A Perception Study of Glottalization in St'át'imcets /ľ/

John Lyon
University of British Columbia

Glottalization of resonants is a feature common to Salish languages, but rare cross-linguistically. One major cue to glottalization is creakiness. This study's goal is to determine the perceptual threshold of creakiness for one resonant in St'át'imcets. Creakiness is simulated by lowering the F0 of the resonant. Two dimensions of creakiness are manipulated: duration and place. Manipulated tokens are presented to the subject, and a judgement is elicited as to whether the token contains a glottalized resonant or a non-glottalized resonant. Results indicate that duration of creakiness is correlated with perception of glottalization. Furthermore, creakiness in /ľ/ is more easily perceived in positions which are canonical in production of /ľ/, arguing for a direct correlation between production and perception. Results show an effect of both place and duration of creakiness for perception of glottalization in /ľ/.

1 Introduction

Maddieson (1984) found that glottalized resonants are rare segments cross-linguistically, occurring in only 20 of the 317 languages he examined. They are likewise relatively understudied.

Production of glottalized resonants includes either a laryngeal constriction, a complete glottal closure, or both. The laryngeal constriction lowers the F0 of the vocal folds, and is perceived as 'creaky voice'. Creaky voice is created by "vocal folds that are tightly adducted but open enough along a portion of their length to allow for voicing". "Looking at the waveforms, the creaky phonation is characterized by irregularly spaced pitch periods and decreased acoustic intensity relative to modal phonation." (Gordon and Ladefoged 2001: 386-87). As shown in Figure 1, the first spectogram also exhibits an irregularly spaced pitch period during articulation of the liquid. The laryngeal constriction of a glottalized resonant is accompanied by an oral articulation characteristic of a corresponding plain resonant. Additionally, "the manner of sub-oral articulation and its timing with respect to the oral articulation(s), vary not only across languages but also within languages depending on factors such as syllable and word position." (Bird & Caldecott, 2004).

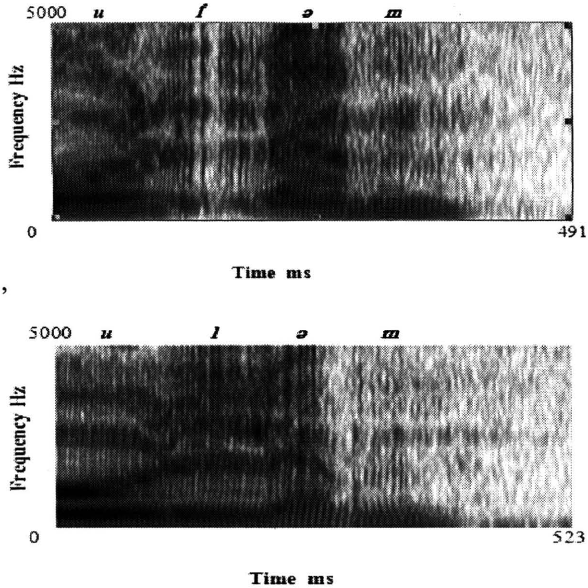

Figure 1. *Spectrograms of [úl̓əm] at the top, and [uləm] at the bottom. Note the wide space between glottal pulses in /l̓/ versus /l/.*

Resonants may either be pre-, mid-, or post-glottalized. With pre-glottalization, the laryngeal constriction precedes the oral articulation. With post-glottalization, the laryngeal constriction follows the oral articulation. With mid-glottalization, the laryngeal constiction occurs simultaneous with (or embedded within) the oral articulation. This variability in timing is in contrast to the articulatory timing of gestures in, for example, glottalized ejectives where the oral articulation must precede the sub-oral articulation in order for the production target to be met. "Under a strong hypothesis based on perceptual recoverability, temporal asynchronies in intergestural timing should generally result from a tendency of speakers to attempt to enhance the perceptibility of each gesture" (Bird, Caldecott, Campbell, Gick & Shaw, in press, 3). Variability in intergestural timing for identical sequences across multiple productions complicates the issue, because different timing strategies result in the same perceptual outcome. The question then arises as to whether one particular production strategy is favored for a specific glottalized resonant in a specific phonological environment, and whether that strategy is also more perceptible.

