OSU Working Papers in Linguistics 50 (1-10)

Perception of Consonant Clusters and Variable Gap Time*

Mike Cahill
cahill@ling.ohio-state.edu

Abstract: In every case in which measurements of labial-velar stops [kp, gb] have been made, it has been found that the labial and velar gestures are not strictly simultaneous, but rather that the velar gesture slightly precedes the labial one (thus [k͡p] and not [p͡k]). One possible explanation for this is that [k͡p] is more perceptually salient than [p͡k]. This paper reports an attempt to test this hypothesis by observing listeners' identifications of [apka] and [akpa] with variable gap times inserted between the consonantal onset and release. The results showed that [apka] was more readily identified than [akpa], effectively showing that perceptual salience cannot be invoked to explain the ordering of velar and labial gestures in labial-velar stops.

INTRODUCTION

Labial-velar stops [kp], [gb] occur in many languages from central and west Africa, where the bulk of them are found, as well as in a handful of languages in and around Papua New Guinea.[1] They are commonly described as having "simultaneous" closure at the labial and velar places of articulation. However, most transcriptions have recorded them as [k͡p] and not [p͡k], and this is no accident. Spectrographic evidence shows that a vowel preceding a labial-velar stop makes a transition into a velar component, and the release of the consonant has labial characteristics, in languages as diverse as Dedua (from Papua New Guinea) and Efik (Ladefoged & Maddieson 1996) and Ibibio (Connell 1994), both from west Africa. Also, Maddieson has presented direct evidence that in Ewe, at least, the labial gesture both starts and ends later than the velar gesture, as in Figure 1 below, taken from electromagnetic articulography data in Maddieson (1993).

* My very great thanks to Keith Johnson, who provided invaluable aid in the design and implementation of every stage of the experiment reported herein.

[1] A few Creole languages of the Caribbean, such as Ndyuka of Surinam, also have labial-velar stops, presumably as a result of African language substrata (Huttar & Huttar 1994).

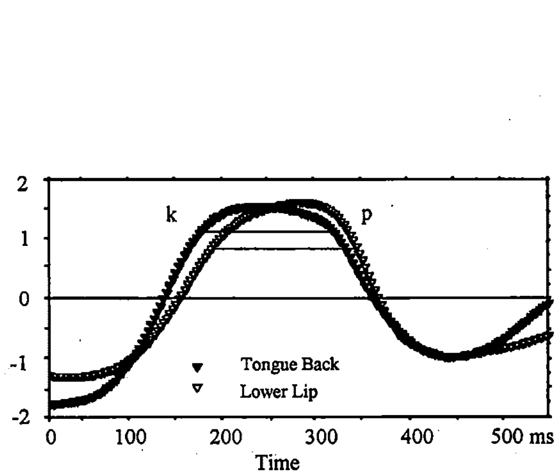

Fig. 1 Coordination of lower lip and tongue back movements in the Ewe word **akpa**. Y-axis is vertical displacement; horizontal lines indicate the likely duration of actual contact of the articulator. (Maddieson 1993)

In this paper, then, [k͡p] and [p͡k] will both refer to articulations that are mostly overlapping, but in [k͡p] the labial gesture follows the velar one, as above, while in [p͡k], the labial gesture would precede the velar one. I will focus on the voiceless stop [k͡p] here, though the discussion is also applicable to the voiced labial-velar stop [g͡b].

One of the questions arising out of research on labial-velar stops is why this partial or incomplete overlap should exist, rather than total simultaneity. Also, why should it be that universally (as far as we know) there is an asymmetry of gestural overlap and that this asymmetry should always be in the same direction?

