Influence of L1 on the production of English voiceless plosive clusters by Norwegian speakers

Jacques Koreman[1], Monica Graff-Wang Andersen[1], Andrea Hexeberg Bjørk[1], Bistra Andreeva[2]
[1] Department of Language and Literature, NTNU, Norway
[2] Department of Language Science and Technology, Saarland University, Germany
jacques.koreman@ntnu.no, {mganders, andrehbj}@stud.ntnu.no, andreeva@coli.uni-saarland.de

Abstract

The realisation of voiceless plosive sequences is investigated in six English and Norwegian clusters. Differences between the two languages are found in the frequency of the release of the first plosive in the cluster. While a burst is nearly always present in Norwegian, it is more variable in English, where it also depends on the cluster. In English, burst durations are shorter and burst intensities are lower than in Norwegian. The total duration of the whole cluster is much shorter in English than in Norwegian.

For Norwegian L2 speakers of English, the cluster durations are intermediate between Norwegian and L1 English. The L2 speakers nearly always produce a release burst for first plosive in the cluster, as in Norwegian. The burst duration is somewhat shorter than in Norwegian, but longer than in English. Burst intensities are higher than in English.

Introduction

The realisation of medial English plosive clusters in words like *ou**tp**ut* and *la**pt**op* can differ from that in similar Norwegian words. In textbooks on English, the first consonant in clusters is often described as lacking an audible release, as shown in Fig. 1 (Davidson-Nielsen, 1977; Ladefoged, 2005; Ashby & Maidment, 2005; Ogden, 2017), as a result of gestural overlap (Browman & Goldstein, 1987), but newer research paints a more variable picture (e.g. Davidson, 2011; Ghosh & Narayanan, 2009). In Norwegian on the other hand, informal observations indicate that speakers tend to pronounce the first consonant (C1) with a clearly audible release, cf. Fig. 2, similar to for example Russian (Zsiga, 2003) or Polish (Rojczyk et al., 2013).

This article compares the pronunciation of voiceless plosive clusters in L1 English (EE) and Norwegian (NN). The investigation also includes their pronunciation in L2 English spoken by Norwegians (NE). The data were collected as part of the bachelor theses of the second and third author.

We investigate the following hypotheses:

1) NN differs from EE in the C1 burst (both with a higher frequency and greater strength), as well as longer cluster durations.
2) Similar differences are expected between EE and NE.

Method

Subjects

Five English speakers (4 Americans and 1 Canadian; 3 female, 2 male) produced English words, while 5 Norwegians (2 female, 3 male; all students of English at NTNU) produced both Norwegian and English words. All speakers were between 20-27 years old and participated in the experiment voluntarily. The recordings for one of the American speakers were excluded, because her pronunciations were hyperarticulated with clear pauses breaking up the clusters in the words, which mainly were compounds.

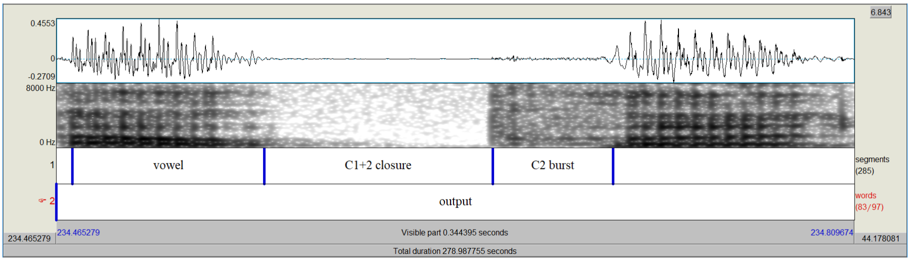

Figure 1. Waveform and spectrogram for the English word OUTPUT (speaker EE4)

