AN ACOUSTIC ANALYSIS OF THE RETROFLEX FLAP

Knut Kvale

Telenor Research
P.O.Box 83
N-2007 Kjeller, NORWAY
E-mail: knut.kvale@tf.telenor.no

Arne Kjell Foldvik

Dept. of Linguistics
University of Trondheim
N-7055 Dragvoll, NORWAY
E-mail: arne.foldvik@avh.unit.no

ABSTRACT

This paper investigates the acoustics of the retroflex alveolar flapped stop [ɽ] and demonstrates that different phonetic contexts systematically affect its realization.

1. INTRODUCTION

Retroflex flap [ɽ] is found in numerous Indian languages, in a number of African languages and in some languages in Australia and Mexico. In Europe retroflex flap occurs in Albanian, North-Western Italian and among speakers in large areas of Norway and Sweden.

Taps and flaps are similar in that the active articulator moves more rapidly than in their non-tapped or non-flapped stop-counterparts. The main difference between tapping and flapping is the direction of the tongue movement after the brief closure.

In Norwegian and possibly also in Swedish retroflex flap was considered sub-standard or non-standard, but with the present trend of more general acceptance of regional and social dialects the status of retroflex flap has risen. The previous low status of [ɽ] possibly explains why little or no research has been done on it. However, for speech technology purposes acoustic studies of retroflex flap are important both to produce natural sounding synthetic speech and to succeed in automatic speech recognition.

Typically the Norwegian [ɽ] is pronounced as a *voiced or partly voiced retroflex alveolar flapped stop*, but we are aware that in some languages [ɽ] may be realized as a retroflex alveolar *lateral* flapped stop. For a *phonemic* discussion of [ɽ], see [1].

2. CONTEXTUAL EFFECTS

In this section we first investigate the realisation of [ɽ] in *logatomes* (i.e. artificial words with a fixed syllable structure), containing diphones for Norwegian text-to-speech synthesis. The logatomes were read singly with an average articulation rate of 100 $^{ms}/_{phoneme}$.

In Norwegian the retroflex flap occurs in pre-, inter-, and post-vocalic position in some contexts, but not word initially nor after the front vowels [i], [e] and [y]. In the present logatome material the retroflex flap was also realised in some of these non-occurring positions.

2.1 Intervocalic position

Figure 1 shows the waveform and the broad band spectrogram of the Norwegian words [ɔ: ɽ ə] and [ɔ: ɾ ə], illustrating some similarities and differences between the retroflex flap and the apical alveolar tap in *intervocalic position*. In both words F_1 in the vowels is relatively constant, whereas the low F_2 of the back vowel [ɔ:] is shifted upwards due to the succeeding alveolar closure. (Alveolar phones are characterized by high F_2 [2]). The closure phase of the retroflex flap is acoustically similar to the corresponding part of the apical alveolar tap [3].

The differences between the tap and the flap are manifested in the F_3 and F_4 of the neighbouring vowels. In the *tap* realisation, the formant trajectories can be smoothly interpolated from the vowel preceding the tap to the following vowel, suggesting that the tongue tip returns to the onset position after the closure. In the *flap* realisation, the tongue makes a brief (alveolar) closure in the passing to another position, resulting in an abrupt change in F_3 across the flap closure. In addition, F_4 before the flap closure turns down, whereas before the tap closure F_4 turns slightly upwards.

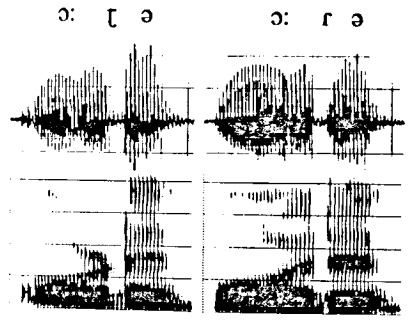

Figure 1 *Waveform and broad band spectrogram of [ɔ: ɽ ə] and [ɔ: ɾ ə] pronounced by a male speaker. Neighbouring vertical lines are 50 ms apart. In the spectrogram the frequency difference between neighbouring horizontal lines is 1 kHz.*

Generally, in isolated words the retroflex flap lowers F_3 and F_4 of the neighbouring vowels, especially in the preceding vowel. For front vowels (with high F_2) the second formant is lowered but is raised for back vowels (with low F_2). The flap has no noticeable effect on the F_1 of the neighbouring vowels.