The present study focuses on the phonemic contrast between /l/ and /l̓/ in St'át'imcets, and the timing differences in /l̓/ perception. I consult data in Bird & Caldecott (2004), as well as measurements of my own data, to determine the favored timing strategy for /l̓/ in St'át'imcets. The results of this perceptual experiment will then compared with the production data, and some light will hopefully be shed on the issue of perceptual recoverability for complex phonemes with variable intergestural timing.

Bird & Caldecott (2004) discuss the distributional properties of glottalized resonants in St'át'imcets, and investigate the relative articulatory timing of glottalization in different syllable positions. Their production data show that while /l̓/ is [+resonant], it patterns with the [-resonant] group in being primarily pre-glottalized. Their Table 5 (p. 331) shows that 68.8% of intervocalic /l̓/ occurences are pre-glottalized, while 31.3% show mid-glottalization. None are post-glottalized. As such, I will consider pre-glottalized position to be the canonical position for the production of intervocalic /l̓/ in St'át'imcets.

This experiment tests a subject's ability to perceive the laryngeal contrast between /l/ and /l̓/ in a minimal pair environment by manipulating the pitch track in a sequence containing /l/, thereby adding creakiness to the /l/ with the expected result that in tokens with sufficient duration of creakiness and a perceptually salient place of glottalization, the token will be perceived as belonging to the pair member with /l̓/.

I am making two specific hypotheses here, both of which may be summarized by my expectation that variations in production will have direct and predictable effects on perception.

The first hypothesis I am making is that the shorter the duration of creakiness, the more likely that the subject will identify the resonant as being non-glottalized. It follows that creakiness of longer duration will more likely result in consistent perception of that creakiness, hence a perception of the resonant as being glottalized.

The second set of hypotheses I am making is that the perceptual threshold of glottalization in terms of duration will be less, when the creakiness cue occurs in a pre-glottalized position. Given that Bird & Caldecott (2004) found that 68.8% of /l̓/ occurences were preglottalized, we might expect that because this is the more common form, it is also more perceptually salient, hence less creakiness will be needed to cue glottalization. Given also that Bird & Caldecott found 0% of intervocalic /l̓/ occurences were postglottalized, I hypothesize that simulated creakiness will have to be of longer duration in post-glottalized position in order to cue creakiness. It is also possible that no amount of creakiness in an intervocalic post-glottalized /l̓/ will result in perception of glottalization, given that it may not occur in naturally spoken St'át'imcets. Also, the perceptual threshold for glottalization in terms of duration will be greater in mid-glottalizated position than in pre-glottalized position, but less than in post-glottalized position. In sum, the production results should mirror the results from this perception experiment.

2 Methods

2.1 Subjects

One female (LT), a fluent speaker of Lower St'át'imcets who is bilingual in English and in her mid 70s participated in this study.

2.2 Bases for Test Tokens and Measurements

Minimal pairs are necessary in a perception experiment in order to isolate one variable (in this case glottalization) and eliminate all other variables which may influence the subject's judgements. The difficulty in finding and testing St'át'imcets minimal word pairs with contrasting glottalized/non-glottalized intervocalic resonants which were familiar to LT forced the researcher to use near minimal pairs. The words [čúntumúləm] *we were told* and [k̓ʷúl̓əm] *to make* served as the bases for comparison in this experiment. Both words have the same VCVC final sequence except that /l/ is glottalized in [k̓ʷúl̓əm] and not in [čúntumúləm]. Note also that both liquids occur in an intervocalic post-stress environment. By extracting the acoustic information from both words, starting from the middle of the /u/ articulation and continuing through to their respective ends, a near minimal environment can be constructed for the purposes of this experiment. Because only the final VCVC sequence [úləm] will be presented to the subject (with varying durations and places of simulated creakiness on /l/), the subject's judgement about the identity of the word to which the sequence belongs should rest only on their perception of whether or not the resonant in the sequence is glottalized.