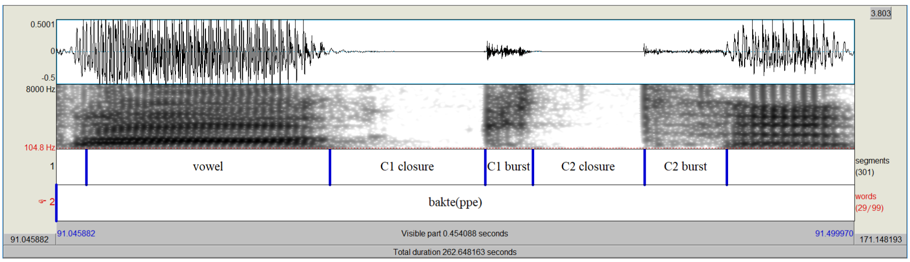

Figure 2. Waveform and spectrogram for the Norwegian word BAKTEPPE (speaker NE5)

Material and recordings

The stimulus material consisted of 48 words divided over six plosive clusters /pt, pk, tp, tk, kp, kt/, cf. Table 1 for English and Table 2 for Norwegian. Many of the Norwegian words are longer than the English words because it was not possible to find suitable shorter words. Only voiceless plosive clusters were used because the C1 bursts are easier to identify in the speech signal.

The words were produced in isolation and were read by the informants from a PowerPoint presentation on a computer screen. The recordings were made in Praat (Boersma & Weenink, 2017) with a 44.1 kHz sampling rate and 16-bit amplitude resolution, using a Shure KSM44 microphone. Recordings were made in a sound-treated studio.

Measurements and statistical analysis

We measured C1 burst duration (including aspiration) and maximum intensity. Additionally, the total closure duration was measured from the beginning of the closure of the first plosive until the release of the closure of the second plosive, including the intervening C1 burst if present. All durational measurements were converted to log values before statistical analysis in order to make their distribution less skewed.

Chi-squared tests in SPSS (IBM Corp., 2017) were used to analyse burst frequencies. Burst and closure durations as well as burst intensity were analysed using linear mixed-effects models in R (R Core Team, 2018). Significance of fixed effects was evaluated by performing maximum likelihood *t*-tests using Satterthwaite approximations to degrees of freedom. The lme4 (Bates et al., 2015) and lmerTest packages (Kuznetsova et al., 2017) were used. To compute the final model, we used the step procedure in R for backwards variable selection. All effects were treatment coded. The lsmeans function was used to perform all pairwise comparisons with Tukey adjustment.

Results

Two attributes of the clusters are discussed in this section. The first is the presence of an audible or visible release burst in the signal. The duration and intensity of the burst were also analysed.

Table 1. English stimulus words for six clusters

/pt/	/pk/	/tp/	/tk/	/kp/	/kt/
LAP TIME	CUPCAKE	JETPACK	KIT KAT	JACKPOT	BACKTALK
LAPTOP	POPCORN	POTPIE	BITCOIN	CHECKPOINT	TRACK TIME
UPTOWN	TOPCOAT	NET PULSE	NUTCASE	BACKPACK	WORK TASK
PEP TALK	SHOPKEEPER	BRAT PACK	MEAT CUBE	LOCKPICKING	COOKTOP
SHOPTALK	DROPKICK	MEAT PIE	FLEET COST	BLACKPOOL	MILK TANK
TOP TEETH	UPKEEP	FOOTPRINT	CATCALL	SICK PAY	WORKTABLE
TIP-TOP	LIP CARE	OUTPUT	SITCOM	BREAKPOINT	NECKTIE
TIPTOE	STOP KEY	PILOT PIN	BATCAVE	COCKPIT	FOLKTALES

Table 2. Norwegian stimulus words for six clusters

/pt/	/pk/	/tp/	/tk/	/kp/	/kt/
KAPTEIN	PAPPKARTONG	MOTPOL	MOTKAMP	LØKPULVER	TAKTERRASSE
OPTIKER	OPPKAST	MATPAKKE	MATKASSE	BAKPÅ	BAKTEPPE
SKEPTISK	POPKORN	SØTPOTET	MATKURS	LAKKPISTOL	BAKTANKE
STOPPTEGN	SKAPKANT	DIETTPENGER	NATTKÅPE	MINKPELS	PRAKTISK
OPPTA	OPPKALLE	ROTPERSILLE	MATKART	STIKKPILLE	HEKTISK
SOPPTUR	POPKUNST	MOTPART	MOTKULTUR	COCKPIT	VEKTER
TOPPTUR	OPPKOK	POTETPULVER	UTKANT	TRYKKPUMPE	DOKTOR
LAPTOP	OPPKOMME	UTPEKE	POTETKAKE	TAKPAPP	KAKTUS

The second attribute investigated is the total closure duration. With more overlap between the articulatory gestures for the two consonants, this duration will become shorter (ignoring the effect of speaking rate).