At least three possibilities exist and are worthy of consideration. One possibility is that there is an "ease of articulation" factor, that is, that [k͡p] requires less effort to produce than [p͡k]. One might argue that with the condyles of the jaw acting as a pivot, especially if a consonant is pronounced after a vowel, it would be more natural for the articulators closer to the pivot point to make contact sooner than those further away; so a consonant made with a place of articulation further back in the mouth would be more likely to precede a consonant made in the front of the mouth. This could be a physiological explanation of the data in Hume (1996), who gives examples from several languages in which metathesis of consonant clusters operates to give an output in which the more posterior consonant precedes the more anterior one. She proposes a phonological constraint in which the more posterior of a consonant cluster pair is favored to precede the other. Arguing from the "ease of articulation" viewpoint, however, is notoriously suspect, since languages of the world abound in sounds which are not simple to produce. Sounds such as implosives, clicks, ejectives, and complex consonant clusters come to mind. A perusal of Ladefoged & Maddieson (1996) yields an abundance of examples. One may be able to argue persuasively and theoretically that certain sounds are, in fact, more difficult to make than others, but the existence of difficult sounds in languages of the the world makes this argument a tendency rather than a robust explanation for the phenomenon. Also, what is judged as "difficult" largely depends on the inventory of sounds in the speaker's native language compared to the language under consideration. I judge the [kp] found in most Ghanaian languages to be fairly difficult, while the Ghanaian takes the [k͡p] in stride but judges the phonetics of American English 'squirrel' to be very difficult.

Another possibility is that since the historical development of labial-velars seems commonly, perhaps always, to be the reflex of a labialized stop, this labial release has been maintained in modern languages. The two main sources of /kp/ historically seem to be sometimes ***pw**, exemplified by Aghem (Hyman 1979), but more often ***kw**, exemplified by the Sawabantu group of languages in western Cameroon, (Mutaka and Ebobissé 1996). These and other examples are examined in Cahill (in prep). In both of these proto-forms, the release is labial, whereas the start of the consonant, at least in the case of ***kw**, is not. The possibility is that the asymmetry present in the proto-form, that is, the labiality being skewed more to the release, is preserved in the synchronic reflexes.

A third possibility is that, perceptually, a [k͡p] is more salient than a [p͡k]. This implies that [k͡p] is easier to perceive in some way than a [p͡k].[2] Again, Hume's constraint favoring consonant clusters with the more posterior consonant occurring first would be a way of expressing this tendency in phonological terms. But, as with the "ease of articulation" possibility, languages abound which have hard-to-hear sounds. These would include creaky and breathy vowels, different fricatives, labial-velars themselves, and a host of others. Similarly to the point made for difficult articulations, sounds which are judged "hard to perceive" are mainly those which do not occur in the hearer's native language.

It is of course possible that more than one of the above factors could be at work here. For example, [k͡p] could have developed for historical reasons, then remained as such for ease of articulation. Each possibility also has its own set of objections. In addition, it is hypothetically possible that [k͡p] exists rather than [p͡k] merely as an accident of language, though the universality of [k͡p] makes this scenario rather dubious. But if we assume that there is a reason (or reasons) behind the asymmetry of labial-velars, then it should be possible to investigate what that reason is, despite any initial difficulties.

The experiment reported here was an attempt to test the third hypothesis. More specifically, the hypothesis this experiment addressed was that the reason why the partial overlap of labial-velars is always skewed in the direction of labial release is that a labial release and velar onset is more perceptually salient than a velar release and labial onset. To test this, we spliced together sequences of [kp] and [pk], with varying gap durations, and tested to see which was more readily identifiable.

The result of the experiment did not support this hypothesis, but showed rather that [pk] was the more salient of the two clusters.

---

2 Chomsky and Halle (1968), in reasoning about perception of multiply-articulated sounds, get the phonetics precisely backwards with respect to labial-velar stops. They write, "The order of release of the different closures is governed by a simple rule. In sounds without supplementary motions [i.e. movement of the glottis during the period of closure- mc], the releases are simultaneous. In sounds produced with supplementary motions, closures are released in the order of increasing distance from the lips. The reason for this ordering is that only in this manner will clear auditory effects be produced, for acoustic effects produced inside the vocal tract will be effectively suppressed if the vocal tract is closed." (1968:324). This predicts labial-velars with a simple pulmonic airstream should release both closures simultaneously, while labial-velars with an ingressive velaric airstream should release the labial closure first. However, in both cases, it is the labial closure which is released last (see Ladefoged 1968, Painter 1970).

