SEGMENTAL INVARIANCE RECONSIDERED

R.G. Daniloff, Purdue University, W. Lafayette, Indiana, USA
M.A.A. Tatham, Languages and Linguistics, University of Essex, Colchester, UK

Stop consonant articulation is context sensitive, Fromkin (1966). Such context sensitivity, often related to coarticulation, is taken as support for complex, high level "encoding", Kozhevnikov and Chistovich (1965). The duration and force of labial closure have been variously shown to be context sensitive to (1) syllable position, (2) voicing, (3) stress and (4) vocalic context. We concluded that previous cinefluorographic, electromyographic, and palatographic studies may have overestimated the extent of context sensitivity by failure to control for tempo, speech effort, task learning, and sentential context. The purpose of the study reported herein was to reassess the context sensitivity of the EMG impulse associated with labial-stop closure.

Procedures

4 adults served as subjects. Each was required to repeat previously tape recorded sentences as he heard them over earphones. The tape recorded sentences were carefully controlled for constant tempo, stress levels, and good phonetic quality. Speech tokens consisted of all combinations of C_xVC, CVC_x, C_xVC_x, 'VCV, and V'CV tokens, C_x = [p,b], C = [d], V = [i,u,ʌ]. CVC items were spoken in the sentence, "He'll spoof the [CVC] again", and VCV items in, "Smell this poof of [VCV] again", such that each token received primary, sentence level stress. Subjects visually monitored their vocal output as they spoke, keeping it at 60-65 dB, SPL. Each subject practiced the task of repeating the tape recorded sentences with as controlled a tempo, vocal output, and desired stress level as possible.

Bipolar silver disc surface electrodes on the upper lip served to detect the EMG pulse on the orbicularis oris for the lip closing gesture for [p,b]. The EMG and Raw Voice signals were recorded, rectified, and integrated, and on a single, 5 channel mingograph output, raw/integrated EMG, voice, and timing signals were displayed. The upper lip-surface electrode array was chosen because: (1) the upper lip is accessible, (2) labial surface EMG signals are quite interpretable-reproducible, (3) myo- and biomechanically

the lip in its closing gesture is a simple system, (4) the labial closure gesture has been extensively studied, (5) surface electrodes potentially yield a better estimate of whole-muscle activity than do needle/hooked wire electrodes, (6) labial closure is reputedly context sensitive.

Subjects received extensive practice beforehand; subjects who could not relax the lips to nearly zero-baseline EMG activity between utterances were rejected. Each of 40 tokens was repeated 26 times for a total of 1040 repeated, randomized sentences spoken at one sitting, with pauses every 10 minutes for relaxation.

Results:

The results of the study are based upon 4 criterion measures: the peak amplitude of the EMG pulse for lip closure, the duration of the EMG pulse, the delay between EMG onset and acoustic burst release, and the delay between peak EMG level and burst release. Data were analyzed using 3 way ANOVAs, with a conservative $\alpha \leq 0.1$. The concise results are shown in the table below. Interpretation of the results is based upon the following assumptions: lip closure for stops is mediated primarily by orbicularis oris contraction (00). M.O.O. force of contraction and the height of the integrated EMG signal are linearly, if not monotonically related. If one of the particular aspects of context-sensitivity being investigated were a part of speakers' linguistic competence, it was our expectation that all 4 speakers should show a statistically significant and similar shift in the labial closure criterion measure associated with that aspect of context since by our estimate, 25 repetitions of each token offered a firm basis for statistical inference.

The peak EMG amplitude measure presumably relates directly to maximum force of M.O.O. contraction. As shown, in every case, one or more subjects for one or more tokens showed a non-significant change in EMG peak amplitude; and in fact, for all 4 contextual effects: voicing, syllable position, vowel, stress - at least one subject showed a reversal of trend, with differences in EMG peak being just opposite those shown for 2 or 3 of the subjects. Thus, there is only a modest trend for voiceless stops in /i/ context to be modestly more effortful, muscularly. Stress and syllable position had no consistent effect upon peak EMG amplitude. Duration of the EMG pulse revealed a strong dependence upon context such

