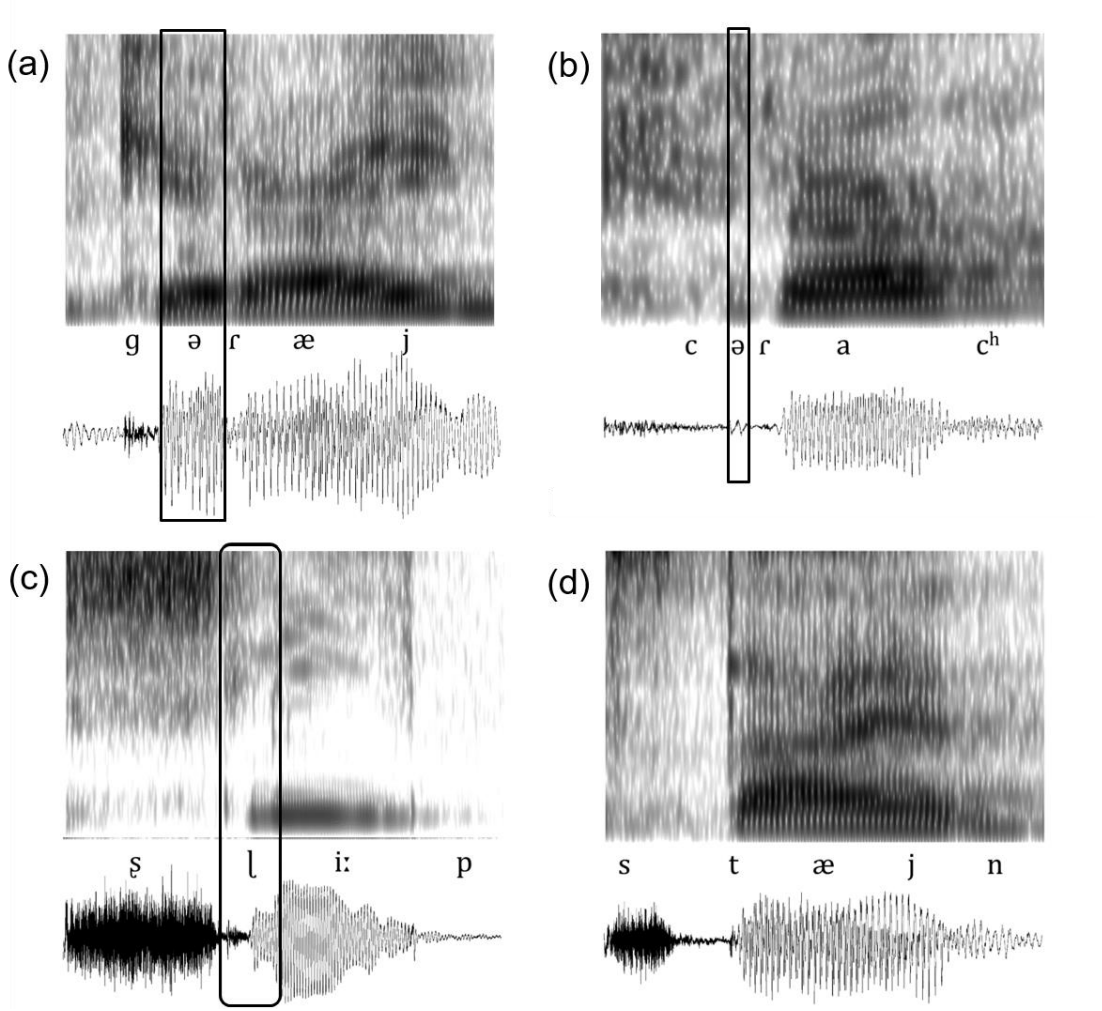

Figure 2: Spectrograms and waveforms for the four different types of realisation of consonant clusters (a, b, c and d), illustrated with the following four Norwegian words: *grei* ('okay')[gə[1]ɾæj], *krakk* ('stool')[cə[1]racʰ], *slipe* ('grind')[[2]ʂɭiːp(ə)] and *stein* ('stone')[[1]stæjn]. The vocalic insertions in (a) and (b) are marked with rectangles, and the possible 'masked' vocalic insertion in (c) is marked with an oval.

It was particularly difficult to identify vocalic insertion in clusters with /l/ as the second consonant because laterals display similar acoustic properties to vowels (voicing and formant structure) and may be relatively long in duration. To count as vocalic insertion before /l/, there had to be a clearer formant structure in the vowel than in the following /l/, with higher energy evident in the higher frequency range, and preferably some discontinuity in amplitude (laterals typically have lower amplitude than vowels), see Figure 3. Sequences of voiceless stops and approximants /pl, kl, tw, kw/ are also difficult to segment because of the post-release aspiration of the voiceless stop, in both languages but especially in English, and by association the tendency for the approximant to devoice, creating an interval of frication which could belong to either, or both. Since it is virtually impossible to determine where the boundary lies in such cases, we did not measure the duration of the individual consonant in these cases, only the duration of the stop closure, and the total duration of the cluster.

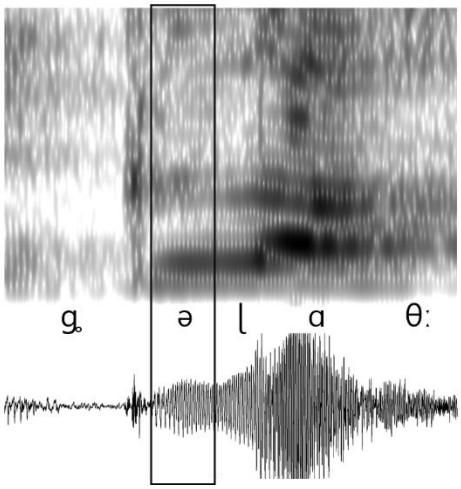

Figure 3: Spectrogram and waveform for a four-year-old child's realisation of a consonant cluster with /l/ as second consonant and a clear vocalic insertion (category (a)), illustrated with the Norwegian word *glass* ('glass')[gə¹l̥ɑθː].

As the results will show, all four realisation types were common in the Norwegian dataset, while the English dataset practically only had realisation type d. The original coders (3 for each language) worked closely together, discussing their ratings with each other and with one of the main investigators when in doubt. Some of the examples were even discussed across the whole cowriter team. Since there were examples of all four realisation types in Norwegian, but very few instances of realisation types a and b in the English data, a reliability scoring was conducted on the Norwegian data only. This was investigated through blind coding of 40 consonant clusters, 20 from the adults and 20 from the children. These were semi-randomly selected, balanced between the four categories, and for the children between the three age groups. The agreement concerning the presence of a vocalic insertion (i.e. an a/b vs. a c/d) was 75%. There was a marginally higher agreement in the child data set (16 of 20) than the adult data set (14 of 20). Looking closer at the four categories described above, there was a clearer agreement for the clusters originally scored as a clear insertion (a) (7 of 10) or definitely no insertion (d) (7 of 10) than for relatively clear (b) (3 of 10) or possible marked insertions (c) (0 of 10). Regarding the selected c category clusters, nearly all (9 of 10) instances were interpreted as definitely no insertion (d) in the blind coding. As for the b category, it was evident that the duration of the inserted vowel played a major role: the original coders accepted durations down to 9 ms as long as other criteria were fulfilled, while for the blind coding no duration shorter than 27 ms was accepted as a b.

To test our hypotheses, we measured the following: Duration of cluster, duration of C1 (if released), duration of release friction of C1 (if present), duration of C2 (if clear boundary), and duration of inserted vowel (if present).