Burst: frequency

Pearson's chi-squared tests of burst frequency reveal that language condition was highly significant, with χ^2 (2, N = 672) = 139.83, $p < 0.001$. The effect was also found when each language pair was analysed separately (EE-NN and EE-NE: $p < 0.001$; NE-NN: $p < 0.01$).

The English speakers did not produce a C1 burst in 72 out of 192 cases, i.e. in 37.5 % of the clusters (see Fig. 3). The Norwegian speakers realized the C1 burst in almost all cases, except 3 out of 240 cases when speaking Norwegian words (NN), and 14 out of 240 cases when speaking English (NE).

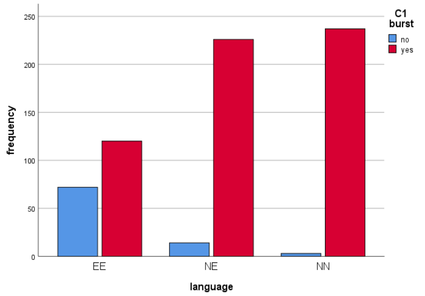

Figure 3. C1 burst frequencies for three language conditions

To investigate the effect of cluster on burst frequency, we performed separate chi-squared tests for each language condition. In these tests the dependency of burst frequency on cluster was evaluated.

For EE, this relation was highly significant, with χ^2 (5, N = 672) = 61.07, $p < 0.001$. As Fig. 4 shows, the order of frequency of C1 bursts was /p/ > /k/ > /t/. Burst were in fact more often absent than

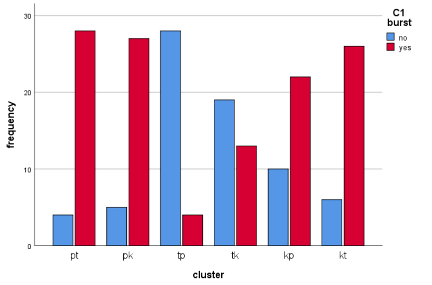

Figure 4. C1 burst frequency for six clusters (EE)

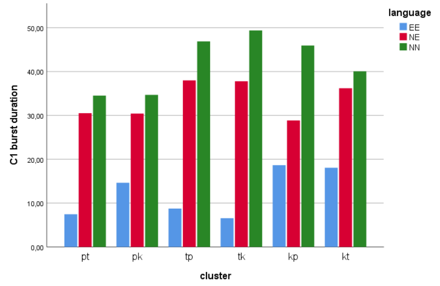

Figure 5. C1 burst durations for three language conditions

present when the first consonant in the cluster was alveolar, i.e. in /tp/ and /tk/.

For NN, there was no effect of cluster. This is not surprising, given the fact that there were only three cases where the burst was not realized (one for /pt/, one for /pk/ and one for /kt/).

Although the number of unreleased C1 bursts was also low for NE (14 out of 240), a chi-squared test showed a significant effect for burst frequency and cluster (χ^2 (5, $N = 672$) = 11,53, $p < 0.05$).

Burst: duration and intensity

A linear mixed-effects model was applied to all the data (for all three languages and all clusters). The final model for C1 burst duration included language and cluster as fixed effects as well as their interaction and the random intercept for speakers, but not for words. The same model was selected when burst intensity was analysed.

Pairwise comparisons for significant variables with lsmeans showed that burst *durations* were significantly different for language, with EE < NE < NN (see Table 3).

Table 3. Results for burst duration and intensity from lsmeans

	lang. pair	df	*t*	p
burst duration	EE-NN	11.6	-8.85	0.001
	EE-NE	11.6	-7.37	0.001
	NE-NN	574.4	-4.66	0.001
burst intensity	EE-NN	9.5	-5.12	0.001
	EE-NE	9.5	-4.78	0.001
	NE-NN	573.5	-1.46	n.s.