	VOICE	POSITION	VOWEL	STRESS
Peak Amplitude EMG	-voice>voice; moderate effect; subject, token dependent	initial=final; weak effect; very subject, token dependent	/i/>/ʌ/>/ɯ/, moderate effect; subject, token dependent	not consistent; weak effect; strongly subject dependent
Duration of EMG Pulse	-voice>voice; strong effect; little subject or token dependence	initial>final; strong effect; small subject, token dependence	not consistent; moderately weak effect; very token, subject dependent	V2 stress>V1 stress; strong effect; small subject dependence
Delay EMG Onset to Burst	voiceless final> voiced final; moderate effect; token, position dependent	initial>final; strong effect; small token, subject dependence	/u/>/i/ weak effect; strongly token, subject dependent	V2 stress>V1 stress; strong effect; small subject dependence
Peak EMG to Burst Onset	voiceless>voiced, moderate effect; subject, token dependent	voiced initial> voiced final; moderately strong effect; strong voicing, moderate subject dependence	not consistent, weak effect; strong token, subject dependence	V2 stress>V1 stress; moderate effect; small subject dependence

that voiceless stops were longer than voiced, initial stops were longer than syllable final stops, and pre-stress-position stops were longer than post-stress stops. In all three cases, tokens for one subject failed to achieve significance. In addition, for the voicing effect, one subject's voiced stops were significantly longer than his voiceless tokens. For the time delay between EMG onset and burst release, the effect of syllable position was fairly strong in that all initial stops began earlier and, in 6 of 8 cases, the difference was significant. For voicing, there was a modestly strong trend for voiceless stops to begin earlier, in 11 of 12 cases, with 10 of the 12 cases being significantly greater. Stress had a strong effect in that the pre-stress stop began earlier for all 4 subjects, significantly so in 3 of 4 cases. Vowels had no consistent effect upon delay. The peak EMG to acoustic release temporal delay measure shows a weak dependence upon voicing; the effect of the vowel upon delay was weak and inconsistent. The effect of stress upon this measure was only moderate in that for all subjects, pre-stress stops had earlier occurring peaks, but these differences were significant in only two of four subjects. The effect of position upon this delay measure was moderately strong in that in all 8 cases, syllable initial vowels had earlier occurring EMG peaks, and in 6 of 8 cases, the differences were significant.

Conclusions

Contrary to the work of Fromkin, syllable final stops were shorter in all cases, significantly so in 6/8 cases, than initial stops. Syllable position had no consistent effect upon the amount of muscle activity for closure, but initial stops began earlier, vis-à-vis onset or peak of EMG and burst release, in all cases, and significantly earlier in 12 of 16 cases. Thus, syllable initial stops are generally longer and earlier in onset, but not muscularly more effortful. Vowel context effects were enigmatic. It was expected that one would have earlier and stronger EMG pulses for [ɪ] than for [ʌ] than for [u]. This was not the case; in the majority of cases, vowel context had non-significant effects on most EMG measures, and even when significant, the direction of the trend varied from subject to subject. Stress was potent as a factor such that pre-stressed stops began earlier in all cases, significantly so in 5 of 8 cases, and in 3 of 4 cases, the EMG

pulse was significantly longer. However, stress had no systematic effect upon the amount of muscle activity needed for closure. Finally, voice as a factor had no systematic, cross subject effect upon amount of muscle activity. With only one reversal (significant), in 9 of 12 cases, voiceless stops were longer than voiced stops, and EMG onset began earlier, vis-à-vis release, in 10 of 12 cases, one exception being a significant reversal. The effect of voice upon the EMG peak to burst release measure was highly variable.

The most startling result was that without exception, at least one subject, for at least one token showed either a non-significant context effect, or a reversal of trend for a given context effect. According to a strict criterion of all subjects and all tokens revealing a significant change in criterion measure, then, not one of the contextual effects investigated is less than idiosyncratic, i.e.: the contextual effects are a trend, but not absolutely a component of linguistic performance. Further analysis showed that subject sex and naïveté had no effect upon the results. Surprisingly, the phonetic shape of the syllable, e.g.: C_xVC_x vs. CVC_x vs. VCV had a profound effect upon all criterion measures, and was probably the single most potent effect found in this study. It is difficult to explain why vowel context did not produce more, and more systematic changes in the size and timing of the EMG patterns for labial closure. It may be the case that coarticulation of stops and vowels, known to be quite a strong effect, is highly idiosyncratic. Or, it may be the case that the transformation of muscle activity into final vocal tract shape is complex, and non-linear, within and across subjects, so that interpretation and comparison of EMG data are more complex than is suspected. It is our conclusion that labial closure as an articulatory gesture is relatively context insensitive as far as amount of muscle activity is concerned. It is context sensitive as far as syllable position, voicing and stress are concerned in that voiceless, initial, pre-stress stops are generally longer and begin earlier: however, certain subjects and tokens violate this trend. We conclude that electromyographic signals, especially vis-à-vis coarticulation, may be more complex to interpret than is presently suspected.

References

Fromkin, V.A. (1966): "Neuromuscular specification of linguistic units", L&S 9, 170-199.

Kozhevnikov, V. and L. Chistovich (1965): Speech: articulation and perception, JPRS 30, 543, U.S. Bureau Commerce, Washington, D.C.