To make a conservative estimate of vocalic insertion, we have in general not included possible masked vocalic insertions (type c) in the calculations where vocalic insertions are involved. This means that the reported incidence of vocalic insertions may be lower than the actual incidence. However, the proportion of type c realisations is roughly the same in the Norwegian and the English data (29% vs. 22%). In view of the low agreement in coding for category b, we also considered excluding those from our

calculations. However, excluding both b and c from the calculations yielded in essence the same significant results as when we only excluded c, so we decided to keep our original calculations for vocalic insertion, with b included.

Statistical analysis

Hypotheses concerning the amount of vocalic insertions (by group and phonetic environment) were investigated through chi-squared ($\chi 2$) tests. Wilcoxon rank sum tests were used to compare durations of clusters, C1 and vocalic insertions across groups, and to investigate the relationship between the duration of the vocalic insertions and the phonetic environment. This method was preferred over t-tests because several of the relevant subsets deviated significantly from a normal distribution. Analyses of how the number and duration of vocalic insertions vary between different phonetic environments involved multiple analyses of the same data set. Here, the Holm-Bonferroni method was used to control the family-wise error rate. All statistical analyses were carried out in R 3.4.4 (R Core Team 2018) using RStudio 1.2 (RStudio Team 2018).

Results

Adult study

Our first hypothesis for adult speech was that Norwegian clusters would exhibit a clearer release in the first consonant and a higher incidence of vocalic insertion between the first and the second consonant than do English clusters. We found that in Norwegian, 30.6% (n=55) of the clusters had a clear or a relatively clear vocalic insertion (realisation types a+b in relation to a+b+c+d), while only one of the clusters in English were of realisation type a (0.6%) and none were of realisation type b. This shows that there is clearly a much higher and statistically significant incidence of vocalic insertion in Norwegian than in English, thus confirming our hypothesis ($\chi 2(1=56.21, p<0.001$). As for realisation type d (no evidence for vocalic insertion), we observed significantly more instances in English than in Norwegian (132 in English vs. 68 in Norwegian, $\chi 2(1)=55.13, p<0.001$), supporting the assumption that in English, it is more common for gestures to overlap (or at least abut). Table 3 gives an overview of the four realisation types by language and clusters.

Table 3

			Norwegian				English			
			Realisation types				Realisation types			
Context		**Clusters**	a	b	c	d	a	b	c	d
Stop+liquid	Stop + /r/	/pr, br, tr, dr, kr, gr/	34	2	10	6	-	-	12	36
	Stop + /l/	/pl, bl, kl, gl/	2	6	11	13	-	-	16	20
Stop+non-liquid		/kw, tw/	1	1	1	5	-	-	4	8
Fricative (non-s)+liquid		/fl, fr/	9	0	2	5	-	-	1	13
S-clusters		/sp, st, sk, sn, sm, sw, sl/	0	0	33	39	1	-	2	55
Total			46	9	57	68	1	-	35	132

Table 3. Number of words with consonant clusters according to cluster type in the adult data.

Our second hypothesis, that clusters are longer in Norwegian (*Mdn*=177 ms, range: 45-387) than in English (*Mdn*=160 ms, range: 79-287), was also confirmed (W=11608, p<0.001). We then looked at whether the clusters with vocalic insertions (realisation types a+b) were longer than clusters without insertions (type d) in Norwegian, but found the opposite to be true: Clusters with vocalic insertion (*Mdn*=158 ms, range: 45-273) in Norwegian are significantly shorter than clusters without insertion (*Mdn*=189 ms, range: 90-387, W=2573, p<0.001, see Figure 4). This somewhat counter-intuitive result will be discussed below.

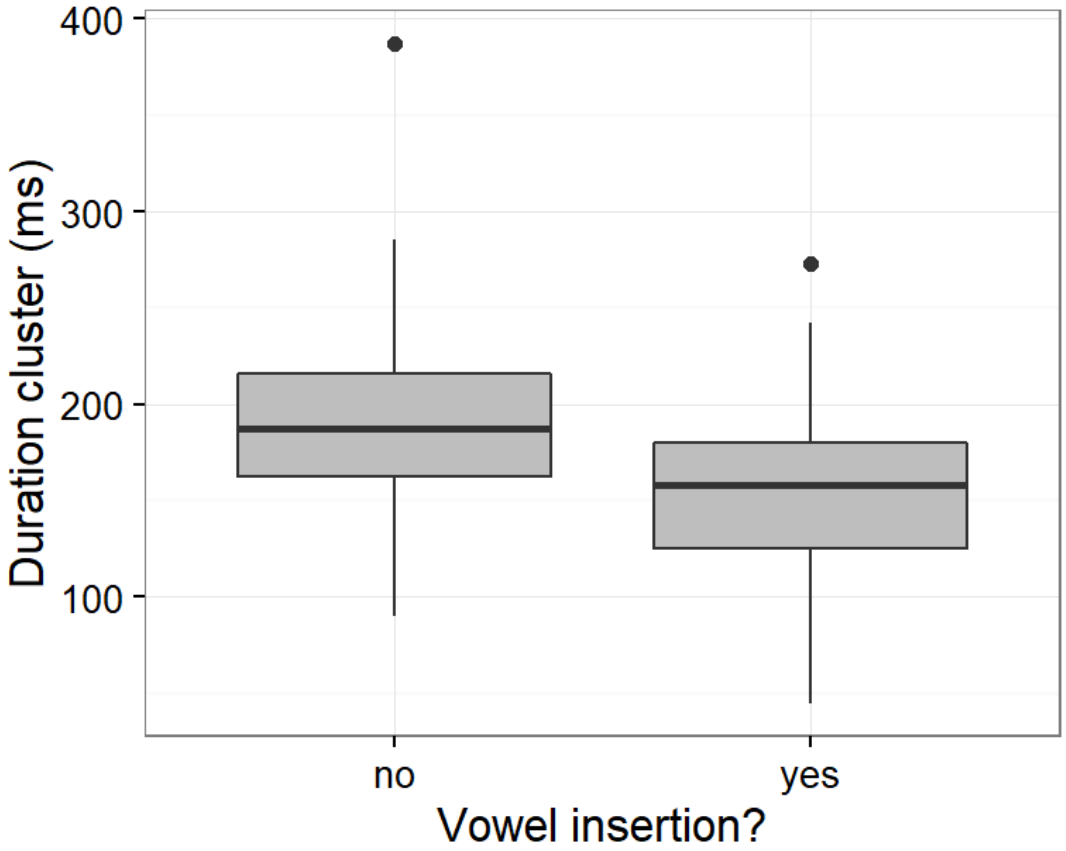

Figure 4: The duration of clusters in adult Norwegian speech without and with vocalic insertions.

Concerning our third hypothesis, that the C1 including the transition (i.e. the release and/or vocalic insertion after C1) would be longer in Norwegian than in English, this is confirmed in the data (Mdn_N=134 ms, $range_N$: 49-359 vs Mdn_E=101, $range_E$=57-209, W=4869, p<0.001). In this calculation, we have excluded the type c productions, as no clear boundary between the two consonants could be established here.

We then checked whether differences in speech rate between speakers of the two languages could explain the cross-linguistic differences in vocalic insertions. Although there was considerable variation in speech rate between the participants, as well as between each of the sentences produced by each participant, no significant difference was found between the languages (W=123, p=0.867). Thus, speech rate cannot explain the clear difference in vocalic insertion between the two languages.

Since there were virtually no vocalic insertions in the English data, our fourth hypothesis concerns only Norwegian. The predictions were that vocalic insertion (realisation types a+b) would be more common a) when C2 was a liquid, and in particular when C2 was a rhotic tap or flap, b) when C1 was a stop, and c) when C1 was voiced.