METHOD

Spectrographic studies of labial-velar stops have shown a release burst characteristic of labial stops, but a transition from the preceding vowel characteristic of velar stops (Ladefoged 1968, Garnes 1975, Connell 1991, 1994, Ladefoged & Maddieson 1996). So splicing together the consonants [k] and [p] gives a reasonable facsimile of a labial-velar stop.

To produce the test sounds, the author recorded several tokens of the syllables [ap], [ak], [ka], and [pa] in a soundproof recording booth using a Marantz 220 tape recorder, with a Sure SM-48 microphone. Representatives of the appropriate tokens were then spliced together using the CSpeech program to form the output tokens [ap-pa], [ap-ka], [ak-ka], and [ak-pa]. (These stimuli will be referred to below with capitals, e.g. APKA.) Silent durations of 0-200 ms, in 25 ms increments, were inserted between the offset of voicing in [aC] and the release burst of [Ca]. For the [ap-ka] and [ak-pa] tokens, intervals were extended to 400 ms as well. A VisualBasic program was set up, so that listeners heard the tokens in random order over headphones, and selected either *apa*, *aka*, *akpa*, or *apka* as the closest to what they heard.

15 listeners participated in the experiment, all undergraduate students from Ohio State University. All but one had English as their mother tongue and were from Ohio. The exception was a Jordanian student whose first language was Arabic, but her responses were not markedly different from the others, so they are included as well.

The listeners were seated in a soundproof booth, with a computer screen in front of them. The program played the token, and the subject used the computer mouse to click on a button on the screen labeled **apa**, **aka**, **akpa**, or **apka**. There was a 2-second interval after they clicked before the next token was played. The 47 tokens were randomized; when one block of 47 trials finished, another block began. The same set of tokens was repeated in this way four times, with a different randomized order each time, for a total of 168 total tokens presented to each subject. The experiment was self-paced, with a token not presented until a response was given to the previous one. The total time for each run of the experiment ranged from 25-40 minutes.

RESULTS

Figure 2 shows the two most common responses to the AKPA stimuli. The main trend is that at shorter gap durations, the AKPA stimulus was perceived as "apa." As the gap duration increased, the perception of "akpa" also increased. The two responses were approximately equal at about 80ms, as measured by the crossover point of the two plotted curves. As expected, virtually all the non-"akpa" responses were "apa", having the same release as the stimulus; therefore, the very few responses which were "aka" or "apka" are not plotted.

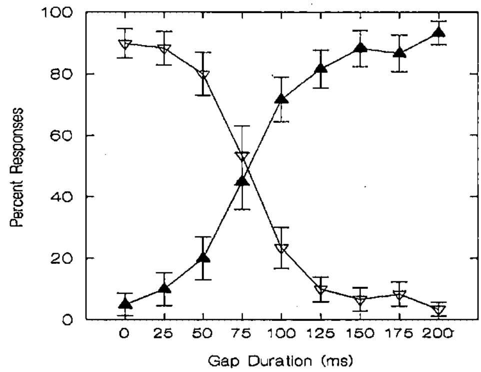

Fig. 2 Responses to AKPA input: filled = "akpa", open = "apa"

Figure 3 shows the responses to the APKA stimuli. Similar to the above, for shorter intervals, the "aka" response was more often given; as the gap interval increased, the "akpa" response was increasingly given. As expected, virtually all the non-"apka" responses were "aka", having the same release as the stimulus; therefore, the very few responses which were "apa" or "akpa" are not plotted in Fig. 3. In comparison to the AKPA stimulus, the APKA stimulus was correctly identified at a shorter gap duration; the crossover from "aka" to "apka" occurred at only 25 ms (compared to 80 ms for AKPA).