10 productions of [k̓ʷúl̓əm] and 9 of [čúntumúləm] were recorded using a Marantz PMD660 digital recorder, set at 44.1 kHz. The subject's productions were contiguous and the words were pronounced in isolation. These productions were analysed using PRAAT, and measurements were taken in order to show that the final VCVC sequences were similar enough acoustically to serve as token bases. The average durations of the [úl̓əm] and [úləm] sequences were measured, and their similar durations indicate that articulatory durations of individual phonemes within the two sequences may also be quite similar. F0 at the point of transition from /u/ to /l/ was also measured, as well as F1 and F2 in the surrounding vowels /u/ and /ə/. Finally, the average duration of the articulations of /l/ and /l̓/ in their respective words were measured.

Given the acoustic similarities between [úl̓əm] and [úləm], it was inferred that by manipulating the average F0 of /l/ (186.7 Hz) in a representative production of [úləm] to half its original value (in the range of 90 Hz), creakiness can be simulated, and [úl̓əm] will be the perceived result (see Figure 2). The [čúntumúləm] production selected as the basis for resynthesis shows continuous voicing, and no evidence of irregular glottal pulsing.

The 10 productions of [k̓ʷúl̓əm] were analyzed (see Table 1) in order to determine how the inter-gestural timing of /l̓/ varies across productions, and to compare the data with Bird & Caldecott (2004). Of the 10 productions of [úl̓əm], the average duration of the primary cues to glottalization (creakiness and/or a full glottal stop) were 107.3 ms. 9 of the tokens show cues of 60 ms or longer. 1 token showed no cue to glottalization. 7 of the 9 tokens which were glottalized were cued by both creakiness and a glottal stop. 2 of the 9 tokens which were glottalized showed only creakiness.

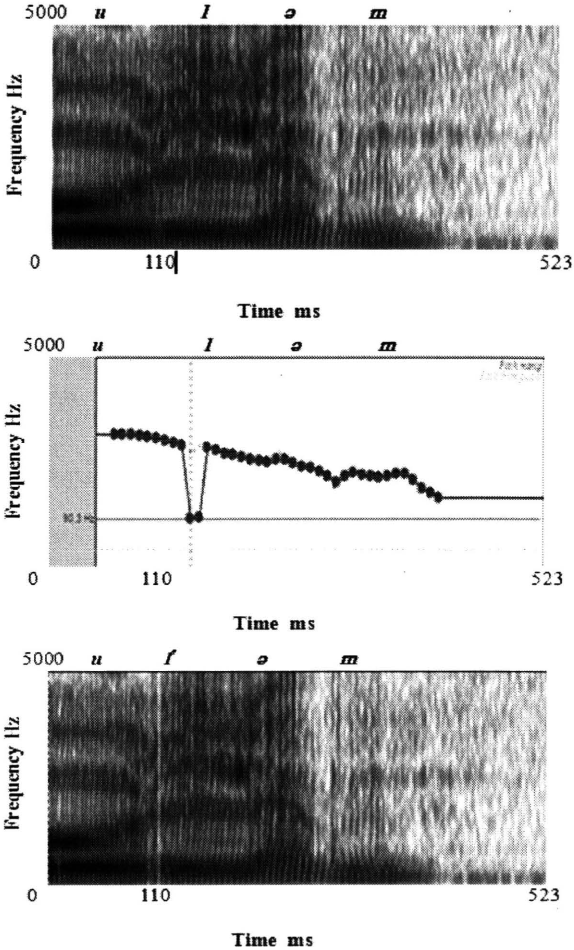

Figure 2. *The top shows a spectrogram of a naturally produced [uləm]; the vertical line at 110 ms is roughly at the [u] to [l] transition. The middle shows the pitch track of [uləm]with 2 pitch points each representing 10 ms, being lowered to simulate creakiness in pre-glottalized position. The bottom shows the result: two light bands beginning at the 110 ms mark indicate a 20 ms simulated creakiness.*

1 production was pre-glottalized, 5 were mid-glottalized, and 3 seem to be post-glottalized, in contrast to the findings of Bird (2004) which showed that pre-glottalized position is favored in production of /l̓/. The 114 ms production had 65 ms of creak, and a 49 ms glottal stop. The cue duration measurement in Table 1 includes both creakiness and the glottal stop, if present. The average durations of glottal cues in pre-, mid- and post- positions are 92ms, 115ms and 131ms respectively.