As for C2, we found that vocalic insertion was seventeen times more common when C2 is a liquid than when it is a non-liquid (χ^2 (1)=46.09, p<0.001). In the clusters where C2 is a liquid, there was vocalic insertion in 51% of the instances, whereas there were

vocalic insertions in only 3% of the clusters (= two instances) with a non-liquid C2. The propensity of vocalic insertion in clusters with C2 as a liquid was mainly carried by clusters with rhotic taps and flaps. If C2 was a rhotic tap or flap, there was vocalic insertion (realisation types a+b) in 71% of the cases, whereas if C2 was another liquid, there was vocalic insertion in only 16% of the instances. This difference is significant (χ^2 (1)=27.47, p<0.001). Thus, vocalic insertion largely, though not exclusively and not always, occurred between a plosive and a tap or a flap.

Looking at the parts of the hypothesis concerning C1, we found that there was vocalic insertion in 50% of the instances with a stop as C1, and in 10% of the instances with a fricative as C1. This difference is significant (χ^2(1)=31.68, p<0.001). Thus, vocalic insertion was much more common when C1 entails a complete obstruction of the vocal tract (as is the case with plosives), suggesting the involvement of aerodynamic constraints. There was also more vocalic insertion in clusters with a voiced C1 than in clusters with an unvoiced C1 (78% vs. 19%, χ^2(1)=44.55, p<0.001).

Until now, we have looked at C1 and C2 separately. Do we get the same results when we look at them in combination? This is an important question, since as exemplified in Table 3, there are different numbers of instances of each cluster type, and the clusters as wholes may behave differently (e.g., /s/-clusters are known to differ from other types of clusters (Yavaş et al, 2008)).

Table 4

		Adults		
Cluster type	**N**	**Clear insertion**	**No clear insertion**	**% with insertion**
voiced stop+tapflap	24	23	1	96
unvoiced stop+tapflap	32	16	16	50
unvoiced fricative+tapflap	10	8	2	80
voiced stop+lateral	12	5	7	42
unvoiced stop+lateral	12	0	12	0
unvoiced fricative+lateral	14	1	13	7
unvoiced stop+non-liquid	12	2	10	17
unvoiced fricative+non-liquid	64	0	64	0

Table 4. Vowel insertion in different cluster types in adults.

As shown in Table 4, in clusters with a voiced stop as C1 and a rhotic tap or flap as C2, there is 96% vocalic insertion. At the other extreme, in clusters with an unvoiced fricative as C1 and a non-liquid as C2 there is no vocalic insertion. These results are both in accordance with all parts of our third hypothesis.

Going into more detail by using Fisher's exact test to compare clusters with C1 voiced stop+C2 tap or a flap and clusters with C1 voiced stop+C2 lateral, there are significantly more vocalic insertions in the first group (p<0.001). Likewise, comparing C1 unvoiced stop+C2 tap/flap with C1 unvoiced stop+C2 lateral, there are significantly more vocalic insertions in the first group (p= 0.002). Similarly, comparing C1 unvoiced fricative+C2 tap/flap and C1 unvoiced fricative+C2 lateral, there is again significantly more vocalic insertion in the first group (p<0.001). All these comparisons confirm that clusters with

taps and flaps are more inducive to vocalic insertions than those with laterals. Turning to comparing clusters with C1 unvoiced fricative+C2 lateral and C1 unvoiced fricative+C2 non-liquid on the one hand, and C1 unvoiced stop+C2 lateral and C1 unvoiced stop+ C2 non-liquid on the other, we find no significant differences. Altogether, these results indicate that vocalic insertions are mainly carried by C2 as tap or flap.

Turning to the manner of articulation of the C1 while controlling for voicedness, we investigated clusters with C1 being a stop or a fricative, respectively, finding significantly more vocalic insertions with C1 as a stop only when C2 was a non-liquid (comparing clusters with C1 unvoiced stop+C2 non-liquid and C1 unvoiced fricative+C2 non-liquid ($p=0.023$)). This difference is mainly carried by the high number of /s/-clusters in the latter group, where no vocalic insertions were found. There was no difference between groups with C1 as stop or fricative when C2 was a tap/flap or a lateral.

Effects of voicing in C1 can only be investigated in clusters with stops, since there are no voiced fricatives in Norwegian. Comparing clusters with C1 voiced stop+C2 tap/flap and C1 unvoiced stop+C2 tap/flap, there is significantly more vocalic insertion when C1 is voiced ($p<0.001$), and the same is found when comparing clusters with C1 voiced stop+C2 lateral and C1 unvoiced stop+C2 lateral ($p=0.037$).

Our fifth hypothesis was that vocalic insertion in Norwegian is optional and of varying duration and can be counted as vocalic intrusion rather than true epenthesis. The Norwegian adults used vocalic insertion occasionally (30.6% on a group level), but none of the adults used vocalic insertions with any phonological consistency. The four Norwegian adults differed in speech rate (4.2–6.0 syllables per second), but all produced between 14 and 16 vocalic insertions, with no correlation between speech rate and the number of or length of insertions.

The durations of the vocalic insertions varied from 9 to 54 ms, with a median of 22 ms. The length of vocalic insertions differed slightly between speakers (with medians ranging from 19 to 24 ms), but the variation was much larger within than across speakers. To see to what extent the vocalic insertions are similar to or different from other unstressed short vowels, we compared the duration of the vocalic insertions with the duration of a set of 46 unstressed short vowels from the same set of speakers. The duration of these unstressed short vowels varied between 25 and 95 ms, with a median of 51 ms. Although the vocalic insertions (*Mdn*=22 ms) were significantly shorter than the unstressed vowels ($W=152$, $p<0.001$), there was an overlap in duration between them: 57% of the measured unstressed short vowels were shorter than the longest vocalic insertion, and 38% of these insertions were longer than the shortest unstressed vowel.

Child study

We hypothesized that all children, across the two languages, would exhibit strategies such as cluster reduction (omission of one element in the cluster) and cluster simplification (modification of one or more of the elements in the cluster), but that the Norwegian children would have more vocalic insertion than the English children. Figure 5 illustrates how both the Norwegian and the English children across the three age groups (aged 2–6 years) displayed all of the three cluster production strategies, but that

the Norwegian children displayed a considerably higher incidence of vocalic insertion than the English children.

Among the cluster productions of the English children, there was no evidence of vocalic insertions at all at two years. There was one instance (i.e. 2% of the total productions) at four years and 6 (13%) at six years, all of the b-type, i.e. relatively clear vocalic insertions. The Norwegian children showed 9 instances (22%) of vocalic insertion at two years (of which 5 were of the a-type), 23 (43%) at four years (of which 15 where of the a-type) and 15 (28%) at six years (of which 6 were of the a-type). The proportion of vocalic insertions increased when reduction disappeared or was sharply reduced. This is obvious, since the children have to produce two consonants in order to have a vocalic insertion between them.