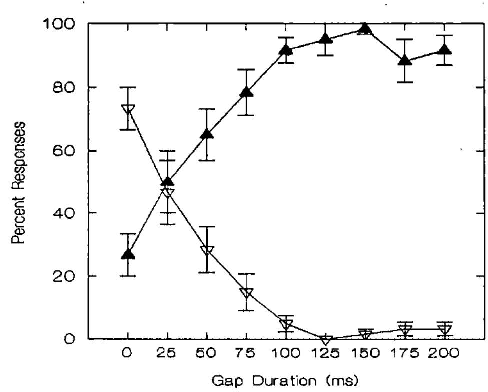

Fig. 3 Responses to APKA input: filled = "apka", open = "aka"

Figure 4 contains the "akpa" and "apka" curves from Figures 2 and 3, showing directly that the "apka" response to APKA was chosen at shorter gap durations than the "akpa" response to AKPA.

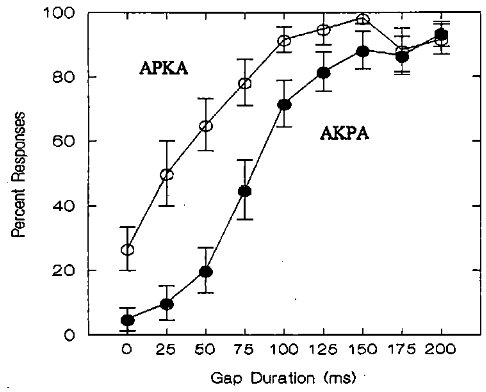

Fig. 4 Responses to KP vs. PK stimuli: filled = AKPA/akpa, open = APKA/apka

A result which was unexpected was that for both the AKKA and APPA stimuli, at very long intervals, subjects occasionally identified the stimulus as a heterogeneous cluster "apka" or "akpa." This is shown in Figs. 5-6. As above, the release consonant of the response was the same as the stimulus for almost all responses. The bulk of the mis-responses was from three subjects, but there was some scattered similar response from others as well.

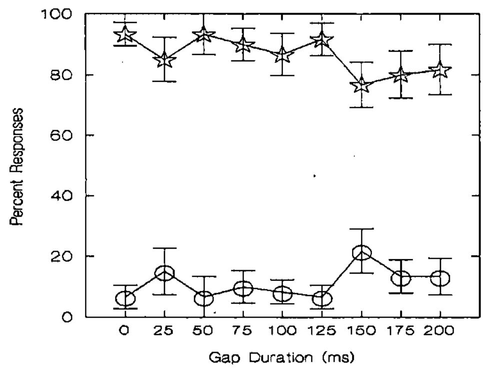

Fig. 5 Responses to AKA stimuli: star = "aka", circle = "apka"

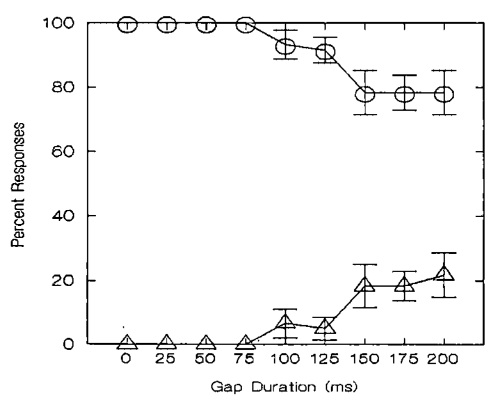

Fig. 6 Responses to APA stimuli: circle = "apa", triangle = "akpa"

DISCUSSION

Several conclusions may be drawn from this data.

First, at long gap durations the subjects sometimes heard a phonetic geminate as a cluster of heterogeneous consonants (Figs. 5-6). In English, geminate stop consonants are rare (the *k:* in *bookkeeper* being one example), and do not contrast with non-geminates in non-compound words at all. Heterogeneous clusters, on the other hand, are more common. The hearers, in common with any speakers of one particular language hearing unfamiliar sounds in a second language, evidently tried to "fit" the unfamiliar sound into their native sound system.