Further analysis within languages shows that burst durations for EE /pt, tk/are shorter than for /pk, kp, kt/. In NN, /tk/ burst duration is longer than for /pt, pk/. In NE, none of the clusters differ in burst duration (Fig. 5).

Comparisons per cluster show that all burst durations are shorter in EE than in NE and NN. In NE, clusters /tk/ and /kp/ have shorter burst durations than in NN (Fig. 5).

Burst *intensity* was also different in the languages, with weaker bursts in EE than in NE and NN (see Table 3).

In EE, burst intensity is weaker in /tk/ than in /pk, kp, kt/, and /pt/ is weaker than /kp/. In NN, /kp/ and /kt/ have stronger burst intensities than /pt, pk/; /kp/ is also stronger than /tp, tk/. In NE, /kt/ is stronger than /pk, pt/, while both latter clusters have a stronger C1 burst intensity than /tp, tk/ (Fig. 6).

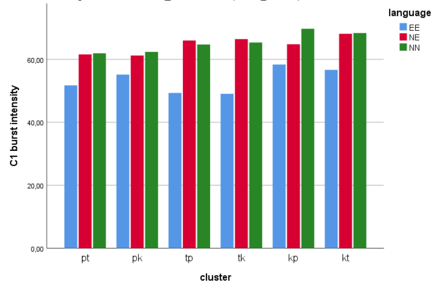

Figure 6. C1 burst intensities for three language conditions

In comparisons per cluster, burst intensity was always lower in EE than in NN. The same was true for the EE-NE language pair, except for the clusters

/pk/ and /kp/ which did not differ. In the comparison NE-NN, there were no differences in burst intensity, except for cluster /kp/ (Fig.6).

Closure duration

Since many clusters were realised without a C1 burst in EE (see Fig. 4), indicating gestural overlap, we first analyse whether this corresponds to differences in the total closure duration for the clusters. A linear mixed-effects model for EE was computed, where cluster, C1 burst, and their interaction were included in the maximum model as fixed effects. It also included random intercepts for speakers and words, and by-speaker random slopes for C1 burst and cluster. After backwards model selection, we ended up with a model including cluster and C1 burst as fixed effects, with random intercepts for speakers. This model showed a significant effect of C1 burst, with $t(188.41) = 2.67$, $p < 0.01$. Clusters with a burst were longer than those without a burst, which is not surprising. Mean closure durations for clusters with a burst were 130 ms (sd = 31.7 ms), while clusters without a burst were 108 ms (sd = 26.4 ms) on average.

Because there are very few observation *with* a burst especially for cluster /tp/ in EE (see Fig. 4), we will present results comparing the three languages based on pooled data with and without a burst (similar results were found for analysis of the data *with* a burst). The maximum model includes the fixed effects language and cluster, and their interaction; random intercepts for speaker and word are also selected. This model could not be reduced.

All languages differ from eachother in terms of closure duration (Table 4).

Table 4. lsmeans results for closure duration

	lang. pair	df	*t*	p
duration	EE-NN	9.8	-5.55	0.001
	EE-NE	9.3	-3.44	0.05
	NE-NN	142.5	-9.13	0.001

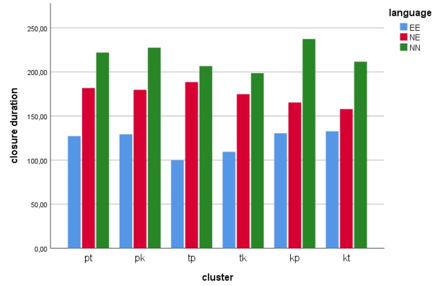

Figure 7. Total closure durations for all clusters in the three language conditions

The differences in the mean values of the languages are shown in Figure 7. The total closure durations are very different in the three language conditions: in EE, they are shortest (121 ms, sd = 31.6 ms), while they are 96 ms longer in NN (217 ms, sd = 61.0 ms); in NE, an intermediate value of 175 ms (sd = 53.1 ms) is found. Norwegians produce the clusters in L2 faster, although the Norwegian words are longer than the English ones.