Glottal cues and place	Cue Duration (ms)
None	10
Creak (mid)	67
Creak (pre)	92
Creak/stop (post)	140
Creak/stop (mid)	114 (65 creak)
Creak/stop (post)	111
Creak/stop (post)	143
Creak/stop (mid)	125
Creak/stop (mid)	131
Creak (mid)	140
Average	107.3

Table 1. *Cue place and duration in 10 measured tokens of [k̓ʷul̓əm]*

Bird (2004) observes that speakers of a language with glottalized resonants vary significantly in their production of these glottalized resonants, not just across speakers, but even across one speaker's productions. These 10 productions of [k̓ʷul̓əm] by LT certainly support this observation.

Based on the variability of cue type and duration shown in Table 1, I assume that the /l̓/ in [k̓ʷul̓əm] may be correctly perceived by either creakiness and a full glottal stop, or by creakiness alone. I also assume that any timing strategy will result in the correct perception of glottalization, although in deference to Bird (2004), I will adhere to my second hypothesis which assumes post-glottalized position as the least favored position for creakiness in an inter-vocalic post-stress environment. Finally, I assume that 67 ms of creakiness is of sufficient length to cue glottalization, and now ask the question: Are smaller durations of creakiness also sufficient to cue glottalization in resonants?

2.3 Experiment Design

A short pre-test was conducted prior to the main experiment to determine if duration and place of glottalization was likely to affect perception.

Comparing the base [úləm] token to the 67 ms creak production of [ul̓əm] (see Figure 1), the 110ms mark in both tokens is the point of transition from /u/ to liquid, and as such this was made the reference point for constructing and testing stimuli with creakiness in pre-glottalized position. Similar reference points were established for mid- and post-glottalized position; 140ms and 170ms respectively.

2.3.1 Stimuli

Using the PRAAT manipulation function, the pitch track of [úləm] was lowered beginning 110 ms from the beginning of the token (at the beginning of the /l/ articulation) to simulate pre-glottalized creakiness (Figure 2). Each pitch point represents 10ms, so in order to simulate a 60 ms creakiness duration, for example, 6 pitch points were lowered to half their original value. The manipulation was then exported as a sound file into PRAAT objects, and saved as WAV files to be used as the test stimuli.

Different durations of creakiness were synthesized, ranging from a minimum of 10 ms to a maximum of 60ms, which is less than the minimum duration of creakiness exhibited in the 10 analysed productions in Table 1.

Two different oral/laryngeal timing strategies were simulated and tested in the pre-test phase: pre- and post- position; while all three timing strategies were simulated and run in the test-phase. As mentioned previously, pre-glottalization is simulated by inserting creakiness 110 ms from the beginning of the test token. Post-glottaliztion is simulated by inserting creakiness 170 ms from the beginning of the token, and is situated towards the end of the /l/ articulation and overlaps into the /ə/ articulation. Mid-glottalization is simulated by inserting creakiness halfway between pre- and post- position at 140 ms. 6 tokens in pre- and post-position were re-synthesized and tested during the pre-test phase, yielding a total of 12 pre-test tokens to serve as stimuli. These 12 pre-test tokens were tested in 3 blocks of trials.

During the test-phase, pre-, mid-, and post-glottalization were tested for the 10, 20, 40, and 60 ms creakiness durations, and only pre-glottalized position for the 30 and 50 ms creakiness durations. This yielded a total of 14 test tokens per trial. The 30 and 50 ms tokens were introduced into the test only after the 4th trial, so these were run only 6 times each and were tested only in pre-glottalized position. A total of 132 tokens were presented to the subject over 3 sessions. 4 trials each were completed during the first 2 sessions, and 2 trials during the final session, making a total of 10 trials. It was determined that attempting to conduct more than four trials in one sitting would likely result in decreasingly reliable judgements due to fatigue on the part of the subject.

2.3.2 Procedure

LT listened to each test token individually through headphones, and her judgements were recorded by the researcher. The stimuli were presented in a different order for each trial. They were not randomized electronically, but were all listed on the PRAAT objects screen, then chosen by the researcher without any pre-meditation. The subject then made a judgement as to whether the stimulus sounded like it came from [k̓ʷúl̓əm] or [čúntumúləm].