Second, since in the short gap durations [kp] was identified as [p], and [pk] was identified as [k], the release is shown to be more crucial than the onset in identifying the nature of the consonants in question (Figs. 2-4). This is consistent with much other research in this area (for a summary, see Pickett, Bunnell, & Revoile 1995 and references cited therein).

Besides the immediate question at hand, this tendency of the release being the key to identifying a consonant helps explain a common diachronic tendency. Labial-velars historically tend to change to simple labials, e.g. ***kp > p, *gb > b** (Cahill in prep and references therein). This can be explained on the perceptual grounds that labial-velars have labial releases and thus tend to be readily perceived as labials.[3]

Third, [kp] was mis-identified at longer gap durations than [pk], or to state it in positive terms, the English speakers in this experiment identified the [pk] cluster more readily than they did [kp] (Figs. 2-4). To what can we attribute this difference? One possibility is that the transition from the vowel into the consonant is the key factor in identifying the consonant, and that [ap] was more salient than [ak]. However, we saw above that the release is more crucial in identifying the consonant. If the release is indeed more crucial, then a reasonable conclusion is that the [p] release of [kp] was less perceptible than the [k] release of [pk].[4] This conclusion effectively falsifies the beginning hypothesis. Recall that the hypothesis, looking for an explanation of why labial-velar consonants in the world's languages are seemingly universally [k͡p] and not [p͡k], postulated that a [kp] would be more identifiable than the reverse [pk]. Our results show exactly the opposite: [pk] is more identifiable than [kp].

Some potential complicating factors exist in this experiment and must be addressed. These include possible interference from English phonotactic statistics, possible asymmetry of the recorded stimuli, and how close the experiment was to the phonetics of real-language [k͡p].

The first factor has to do with the relative frequencies of [kp] vs. [pk] in English words. If [pk] is significantly more frequent than [kp], we might expect listeners to choose the more familiar [pk] and thus skew the results in favor of that selection. However, a search of the MRC database of approximately 150,000 English words showed 3 with [kp] and 7 with [pk] between vowels. This is a difference, to be sure, but not enough to account for the patterns found in this experiment.[5]

---

[3] The same point is made in Connell (1991, 1994).

[4] However, in Figs. 4-5, [p] was <u>never</u> mis-identified at short gap intervals, while [k] was. So do Figs. 4-5 show that [p] is <u>more</u> perceptible? I do not believe so, because of the very small numbers involved in the mis-identification.

[5] The words were: [kp] - jackpot, pickpocket, stockpile, [pk] - napkin, pipkin, shopkeeper, tipcat, upcast, upcountry, upkeep. I can add "backpack" and perhaps others to the [kp] list, but the fact remains that both [VkpV] and [VpkV] are relatively rare patterns in English, and occur almost exclusively with compound words. A further check shows that this pattern of relatively rare, mostly compound, words holds for most other

Another question to consider is were the recorded segments [pa] and [ka] essentially equal in peripheral properties that could affect perceptibility, or not? For example, in the spectrograms of [apka] and [akpa] (Fig. 7), the aspiration of [ka] is longer than that of [pa]. This probably does not contribute to the higher salience of [ka] in this experiment, since both cross-linguistically and specifically in English, a [k] followed by vowel has a significantly longer VOT than a [p] or [t] (Lisker & Abramson 1964). Thus the stimuli for this experiment fit within normal ranges. Other factors, such as loudness, also do not appear relevant.