For the retroflex flap, the curling up of the tongue tip before the forward flap movement takes place shows up in the spectrogram as a fairly long formant transition before the voicebar portion. Typically the whole preceding vowel was diphthongized. Although these formant transitions are important for the perception of [ɽ], they are not included in the [ɽ] segment. (This segmentation approach is similar to the one used for plosives which are defined as starting when the closure begins, and not when the formants of the preceding vowel change [4]).

2.2 Succeeding a consonant

With [ɽ] in a C_V context, an interval of voicing and formant structure is seen in the spectrogram *before* the apex touches the alveolar ridge.

When a bilabial stop or nasal precedes the retroflex flap, the tongue tip is free to curl up and back during the bilabial closure phase. When the built up pressure for the stop is released the tongue tip can flap forward quickly on the egressive airflow. In these consonant clusters the period of voicing between the consonant and the brief flap closure was relatively short (i.e. the duration of the epenthetic schwa, [ə], was about 30 ms after [p] and 50 ms after [b] and [m]).

Also with labiodental fricatives preceding the retroflex flap, the period of voicing between the fricative and the flap closure was short, resulting in an [ə] of about 40 ms duration.

When [ɽ] was preceded by alveolar, postalveolar or retroflex consonants or the apical alveolar tap, the duration of the schwa before the flap closure became noticeable longer; from about 80 ms after [s], up to 130 ms after [ɳ]. When articulating these consonants the tongue tip is engaged in making a closure or a constriction and cannot be curled backwards for the flap till the first closure or constriction is released. These combinations are therefore articulatorily inconvenient and may explain why they do not occur in Norwegian.

When retroflex flap is articulated after the alveolar stops, especially [d] and [n], the [ə] shows up in the spectrogram with a

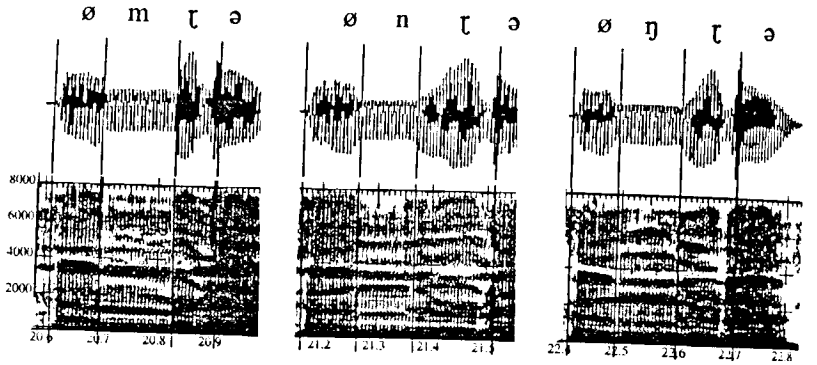

Figure 2 *Bilabial, alveolar and velar nasals preceding the retroflex flap.*

characteristic shift in F_3 from about 2500Hz immediately after the alveolar stop release to about 1700 Hz, (or close to F_2), before the flap closure.

Velar stops involve the dorsal part of the tongue, allowing the tongue tip to start curling backwards during the velar closure phase. Thus, the duration of schwa was only about 50 ms after [k], but longer for [ŋ] (70 ms) and particularly for [g] (80 ms).

Typically, the waveform envelope of [ə] decreases evenly towards the [ɽ] closure, whereas the closure release gives an abrupt change in the waveform, as exemplified in *figure 2*. Notice also the characteristic jumps in F_3 and F_4 from the [ə] before the [ɽ] closure to the following neutral vowel.

2.3 Preceding a consonant

With [ɽ] in a V_C context, the tongue has finished its flapping movement and is free to start its movement towards the consonantal target. Usually the voicing is not turned off after the flap closure release, so a schwa-type vowel appears *after* the flap closure. However, the intensity and duration of this [ə] is much less than when the consonant precedes the retroflex flap. Thus, for [ɽ] in this phonetic environment the duration of [ə] varied from about 40 ms before [b] and [d]) to 70 ms (before [ʃ]). See *figure 3*.