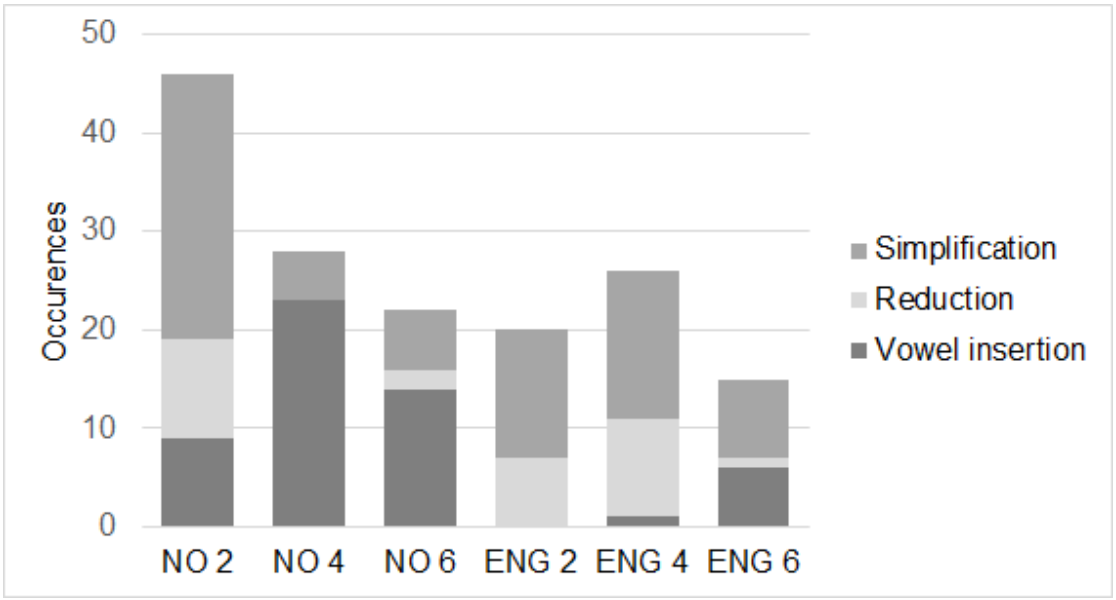

Figure 5: Number of vocalic insertions, reductions and simplifications in the cluster productions of Norwegian and English speaking two-, four- and six-year-olds.

We then hypothesized that the children would mirror the adult patterns. This hypothesis was supported by the data: The Norwegian children as a group had a vocalic insertion in 31.7% (n=47) of the targeted consonant clusters, comparable to the Norwegian adults' 30.6% (realisation types a+b in relation to a+b+c+d). The English children inserted vowels in only 5.6% (n=7) of their targeted consonant clusters. This is more than the English adults' 0.6%, but far less than the Norwegian children. The cross-linguistic difference between children is significant (χ^2=27.294, df = 1, p<0.001).

These figures indicate that the children do reproduce the adult patterns, but note that if we exclude the clusters they reduced to a single consonant (with no possibility for an insertion), the proportions are higher among both groups of children. The Norwegian children inserted a vowel in 54.8% of the clusters they produced as clusters; the English

children 15.4%. There is nevertheless still a notable difference between the two language groups.

Table 5

			Norwegian				English			
			Realisation types				Realisation types			
Context		**Clusters**	**a**	**b**	**c**	**d**	**a**	**b**	**c**	**d**
Stop+liquid	Stop+/r/	/pr, br, tr, dr, kr, gr/	13	11	12	2	-	2	9	13
	Stop+/l/	/pl, bl, kl, gl/	3	2	17	-	-	3	9	16
Fricative (non-s)+liquid		/fl, fr/	4	4	14	3	-	2	2	13
S-clusters		/sp, st, sk, sn, sm, sw, sl/	6	4	3	50	-	-	8	47
Total			26	21	46	55	-	7	28	89

Table 5. Number of words with consonant clusters according to cluster type in the child data.

Table 5 gives an overview of the four realisation types by language and produced clusters. Since the English participants still produced far fewer vocalic insertions (one in the adults and seven in the children), the following paragraphs will focus on Norwegian only. In Norwegian adult speech, we found more vocalic insertion when C1 is a stop; the same tendency was found among the Norwegian children: 45.3% of the consonant clusters with stops as a C1 had vocalic insertion, while 18.5% of the clusters with C1 as a fricative had vocalic insertion. This difference is significant ($\chi^2(1)=11.813$, df=1, p=0.001).

We also found more vocalic insertion in the Norwegian adults' speech when C1 was voiced, and again the same tendency was found for the children: 62.5% of clusters with a voiced C1 had a vocalic insertion, while 21.0% of clusters with an unvoiced C1 did. The difference is significant ($\chi^2(1)=19.151$, $p<0.001$). Whereas the Norwegian adults produced longer vocalic insertions after a voiced C1, there was no difference in the duration of vocalic insertions after voiced and unvoiced C1s among the Norwegian children (Mdn_V=43 ms, range$_V$: 8-103, Mdn_U=47 ms, range$_U$: 18-162, W=243, p=0.694).

Similar to the adults, who showed more vocalic insertion when C2 was a liquid, and most if C2 was a rhotic tap or flap, the Norwegian children produced vocalic insertion in 41.6% (n=37) of the clusters with a liquid C2, and only in 13.4% (n=9) where C2 was non-liquid. Even though this difference is significant ($\chi^2(1)=13.236$, p<0.001), the children produced more insertions in clusters with non-liquids as C2 than adults did (but carried by one cluster type only:[sʊ]). Among the liquids, vocalic insertions were far more common when C2 was a tap (57.1%, n=28) than when C2 was a lateral (22.5%, n=9, $\chi^2(1)=9.502$, p=0.002). (None of the children produced flaps.) When we consider the few vocalic insertions produced by the English children (seven instances in total), they all had a liquid as C2.

Table 6

		Children		
Cluster type	**N**	**Clear insertion**	**No clear insertion**	**% with insertion**
voiced stop+tapflap	23	16	7	70
unvoiced stop+tapflap	17	8	9	47
unvoiced fricative+tapflap	9	4	5	44
voiced stop+lateral	9	4	5	44
unvoiced stop+lateral	15	1	14	7
unvoiced fricative+lateral	16	4	12	25
unvoiced stop+non-liquid	0	NA	NA	NA
unvoiced fricative+non-liquid	67	9	58	13

Table 6. Vowel insertion in different cluster types in children.

Table 6 shows the amount and proportions of vocalic insertions in the cluster combinations included in this study. Comparing the figures in tables 4 and 6, we can observe that similarly to the adults, the children have the most vocalic insertions in clusters with a voiced stop as C1 followed by a tap or a flap as C2. Unlike the adults, the children have some instances of vocalic insertions in all attested cluster types, indicating that indeed, vocalic insertion is a common strategy in children's cluster productions.

As is apparent in the table, the children are less consistent than the adults, and only two differences between cluster types were significant: 1) There are significantly more vocalic insertions in clusters with C1 unvoiced stop + C2 tap or flap than in clusters with C1 unvoiced stop + C2 lateral. 2) There are significantly more vocalic insertions in clusters with C1 voiced stop + C2 lateral than with C1 unvoiced stop + C2 lateral. For the other comparisons where we found significant differences in the adult data, the results go in the same direction, although not to a significant degree. Altogether, this suggests that the children's cluster patterns mirror those of the adults to some degree.

Our third hypothesis was that due to children speaking more slowly than adults, as well as having difficulties with articulatory timing (Payne et al., 2017), the vocalic insertions would be longer in children's than in adults' speech. Figure 6 confirms this expectation, as vocalic insertions have a significantly longer duration in the Norwegian children's words (*Mdn*=43, range: 8-162) than in the Norwegian adults' (*Mdn*=22, range: 9-54, W=413, p<0.001).