During the first 2 pre-trials, LT was played each token as many times as necessary in order for the subject to feel confident about her judgement. During the last pre-test trial and throughout the entire test phase, the researcher selected the token to be played as before, but LT controlled the number of times she exposed herself to the stimulus. In other words, she pushed the "play" button herself. Given the subject's age and hearing impairment, it was judged preferable for the subject to be able to hear each stimulus as many times as she felt necessary to give a confident judgement, rather than control for the number of stimulus presentations, and have the subject be unsure of her judgement. LT often heard the stimuli three or more times before making a judgement.

There are two problems with the pre-test tokens which were remedied in the actual test tokens. First, pre-test tokens did not start at an acoustic 0 point, where the wave form crosses the x axis. As a result, a slight "pop" precedes the /u/ vowel in each token. To guard against this "pop" possibly causing an echo effect or shifting perception towards glottalization, in the actual test data, all tokens start at an acoustic 0 point. Second, in the pre-test tokens, the creakiness of the 40 ms pre-glottalized token, for example, starts at the 110ms pitch point. In the actual test, while the creakiness of the 60 ms pre-glottalized token starts at the 110 ms mark, the creakiness of the 40 ms pre-glottalized token starts at the 120ms mark, and the pre-glottalized 20 ms token starts at the 130ms mark. Thus, a 20 ms creakiness token is *centered* within the 40 ms token, and likewise is the 40 ms token centered within the 60 ms token. This is in contrast to the pre-test stimuli, which all started from the beginning of the reference mark (i.e. 110 or 170 ms), regardless of duration. The choice as to whether to center the creakiness or not may be purely methodological, however it seems intuitively preferable for the three glottalized positions to have centers of reference, rather than reference points skewed to the left or right. For the two reasons discussed above, I have chosen not to combine the results of the pre-test and actual test.

3 Results

The results of the pre-test (Figure 3) generally confirm hypothesis 1, namely that shorter duration of creakiness, the more judgements that the token is non-glottalized. A grey area is manifest at between 30-40 ms creakiness, where judgements were most variable. Although tokens were judged to be glottalized 33% and 66% of the time for simulated pre-glottalization of 10 and 20 ms respectively, the same durations in simulated post-glottalized tokens were *never* perceived as being glottalized.

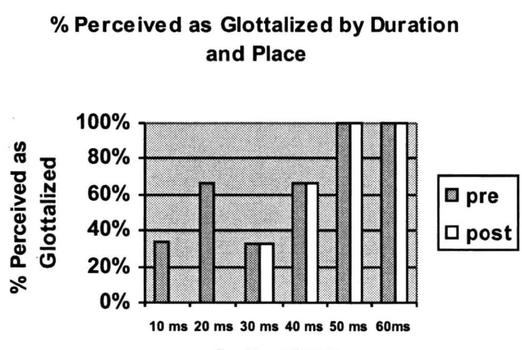

Figure 3. *shows the results of the 3-trial pretest.*

The results for the actual test are given in Figure 4. It can be seen that for mid-glottalized position, positive judgements climb linearly from 20% at 10 ms creakiness duration to 50% with 20 ms, 70% with 40 ms, and 90% with 60 ms. Increasing creakiness duration in post-glottalized position did not result in an increase in the number of positive responses. The results for pre-glottalized position are more variable.

A chi-square test was calculated using JMP software in the UBC ISRL. The probability that place of glottalization effects the perception was determined to be significant (χ^2=13.47; p = 0.0012). The probability that duration of glottalization affects the perception of /l̓/ was also determined to be significant at

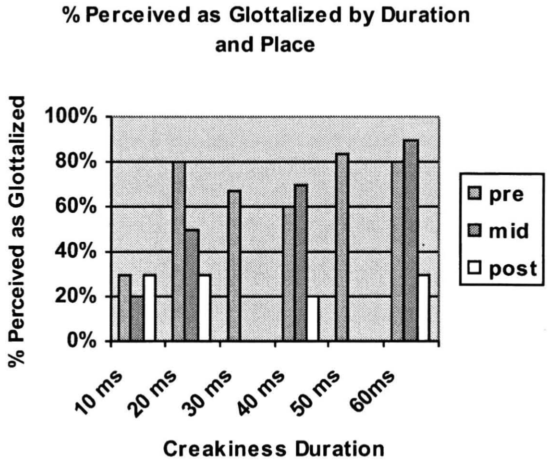

Figure 4. *shows the results of test-phase. Note that increasing duration results in increased rate of perception. Post-glottalized position is perceived significantly fewer times.*

(χ^2=7.89; p < 0.005). The probability that place and duration conspire to effect perception was found, during the final analysis, to be not significant (χ2=5.58; p = 0.062).