When an apical alveolar tap succeeds the retroflex flap, the tongue tip moves up to produce the tap after the [ɽ] closure release. This take time, and the epenthetic schwa became rather long (about 80 ms). Since this is articulatorily inconvenient, the [ɽ] + [ɾ] combination only occurs across morpheme boundaries in Norwegian.

The F_3 and F_4 usually turn down towards the flap closure, yielding a characteristic jump in these formants to the following epenthetic schwa.

2.4 Continuous speech

In continuous speech the contextual effects between the retroflex flap and the neighbouring sounds are in principle similar to those for the logatomes. However, intervocalically the closure phase of [ɽ] may be very brief or even non-existing. Thus, in these cases only the change in formant-transitions in the neighbouring vowels convey the perceptual cues for [ɽ].

When [ɽ] is preceded or succeeded by a consonant, an extra schwa may appear, e.g. the word "flaske" (=bottle) was realised with:

- a clearly articulated, short epenthetic [ə] and a distinct [ɽ] closure phase,

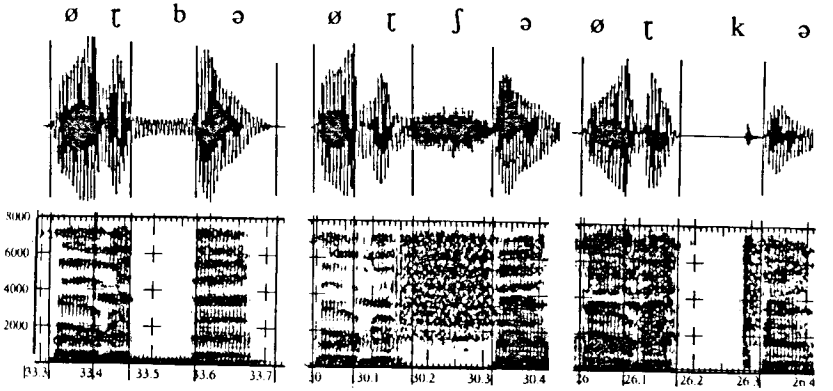

Figure 3 *Bilabial, post-alveolar and velar consonants succeeding the retroflex flap.*

- a very weak [ə] and no [ɽ] closure phase,
- no extra [ə] and partly or totally devoiced [ɽ] with no closure phase.

In the latter case the F_3 of the following back, open vowel was lowered.

The differences between the logatome and continuous speech material were *non-systematic* in that the same person might pronounce words like "flaske" in all the three ways described.

3. CONCLUSIONS

The contextual effects of the retroflex flap can be grouped into three main phonetic environments:

(*i*) [ɽ] in V_V context has an lowering effect on F_3 and F_4; especially in the preceding vowel,

(*ii*) [ɽ] in C_V context is realised with an epenthetic schwa type vowel *before* the flap closure,

(*iii*) [ɽ] in V_C context is realised with a schwa *after* the flap closure.

In (*iii*) the epenthetic schwa is significantly weaker and shorter than in (*ii*).

In continuous speech we found the same contextual effects of [ɽ] as in the logatomes, but with reduction effects: partly devoiced [ɽ], weaker and shorter [ə] and shorter or non-detectable closure phase.

Since the closure phase of the retroflex flap is acoustically very similar to that of the apical alveolar tap [ɾ], we suggest applying the same segmentation approach to these sounds. That is, in linear, phonemic segmentation the epenthetic schwa should be included in the [ɽ]-segment.

We believe that in addition to acquiring more acoustic-phonetic knowledge, such acoustic analyses of speech sounds are important for automatic speech recognition of different dialects and sociolects and for more natural sounding Norwegian text-to-speech synthesis.

REFERENCES

[1] Sandøy, H. (1992), *Norsk dialekt-kunnskap*, Novus forlag.

[2] Zue, V.W. (1989), *Speech Spectrogram Reading - An Acoustic Study of English Words and Sentences*, Course at the University of Edinburgh.

[3] Kvale, K., Foldvik, A.K., (1992), "The multifarious r-sound", Proc. International Conference on Spoken Language Processing, pp. 1259-1262.

[4] Kvale, K. (1993), *Segmentation and Labelling of Speech*, Doctoral thesis, Norwegian Institute of Technology.