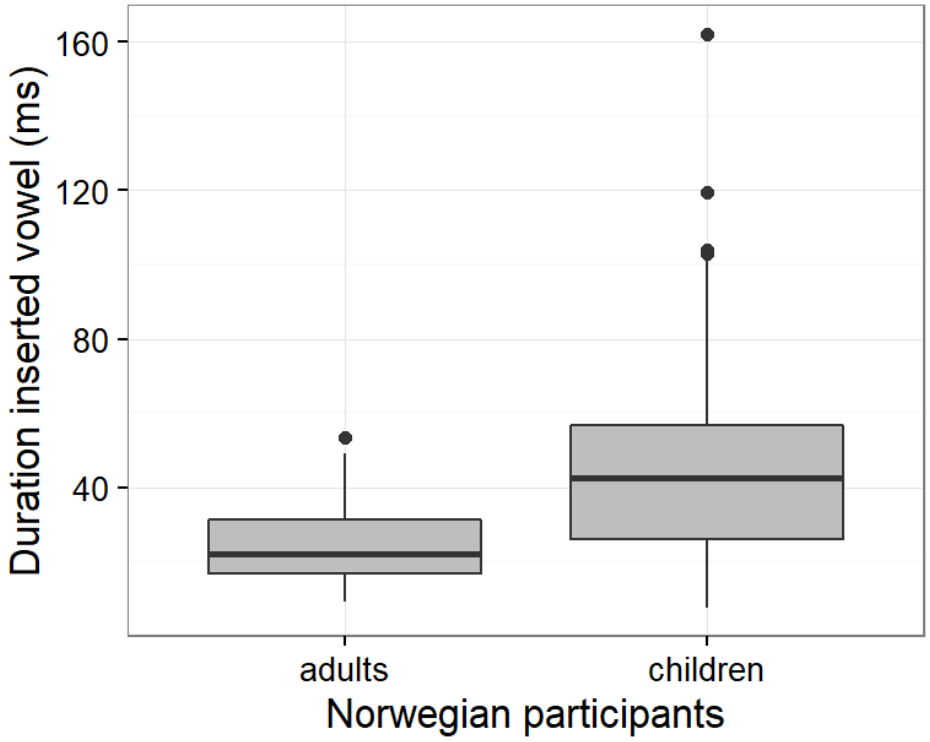

Figure 6. Duration of the vocalic insertions produced by the Norwegian adults and children.

For the adults, we found that the C1 including the transition (i.e. the release and/or vocalic insertion after C1), was longer in Norwegian than in English. As illustrated in figure 7, the same pattern was found in the children (Mdn_N=143, $range_N$: 19-514, Mdn_E=114, $range_E$: 8-276, W=1977, p<0.001), only with larger variation.

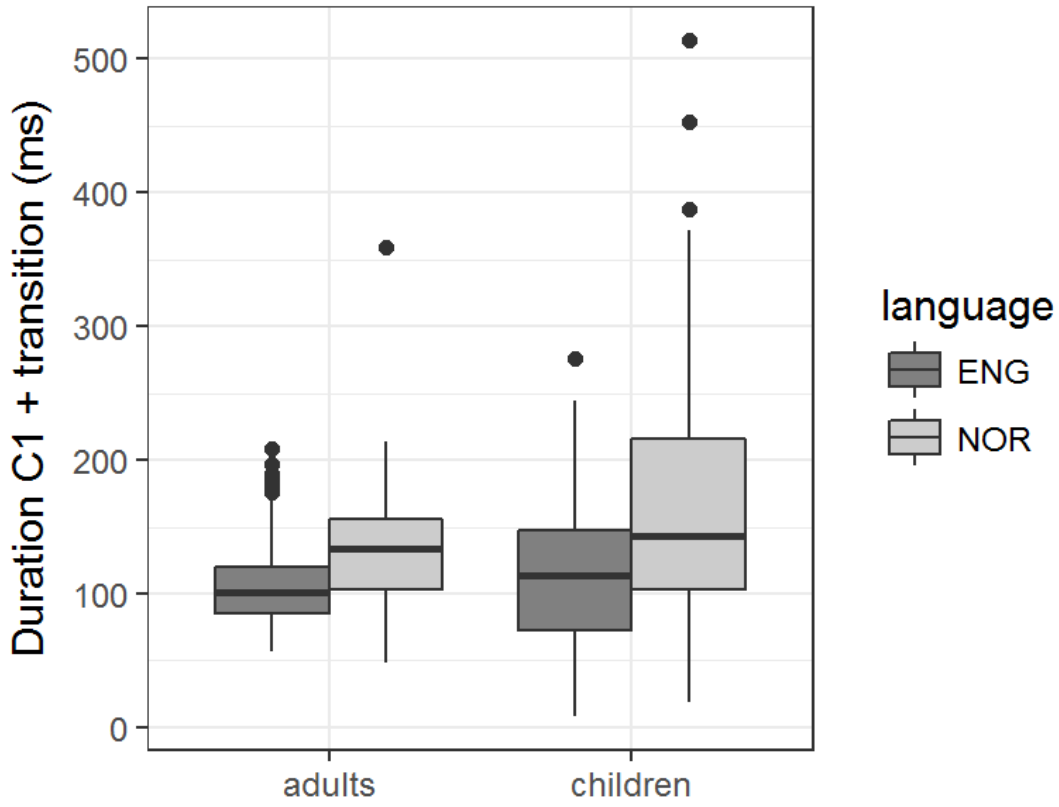

Figure 7: Cross-linguistic comparison of the duration of C1 and transition between the consonants in a cluster in adults as well as children.

Discussion

Our point of departure for this study was the observation that according to earlier studies (Simonsen, 1990; and Simonsen, Garmann & Kristoffersen, 2019), for Norwegian-learning children, a salient strategy in the production of initial consonant clusters is vocalic insertion. This strategy does not seem to be as salient in many other languages; in particular, it has been reported as an infrequent pattern for English-speaking children (McLeod et al., 2001). One possible reason for this cross-linguistic difference could lie in the speech production behaviour of adults for the two languages and the role this plays in shaping the acquisition pathway; specifically, Norwegian has been reported to have an open transition between consonants, while English has a closer transition with considerable overlap between adjacent consonants.

In our comparison of the production of consonant clusters in Norwegian and English adult speech, we found clear evidence for an open transition in Norwegian, but not in English, supported by the finding that overall, clusters had a longer duration in Norwegian than in English. Taking into account the general difficulty in segmentation, we found that the first segment in a consonant cluster was mostly released before the beginning of the articulation of the second consonant in Norwegian, while this was not so in English. So as not to exaggerate the number of vocalic insertions, we set up strict criteria for their identification concerning amplitude, voicing, formant structure, and duration, with the result that vocalic insertions would, if anything, be *under*-reported in

both languages. Even so, we found a large proportion of vocalic insertions in Norwegian, in contrast to virtually none in the English data, as we hypothesised.

With no significant difference in speech rate between Norwegian and English speakers, and no relationship between speech rate and the number of vocalic insertions in the Norwegian participants, we assume that open transition in Norwegian clusters and resulting vocalic insertion is not an artefact of speech rate, but rather of the particular temporal coordination of the gestural events.

Vocalic intrusion, not epenthesis

The duration of the vocalic insertions in Norwegian adult speech varied from 9 to 54 ms. As they are relatively short and optional in occurrence, we argue that the vocalic insertions in Norwegian cluster productions are not true phonological epentheses according to the definition by Hall (2006), but vocalic *intrusions*, resulting from flexibility in gestural timing. Although the vocalic intrusions were shorter than the unstressed short vowels, more than half of them overlapped in duration with the unstressed short vowels. This overlap indicates that at least some of the vocalic intrusions are not only perceptually salient for children, but that in terms of duration, they could plausibly even be interpreted as lexical vowels by the infant listener.

We do not know how long a vowel needs to be for it to be perceived as such, and this might differ between individuals depending on for example the phonetic training and the listener's awareness of vocalic intrusion in Norwegian. This question cannot be answered through our study. However, further investigation that examines not only adults' perception of vocalic intrusions, but also children's perception of intrusive vowels in adult speech, could throw more light on the processes by which infants map the linguistic input they are exposed to onto their emerging phonological representations, and how language-specific detail can critically shape these processes. Further investigation would also reveal cross-linguistic differences in perceptual expectations in this regard, since the language-specific nature of prosodic-phonetic biases predicts different expectations in the auditory domain too. For adults, we can assume that cross-linguistically, different speech patterns at this level of phonetic detail are accompanied by subtle, but systematic differences in perceptual thresholds. We would for example expect English listeners to be more sensitive to vocalic intrusion in clusters than their Norwegian counterparts, since this does not typically happen in English speech production, at least not for speech in 'normal' registers and contexts. Anecdotally, native speakers of Norwegian tend not to be aware of the intrusive vowel, and report difficulty in perceiving it even when the articulation is pointed out to them. Furthermore, the vocalic intrusions do not appear to affect stress assignment. This is additional evidence for the vowels being interpreted as intrusive rather than true epenthetic vowels according to Hall's (2006) definitions. In a perceptual study of English, Portnoy (2018) and Portnoy and Payne (2018) report a varying threshold for the perception of a vocalic interval along a continuum from CC to CVC, as a function of phonetic, grammatical and lexical factors. While it is to be expected that perception of vocalic intrusion in Norwegian is also mediated by higher order factors, there is also reason to predict different perceptual thresholds, with Norwegians requiring a longer duration of vocalic interval in order to perceive one.