Looking at the differences between the languages within each cluster, the general picture is that all clusters are realised differently in the three language conditions (see Fig. 7), as might be expected given the clear differences between the closure durations in the languages in general, as reported above. The only exceptions are /kp/ and /kt/, for which the closure durations do not differ in EE and NE; for /tp/ and /tk/, the closure durations do not differ in NE and NN.

An analysis of clusters within each language shows that closure durations of the different clusters mainly differ in EE. The pattern is quite clear: the closure durations of /tp/ and /tk/ do not differ from eachother, but is at least 27 ms shorter for /tp/ than for the four remaining clusters, and it is at least 18 ms shorter for /tk/ (see Fig. 7). The only exception to this pattern is that /tk/ does not differ from /pt/.

In NE and NN on the other hand, none of the clusters differ from each other with respect to their duration. There is one exception in each of the two language conditions: in NN, /tk/ is 38 ms shorter than /kp/, and in NE the cluster /tp/ is 31 ms longer than /kt/ (see Fig. 7).

Conclusion and discussion

The results reported in this study confirm our hypotheses. L1 English (EE) and Norwegian (NN) showed clear differences in all our measures. While in Norwegian (NN) the first plosive in the cluster nearly always has an audible and visible release burst (except for three cases), a release burst occurs in less than two thirds of the cases in L1 English. This is an effect of gestural overlap, which also results in much shorter total closure durations for the cluster. When there *is* a burst, its duration is shorter and its intensity is lower than in Norwegian.

Moreover, the same clusters behave differently in the two languages with respect to all measurements. For example, unlike in Norwegian, English /tp/ and /tk/ have shorter total closure durations and lower C1 burst frequencies than other clusters. This confirms our expectations, as English speakers often replace syllable-final /t/ with a glottal stop (Davidson, 2011). No clear differences were found between front-to-back compared to back-to-front clusters.

When speaking English (NE), Norwegians have shorter closure durations than in their native language. This is not due to more gestural overlap, however, since they produce an audible and visible release almost as often as when they speak Norwegian. A syllable-final /t/ is never replaced by a glottal stop. Their C1 burst durations tend to have intermediate values between L1 English (EE) and Norwegian (NN); their burst intensities do not differ from NN.

We are presently investigating the effect of the presence of a C1 burst and of cluster duration on the perception of a foreign accent. First results indicate that native English listeners do hear the differences between the manipulated stimuli, but do not perceive them as an indication of foreign accent.

References

Ashby, M., & Maidment, J. (2005). *Introducing Phonetic Science.* Cambridge: Cambridge University Press.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *J. Stat. Softw. 67(1),* 1-48.

Boersma, P., & Weenink, D. (2017). Praat: doing phonetics by computer, v. 6.0.35. http://www.praat.org/.

Browman., C. P., & Goldstein, L (1987). Tiers in Articulatory Phonology, with some Implications for Casual Speech. *Haskins Laboratories SR-92.*

Davidson (2011). Characteristics of stop releases in American English spontaneous speech. *Speech Communication 53,* 1042-1058.

Davidsen-Nielsen, N., (1977) *English Phonetics.* Copenhagen: Gyldendal Norsk Forlag A/S.

Ghosh., P. K., Narayanan., S., S (2009). Closure duration analysis of incomplete stop consonants due to stop-stop interaction. *JASA 126(1),* EL1-7.

IBM Corp. (2017*). IBM SPSS Statistics for Windows,* v. 25.0. Armonk, NY: IBM Corp.

Ladefoged, P. (2005). *A Course in Phonetics.* Belmont, CA: Wadsworth Publishing.

Kuznetsova, A., Brockhoff, P. B., Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *J. Stat. Softw. 82(13),* 1-26.

Ogden, R., (2017). *Introduction to English Phonetics*. Edinburgh: EUP.

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

Royczyk, A., Porzuczek, A., & Bergier, M. (2013). Immediate and Distracted Imitation in Second-Language Speech: Unreleased Plosives in English. *Research in Language 11(1)*, 3-18.

Zsiga, E. C. (2003). Articulatory timing in a second language. Evidence from Russian and English. *Stud. Second Lang. Acquis. 25,* 399-432.