4 Discussion

4.1 Discussion of Results

Pre-test results generally confirm Hypothesis 1. Greater duration of creakiness results in more [k̓ʷul̓əm] judgements.

Looking at Figure 3, it is interesting that on 2 out of 3 pre-trials, LT perceived the 20 ms pre-glottalized token as belonging to glottalized [k̓ʷul̓əm], whereas the 20ms post-glottalized token was consistently perceived as belonging to non-glottalized [čúntumúləm]. Given Bird's observation showing that intervocalic post-glottalized /l̓/ is perhaps not realised in St'át'imcets speech, the pre-test results support Hypothesis 2, in that the 20ms of creakiness is able to perceptually cue glottalization in its naturally occurring pre-glottalized position, but is not sufficient in the unnatural post-glottalized position. Similar, though less striking results, are shown for the pair of 10ms tokens.

The results of the actual test (Figure 4) show that the generalization formulated as Hypothesis 1 is supported. The longer the duration of creakiness, the more often the test stimulus is perceived as being glottalized. Figure 4 shows that for a 60 ms creakiness duration in pre-glottalized position, 80% of the tokens were perceived as being glottalized, versus only 30% with a 10 ms creakiness. It was not expected, however, that a pre-glottalized 20 ms creakiness duration would result in more positive judgements than 40 ms. It is likely that more trials will result in a linear correlation, and that a positive correlation between positive judgements and increasing duration will continue to hold.

For mid-glottalized position, positive judgements climb linearly from 20% at 10 ms creakiness duration to 50% with 20 ms, 70% with 40 ms, and 90% with 60 ms. This is as expected.

Perhaps most interesting is the fact that longer durations of creakiness in post-glottalized position did *not* result in more judgements of glottalization. This speaks strongly in favor of Hypothesis 2, and in favor of Bird & Caldecott's (2004) data which does not show intervocalic /l̓/ occurring with post-glottalization. If post-glottalized /l̓/ is either not naturally produced or at least non-canonical, this fact might also explain why it is not easily perceived.

Although in the final analysis it was found that there is no significant combinatorial effect of place and duration on perception of glottalization, a preliminary chi-square test which was run after trial 8 showed that the effect *was* significant (p = 0.047)), so it is possible that additional trials will show that there is in fact a combined effect of place and duration on the perception of creakiness as a cue to glottalization.

4.2 Future Research and Implications

Subsequent research should attempt to test judgements between *natural* minimal pair words, instead of near minimal pairs as this experiment has done. It will likely be much easier for the subject to come to a judgement, since they will not have to mentally place the stimulus within its larger, corresponding word. Ideally, lexical and grammatical class should also be controlled for. The [úl̓əm] in [k̓ʷúl̓əm] consists of the final VC of a root √k̓ʷúl̓ followed by an intransitivizer morpheme *-əm,* while the [úləm] in [čúntumúləm] consists of the final VC of 1st person plural passive ending followed by a passive morpheme. It is also important that more than 10 productions of the base words be analysed. It may be that the productions which I assume are post-glottalized are simply anomalies, or variants of mid-glottalized productions. Finally, to say anything definitive about the perception of glottalized resonants in St'át'imcets, it will be necessary to test all glottalized resonants in all possible syllabic environments.

There are both perceptual and articulatory pressures which determine the timing of laryngeal gestures. Perception is made more difficult if the creak obscures place-of-articulation information in a CV or VC transition. (Howe and

Pulleyblank 2001). Scott (2006) notes a conflict between these pressures for resonants in intervocalic position. Intervocalically, onset cues are the most important for perception (Howe and Pulleyblank 2001) so to cue a resonant as being glottalized, we would expect pre-glottalization. "On the other hand [in an intervocalic post-stress] sequence articulatory pressure would want the creak to be further away from the stressed vowel and so we would expect a sequence of voice-creak on the sonorant in this situation. Van Eijk (1997) claims that this is the situation in St'at'imcets." (Scott, 2). It is important to address these "pressure conflicts". Further perceptual studies may help answer the question as to whether the acoustic salience of a resonant onset is more important as a cue to glottalization than yielding to natural articulatory pressure, which in this case would move the creakiness in /l̓/ as far away from stressed /ú/ as possible. Figure 4 shows that pre- and mid- positions are favored more or less equally, while post- position definitely is not. This could support the hypothesis that preserving onset cues is more important than yielding to articulatory pressure.