The incidence of vocalic intrusion in Norwegian is gradient and clearly shaped at least in part by articulatory considerations. We did not find vocalic intrusions in all contexts: They are overwhelmingly more prevalent in clusters with a liquid as a second consonant, and more so if the second consonant is a rhotic tap or flap than if it is a lateral. Contrary to our predictions, clusters with vocalic intrusions were significantly shorter in duration than clusters without vocalic intrusions. We advance the following explanation: We found the majority of our vocalic intrusions in clusters with taps and flaps, which are both short, leading to short clusters even when the vocalic portion is included.

It is possible that vocalic intrusion in itself is a consequence of the shortness of the tap (and the flap) in Norwegian. Bradley (2007) referring to Walsh (1997) argues that to enhance perceptibility, cross-linguistically, taps tend to be flanked by vowels. Turning the argument from perception to production, as we raised earlier on, the tap involves a rapid, ballistic movement from a neutral tongue position to an apical closure and back to the neutral tongue position again. This movement may result in the articulation of a neutral vowel preceding (and possibly following) the tap. The production of a flap also involves a preceding neutralisation of the tongue position. Arguably, the vocalic intrusion may also play a role in maintaining a form of temporal "balance" among clusters from an auditory perspective, by making up the "missing" time associated with very short articulations.

We hypothesized, in accordance with previous findings by Davidson and colleagues, that vocalic intrusions would be more common when the first consonant is a stop, and especially when that stop is voiced. Looking at C1s separately, this was indeed the case. However, our analyses of combinations of C1 and C2 indicate that voicing is more important than degree of obstruction of the C1, and that C2 as tap or flap carries the most weight: The prototypical Norwegian vocalic intrusion occurs in a cluster with a voiced stop as C1 and a tap or a flap as C2. Even so, there is a considerable amount of vocalic intrusions in other contexts (e.g when C2 is a lateral), among both adult and child speech in Norwegian, but not in English. These patterns indicate that there are very clear language-specific biases related to transition type. These biases channel speech behaviour in two distinct ways for the two languages examined, and are salient enough also to shape the acquisition process.

Vocalic intrusion in children's speech

As hypothesized, both Norwegian- and English-speaking children use various production strategies when targeting initial consonant clusters, and Norwegian children have many more vocalic intrusions than English children. Our analyses of adult's and children's speech suggest that the children overall mimic the adults in their production of vocalic intrusions. Thus, vocalic intrusion cannot be reduced to a strategy children employ to overcome the challenge of producing consonant clusters. Quite the opposite is true: this is the way Norwegian is spoken.

While both Norwegian and English children show similar patterns as the adults speaking the same language, both groups of children produce a higher proportion of vocalic intrusions than the adults. In addition to mimicking the adults, the children seem to use vocalic intrusion somewhat more and with a longer duration than the adults. Thus, vocalic intrusion appears to be a strategy as well, reflecting that a CVC sequence is easier

for a young child to produce than a CC sequence. For both languages, children struggle with articulating two consonants and the transition between them in a cluster. For English children, the transition is close, and they eventually succeed relatively well at this. For Norwegian children, the transition is more open, and in certain contexts, in particular with liquids as the second consonant, the adults often produce a vocalic intrusion between the consonants. This phonetic detail is a prosodic-phonetic bias that the Norwegian child is exposed to and must learn (Payne, 2016), and one which at the same time simplifies the cluster production for the children. Thus, while vocalic intrusion is a kind of simplification strategy, it is one which is language-appropriate and cannot in itself be counted as a matching error for Norwegian children.

The child data provide strong evidence for the language-specific nature of the acquisition pathway. Interestingly, in Norwegian, the strategy is displayed most evidently at 4 years, and we conjecture that this *increase* in incidence from 2 to 4 years is due to the fact that the 4-year-olds attempt to produce the full cluster (both elements) more than the 2-year-olds, and thus employ the vocalic intrusion strategy more. The development is probably facilitated by the greater coordinatory abilities acquired by this age, compared to those acquired by two-year-olds. The English-learning children, by contrast, show a marked lack of vocalic intrusion in their cluster production attempts, preferring to use simplification or reduction instead, with reductions continuing well into the fourth year. Curiously, however, there is a spike of vocalic intrusion adoption at age 6 – a spike which is nevertheless lower than the lowest incidence in Norwegian-learning children (at age 2). We conjecture that this localised spike reflects a greater effort at this age to produce clusters more fully, and note that it accompanies a sharp decline in actual reductions at this age.

Thus, we see evidence for an integration of universal articulatory challenges and language-specific strategy biases: Cluster production is difficult and children adopt a variety of strategies at early stages in acquisition, independent of the target language. At the same time fine phonetic detail of cluster production in the ambient adult speech influences the degree to which each of the strategies are applied. Reductions are adopted in both languages early on, but are greatly dispreferred much earlier on in Norwegian (much diminished at age 4) than in English (much diminished only at age 6). Simplification is also very prevalent in both languages at an early age, but is dramatically diminished in Norwegian by 4 years, while remaining prevalent in English even by age 6 (perhaps mirroring the propensity for cluster assimilations in adult speech). Vocalic intrusion is present from an early age in Norwegian and increases in incidence with age, becoming the main strategy for four- and six-year-olds, while in English it is largely avoided until the age of 6.

Articulatory Phonology provides a good framework for modelling the phenomenon: the vocalic gesture which is assumed to always be present during articulation emerges in cases where the consonantal gestures do not overlap. This is particularly evident in the articulation of voiced stops plus taps and flaps, and in a language with a generally open transition like Norwegian. For children, this tendency is exaggerated through their slower articulation and problems with the timing of gestures. It is also possible that, for a period during the acquisition process, they interpret vocalic intrusion as more phonologically embedded. Gestural frameworks, such as Articulatory Phonology, help model what is happening at the speech production level, and how this differs cross-

linguistically, without recourse to claims of differences at the more abstract phonological level. In our view, abstract phonological structures are emergent, and crucially based not only on gestural coordination patterns, but also on auditory patterns in the input. More detailed examination of formant structures in vocalic intrusions, and follow-up perceptual experiments will hopefully elucidate these abstract structures further.

Conclusion

The comparison of initial clusters in English and Norwegian adult speech clearly demonstrates that indeed, English has a close transition, while Norwegian has an open transition between the consonants. Our study indicates that in the acquisition of consonant clusters, the phonetic realisations of clusters in the ambient languages – more specifically whether there is an open or a close transition between consonants – plays a crucial role in the strategies employed by children. For languages with an open transition, the intrusion of a short vowel between the consonants in a cluster, which occurs relatively sporadically, but in a more widespread and systematic manner within specific phonetic contexts, is clearly perceived by the children and used in their productions. This language-related tendency is then strengthened through the child-specific tendency of slower articulation and difficulties with gestural timing, allowing for vocalic intrusions in the children's speech not only in Norwegian, but also in English.