It would be interesting if these results could be replicated with fluent speakers of St'ất'imcets other than LT. Similar results in other speakers would suggest that these findings do not result from idiosyncracies, but are instead shared by all speakers of the language. It will also be necessary, at some stage, to test non-speakers or second language learners of St'ất'imcets to determine to what extent these cues to perception of glottalization are learned, and to what extent they are the result of innate articulatory pressures. For example, if non-speakers of St'ất'imcets, who speak a language lacking a phonemic distinction between /l/ and /l̓/, perceived creakiness equally well (or poorly) in canonical (pre- and mid- position) and non-canonical positions (post- position), it might support a hypothesis that a child loses the ability to perceive these place distinctions at some point during their language development because there is no phonemic distinctions between /l/ and /l̓/, and canonical versus non-canonical timing strategies are therefore not salient.

Löfqvist (1984) notes that "speech perception is a goal directed behavior. Speech movements are executed to produce an acoustic signal that can transmit linguistic information. If relative timing is an important part of motor control, it is reasonable to assume that traces of relative timing can be found in the acoustic signal, and would thus play a role in the perception of speech" (288). As seen in glottalized resonants in St'ất'imcets, relative timing varies considerably from both an articulatory and an acoustic standpoint. If the same linguistic information can be transmitted through different acoustic signals, and there is no one-to-one correspondence between production and perception for segments with variable timing, then this fact certainly needs to be addressed in any integrated model of speech perception and production.

5 Conclusions

Although this experiment is very limited in its scope, it does add much-needed perceptual information to our knowledge base of glottalized resonants and their production. As an initial effort in the direction of understanding the relationship between production and perception of these complex articulations with multiple, varying strategies, it is hoped that this experiment will provide impetus for subsequent work.

In summary, this experiment has found that duration of creakiness is correlated with the likelihood that the stimulus will be perceived as being glottalized. This experiment has also found that place of creakiness has a positive correlation with perception of glottalization. More specifically, the production data from Bird & Caldecott (2004) showing no intervocalic post-glottalized /l̓/ productions seems to be mirrored, to some extent, by the perception data of the present study which shows that post-glottalization of intervocalic /l̓/ is also not easily perceived.

References

Bird, Sonya & Caldecott, Marion. 2004. *Timing Differences in St'át'imcets Glottalised Resonants: Linguistic or Bio-Mechanical.* Proceedings of the 10th Australian International Conference on Speech Science & Technology, Macquarie University, Sydney, December 8-10, 2004. p. 328-333.

Bird, S. Caldecott, M. Campbell, F. Gick, B. & Shaw, P. in press. *Oral-Laryngeal Timing in Glottalized Resonants.* Journal of Phonetics.

Caldecott, Marion. 2000. *unpublished festschrift on glottalised resonants in St'át'imcets.*

Gerfen, Chip & Baker, Kirk. 2005. *The Production and Perception of Laryngealised Vowels in Coatzospan Mixtec.* Journal of Phonetics. vol 33. p. 311-344.

Gordon, Matthew and Peter Ladefoged. 2001. *Phonation types: a cross-linguistic overview.* Journal of Phonetics 29: 383-406.

Howe, Darin and Douglas Pulleyblank. 2001. *Patterns and timing of glottalisation.* Phonology 18: 45-80

Löfqvist, Anders & Yoshioka Hirohide. 1984. *Intrasegmental Timing: Laryngeal-Oral Coordination In Voiceless Consonant Production.* Speech Communication 3: 279-289

Scott, Mark. 2006. *Glottalized Sonorants in St'át'imcets.* UBC Linguistics Field Methods Class Research Paper.

van Eijk, Jan. 1997. *The Lillooet Language: Phonology, Morphology, Syntax.* UBC Press.

John Lyon
johnlyon@interchange.ubc.ca