These findings provide clear evidence of an integration of the universal articulatory challenges associated with consonant cluster production and the prosodic-phonetic biases that arise due to systematic cross-linguistic differences in the phonetic realisation of those clusters. These language-specific phonetic biases are reflected in the strategies that are adopted at early stages in acquisition which mirror fine phonetic detail of cluster production in the ambient adult speech.

References

Barry, M. (1991). Temporal modelling of gestures in articulatory assimilation. *Proceedings of the XXIIth Congress of Phonetic Sciences*, vol. 4, 14–17.

Bernhardt, B. M., & Stemberger, J. P. (2018). Tap and trill clusters in typical and protracted phonological development: Conclusion. *Clinical Linguistics & Phonetics* 32(5-6), 563-575.

Bernthal, J. E. and Beukelman, D.R. (1978). Intraoral Air Pressure During the Production of /p/ and /b/ by Children, Youths, and Adults. *Journal of Speech, Language, and Hearing Research*, 21, 361–371.

Boersma, P. & Weenink, D. (2016). *Praat: doing phonetics by computer [Computer program]. Version 6.0.20*, retrieved 3 September 2016 from http://www.praat.org/.

Bradley, T. G. (2007). Morphological derived-environment effects in gestural coordination: A case study of Norwegian clusters. *Lingua*, *117*(6), 950–985.

Browman, C. P., & Goldstein, L. (1990). Tiers in articulatory phonology, with some implications for casual speech. *Papers in laboratory phonology I: Between the grammar and physics of speech*, 341-376.

Browman, C., & Goldstein, L. (1991). Gestural structures: Distinctiveness, phonological processes, and historical change. In *Modularity and the motor theory of speech perception: Proceedings of a conference to honor Alvin M. Liberman* (pp. 313–338). Erlbaum Hillsdale, NJ.

Browman, C.P., & Goldstein, L. (1992a). Articulatory phonology: An overview. *Phonetica*, 49(3–4), 155–180.

Browman, C. P., & Goldstein, L. (1992b). "Targetless" schwa: an articulatory analysis. *Papers in laboratory phonology II: Gesture, segment, prosody*, 26-56.

Byrd, D. & Tan, C.C. (1996). Saying consonant clusters quickly, *Journal of Phonetics*, 24, 262–282.

Byrd, D. (1996). Influences on articulatory timing in consonant sequences. *Journal of Phonetics*, *24*(2), 209–244.

Catford, J. C. (1977). *Fundamental problems in phonetics*. Edinburgh: Edinburgh University Press.

Catford, J. S. (1988). *A practical guide to phonetics*. Oxford: Clarendon Press.

Davidson, L. (2000). Hidden rankings in the final state of the English grammar. Baltimore: Johns Hopkins, ms.

Davidson, L. (2005). Addressing phonological questions with ultrasound. *Clinical Linguistics and Phonetics*, 19:6/7, 619–633.

Davidson, L. (2010). Phonetic bases of similarities in cross-language production: Evidence from English and Catalan. *Journal of Phonetics*, *38*(2), 272-288.

Davidson, L., & C. Wilson. (2016). Processing non-native consonant clusters in the classroom: Perception and production of phonetic detail. *Second Language Research* 32:4, 471–502.

Endresen, R. T. (1991). Fonetikk og fonologi. *Ei elementær innføring*, 2nd ed. Oslo: Universitetsforlaget.

Gafos, A. (2002). A grammar of gestural coordination. *Natural Language and Linguistic Theory* 20 (2), 269–337.

Gafos, A., & Goldstein, L. (2012). Articulatory representation and phonological organization. In A. C. Cohn, C. Fougeron, M. K. Huffman (Eds.). *Handbook of Laboratory Phonology.* Oxford: Oxford University Press, 220-231.

Gósy, M. (2008). "R" hangok: Kiejtés, hangzás, funkció [R sounds: Pronunciation, sounding, function]. *Magyar Nyelvör, 132*, 1–17.

Hall, N.E. (2003). *Gestures and segments: Vowel intrusion as overlap* (Doctoral dissertation).

Hall, N.E. (2006). Cross-linguistic patterns of vowel intrusion. *Phonology*, *23*(3), 387–429.

Hall, N. (2013). Acoustic differences between lexical and epenthetic vowels in Lebanese Arabic. *Journal of Phonetics*, *41*(2), 133-143.

Hardcastle, W. J. (1985). Some phonetic and syntactic constraints on lingual coarticulation during /kl/ sequences, *Speech Communication*, 4, 247–263.

Hardcastle, W., & Roach, P. (1977). An instrumental investigation of coarticulation in stop consonant sequences. In H. Hollien, & P. A. Hollien (Eds.), *Current Issues in the Phonetic Sciences* (pp. 531–540). Amsterdam: John Benjamins.

Ignatova, D., Bernhardt, B.M., Marinova-Todd, S., & Stemberger, J. P. (2018). Word-initial trill clusters in children with typical versus protracted phonological development: Bulgarian. *Clinical Linguistics & Phonetics 32* (5-6), 506-522.

Imbrie, A.K.K. (2005). *Acoustical study of the development of stop consonants in children*. Doctoral thesis, Massachusetts Institute of Technology.

Koenig, L.L. (2000). Laryngeal Factors in Voiceless Consonant Production in Men, Women, and 5-Year-Olds. *Journal of Speech, Language, and Hearing Research* (43) 1211–1228.

Kristoffersen, G. (2000). *The Phonology of Norwegian*. Oxford: Oxford University Press..

Kristoffersen, K.E., & Simonsen, H.G. (2006). The acquisition of# sC clusters in Norwegian. *Journal of Multilingual Communication Disorders*, *4*(3), 231–241.

Kwon, H. & Chitoran, I. (2016). Cross-linguistic differences in articulatory timing lag in consonant cluster perception, *The Journal of the Acoustical Society of America*, Volume 140, Issue 4.

Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages* (Vol. 1012). Oxford: Blackwell.

Lundeborg Hammarström, I. (2018). Word-initial/r/-clusters in Swedish speaking children with typical versus protracted phonological development. *Clinical Linguistics & Phonetics 32* (5-6), 446-458.

Másdóttir, T. (2018). Word-initial/r/-clusters in Icelandic-speaking children with protracted versus typical phonological development. *Clinical Linguistics & Phonetics 32* (5-6), 424-445.

McLeod, S., J. van Doorn, V. A. Reed (2001). Normal Acquisition of Consonant Clusters. *American Journal of Speech Language Pathology.* Vol. 10*: 99–110.*

Moen, I., Simonsen, H. G., Lindstad, A. M., & Cowen, S. (2003, August). The articulation of the East Norwegian apical liquids/ɭ ɾ ɽ /. In *The 15th International Congress of Phonetic Sciences* (pp. 1755-1758).

Nolan, F. (1992). The descriptive role of segments: evidence from assimilation. In Docherty, G. & Ladd, D.R. (Eds.), *Laboratory Phonology 2*, 261-280. Cambridge: CUP.

Ozbič, M., Kogovšek, D., Stemberger, J.P., Bernhardt, B.M., Muznik, M., & Novšak Brce, J. (2018). Word-initial rhotics in Slovenian 4-year-olds with typical versus protracted phonological development. *Clinical Linguistics & Phonetics 32* (5-6)*,* 523-543.

Payne, E. (2016). Prosodic-phonetic biases and the construction of phonological markedness. In Enger, H.-O., Knoph, M.I.N.; Kristoffersen, K.E., & Lind, M. (Eds.) *Helt fabelaktig! Festskrift til Hanne Gram Simonsen på 70-årsdagen*. Oslo: Novus, 181–198.

Payne, E., Post, B., Garmann, N. G., & Simonsen, H. G. (2015). VC timing acquisition: Integrating phonetics and phonology. *The Scottish Consortium for ICPhS*.

Payne, E., Post, B., Garmann, N. G., & Simonsen, H. G. (2017). The acquisition of long consonants in Norwegian. In Kubozono, H. (Ed.): *The Phonetics and Phonology of Geminate Consonants* (Vol. 2). Oxford University Press, 130–162.

Perez, D., Vivar, P., Bernhardt, B. M., Mendoza, E., Ávila, C., Carballo, G., ... & Vergara, P. (2018). Word-initial rhotic clusters in Spanish–speaking preschoolers in Chile and Granada, Spain. *Clinical Linguistics & Phonetics 32* (5-6), 481-505.

Portnoy, E. (2018). The perception of a vowel in consonant clusters. *Unpublished BA dissertation*, University of Oxford, 1–60.

Portnoy E., & Payne E. (2018). The role of phonotactics and lexicality on the perception of intrusive vowels. Paper presented at the *Representing Phonotactics Workshop,* LabPhon, Lisbon June 2018.

Prieto Vives, P., & Bosch Baliarda, M. (2006). The development of codas in Catalan. *Catalan Journal of Linguistics*, 5, 237-272.

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

RStudio Team (2018). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA. http://www.rstudio.com/.

Ramalho, A. M., & Freitas, M. J. (2018). Word-initial rhotic clusters in typically developing children: European Portuguese. *Clinical Linguistics & Phonetics 32* (5-6), 459-480.

Saffran, J. R. (2003). Statistical language learning: Mechanisms and constraints*. Current directions in psychological science*, 12(4), 110–114.

Simonsen, H.G. (1990). Barns fonologi: System og variasjon hos tre norske og ett samoisk barn. (Unpublished doctoral dissertation, University of Oslo).

Simonsen, H.G., Garmann, N.G. & Kristoffersen, K.E. 2019. Consonant clusters in the speech of children with 5p deletion syndrome. Hognestad, J.K., Kinn, T. & Lohndal, T. *Fonologi, sosiolingvistikk og vitenskapsteori. Festskrift til Gjert Kristoffersen.* Oslo: Novus forlag, p. 295-314.

Simonsen, H. G., Moen, I., & Cowen, S. (2008). Norwegian retroflex stops in a cross linguistic perspective. *Journal of Phonetics*, *36*(2), 385-405.

Stathopoulos, E. T., & Sapienza, C. (1993). Respiratory and laryngeal measures of children during vocal intensity variation. *The Journal of the Acoustical Society of America*, *94*(5), 2531-2543.

Stemberger, J. P., & Bernhardt, B. M. (2018). Tap and trill clusters in typical and protracted phonological development: Challenging segments in complex

phonological environments. Introduction to the special issue. *Clinical Linguistics & Phonetics 32* (5-6), 411-423.

Steriade, D. (1990). Browman and Goldstein's paper. *Papers in laboratory phonology I: Between the grammar and physics of speech*, 382–397.

Tar, É. (2018). Word-initial tap-trill clusters: Hungarian. *Clinical Linguistics & Phonetics 32* (5-6), 544-562.

Vanvik, A (1971). The phonetic-phonemic development of a Norwegian child: *NTS* 24, 269–325.

Vihman, M. M., & Velleman, S. L. (2000). The construction of a first phonology. *Phonetica*, 57(2-4), 255–266.

Walsh, L. (1997). The phonology of liquids. *University of Massachusetts, Amherst: Doctoral dissertation*.

Wilson, C., & Davidson, L. (2013). Bayesian analysis of non-native cluster production. In *Proceedings of NELS* (Vol. 40).

Yavaş, M., Ben-David, A., Gerrits, E., Kristoffersen, K.E., & Simonsen, H.G. (2008). Sonority and cross-linguistic acquisition of initial s-clusters. *Clinical Linguistics & Phonetics*, *22*(6), 421–441.

Zsiga, E. C. (1994). Acoustic evidence for gestural overlap in consonant sequences. *Journal of Phonetics, 22*, 121–140.

APPENDIX

Table I: Number of word tokens in the adult data by language and consonant cluster.

Table I

Adults	bl	pl	br	pr	tr	dr	kl	kr	gl	gr	fl	fr	kw	tw	sp	st	sk	sl	sn	sm	sw	Total	Total
NOR	8	8	12	8	12	8	8	8	8	8	8	8	4	8	12	12	12	8	12	8	8	188	366
ENG	8	8	12	8	12	8	8	8	12	8	8	8	4	8	7	8	11	12	8	4	8	178	

Table II: Number of word tokens in the child data by language, age group and consonant cluster.

Table II

Children	Age	pl	br	tr	kl	kr	gl	fl	fr	sp	st	sk	sn	sm	sw	Total	Total	Total
NOR	2.5	2	6	3	1	2	3	5	3	3	3	5	3	3	4	46	156	300
	4	3	8	3	3	3	3	5	3	2	3	6	3	4	6	55		
	6	3	9	3	3	3	3	6	3	3	3	6	3	2	5	55		
ENG	2.5	3	3	3	3	3	6	3	3	3	3	3	3	6	3	48	144	
	4	3	3	3	3	3	6	3	3	3	3	3	3	6	3	48		
	6	3	3	3	3	3	6	3	3	3	3	3	3	6	3	48		

Table III: Target words in the material given to each group.

Table III				
Cluster type	**ENG adults**	**NOR adults**	**ENG children**	**NOR children**
bl	blab	blass		
	black	blod		
br	braise	brann		brann
	break	brannmann		brannmann
	breakfast	brysk	breakfast	brød
dr	dregs	drikke		
	drive	drodle		
fj		fjær		
fl	flag	flagre	flag	flagg
	flak	fly		flaske
fr	frog	frakk	frog	frosk
	frump	fred		
gl	glass	glass	glass-of-milk	glass
	glove	glidelås	glove	
	glut			
gr	grass	grei		
	grot	grind		
kl	clods	kladd		
	clothes	klær	clothes	klær
kr	crag	krakk		krakk
	cry	krig	cry	
kv	quail	kveld		
pl	plait	pleier	plaster	plaster
	play	plogfure		
pr	prat	presang		
	price	problem		
sk	scapegoat	skapt		
	school	skole	school	skole
	skulk	skur		sko
sl	sleep	slipe		
	sleet	slutt		
	slept			
sm		smile	smile	
	smoke	smokk	smoke	smokk
sn	snails	snakke	snail	
	snide	snekre		
		snørr		snørr
sp	spook	spade		spade
	spoon	spikre	spoon	
		spill		
st	stooge	staur		
	stool	stein	stool	stein
		stor		
sv	sweet	svane	sweet	svane
	swig	svar		sverd
tr	track	trekke		
	tractor	traktor	tractor	traktor
	traipse	tro		
tv	twice	tvil		
	twine	tvinne		

Table IV: List of words selected in adult's speech for measuring short, unstressed vowels in Norwegian.

Word	**Realised short vowel**
dagt**i**lbud	(t)i(l)
flagr**e** da	(r)ɛ(d)
h**e**likopter	(h)ɛ(l)
hel**i**kopter	(l)i(c)
kunn**e** snakke	(n)ə(s)
kyll**i**ng	(l)i(ŋ)
matpakk**e** med	(c)ə(m)
m**e**d sprit	(m)ə(s)
rens**e** såret	(s)ə(s)
smil**e** da	(l)ɛ(d)
strømp**e**bukse	(p)ə(b)
sår**e**t med	(r)ə(m)

FOOTNOTES

1. In the literature, the terms ‘epenthesis’, ‘vowel/vocalic insertion’ and ‘vowel/vocalic intrusion’ are terms used for the phenomenon treated here. In the introductory part, we use the terms applied in the articles referred to. In the discussion, we turn to the status of such elements, and whether they appear to be phonetic or phonological in nature.