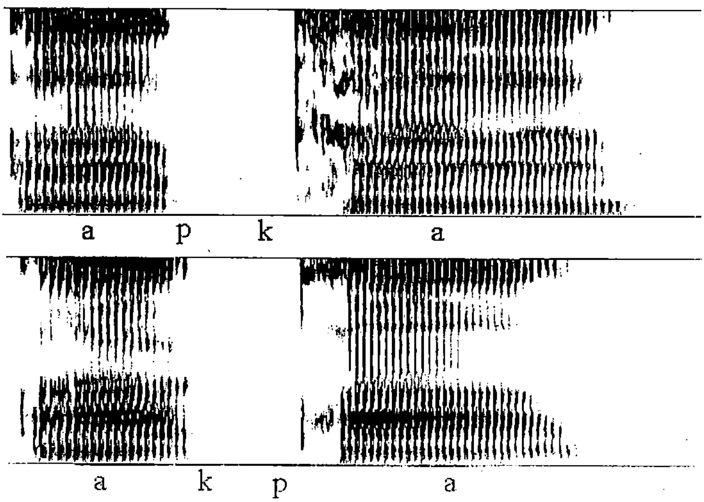

Fig. 7 Spectrograms of the [apka] and [akpa] stimuli

A further question is whether the vowel quality had any impact on the results. Would the same results be found with other vowels besides [a]?

A more serious concern is how similar the spliced [kp] in this experiment is to a real-language [k͡p]. There are at least two differences between the spliced [kp] here and at least some [k͡p] in natural languages. Quite often, literature sources note that [kp] is unaspirated, unlike simple voiceless stops (e.g. Smith 1967). The input sounds for this experiment had normal English aspiration (see Fig. 7). Also, there is often a suction mechanism for producing labial-velars (Ladefoged 1968), which was absent in this experiment. These factors could conceivably have an impact on the results.

---

stop-stop sequences, in whatever combination of voiced or voiceless, with the exception of [kt] and [pt], which are relatively more common.

CONCLUSION

In this experiment, [pk] was shown to be more perceptually salient than [kp]. If this result is a universal of human language perception, this conflicts with the universal of language production that in languages with labial-velars, [k͡p] is preferred over [p͡k]. In the introduction, we discussed three possible factors which could account for the universality of [k͡p]: first, perceptual salience, which was tested in this experiment; second, a relic of labial-velars' historical development from labialized stops (e.g. k^w); and third, greater ease of articulation for [k͡p] than [p͡k]. We have shown here that the bias of perceptual salience is actually towards [pk], not the reverse. We conclude, then, that the perceptual salience of a velar release is not a factor in the universality of [k͡p], but must be overridden by the historical or articulatory factors previously discussed. Future experiments and studies are needed to focus on these influences.

REFERENCES

Cahill, Mike. (in preparation) Labial-velars: Phonetics, History, and Phonology. ms., OSU.

Connell, Bruce. (1991). Accounting for the reflexes of labial-velar stops. In Rossi, M. (ed.) Proceedings of the XIIth ICPhS, Aix-en-Provence, Vol. 3:110-113.

Connell, Bruce. (1994). The structure of labial-velar stops. Journal of Phonetics 22:441-476.

Garnes, Sara. (1975). An Acoustic Analysis of Double Articulations in Ibibio. Proceedings of the Sixth Conference on African Linguistics, OSU Working Papers in Linguistics No. 20.

Hume, Elizabeth. (1996) A Non Linearity Based Account of Leti Metathesis. ms, OSU.

Huttar, George L., and Mary L. Huttar. (1994). Ndyuka. London: Routledge.

Hyman, Larry M. (1979). Phonology and noun structure. In Hyman, Larry M. (Ed) Aghem Grammatical Stucture. Southern California Occasional Papers in Linguistics No. 7. USC.

Ladefoged, Peter. (1968). A Phonetic Study of West African Languages. (2nd ed.) Cambridge University Press.

Ladefoged, Peter, and Ian Maddieson. (1996). The Sounds of the World's Languages. Cambridge, MA: Blackwell Publishers.

Lisker, Leigh, and Arthur S. Abramson. (1964). A cross language study of voicing in initial stops: Acoustical measurements. Word 20:384-422.

Maddieson, Ian. 1993. Investigating Ewe articulations with electromagnetic articulography. Forshungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München 31:181-214.

Mutaka, Ngessimo M., and Carl Ebobissé. (1996) The formation of labio-velars in Sawabantu: evidence for feature geometry. ms., to appear in Journal of West African Languages.

Painter, Colin. (1970). Gonja: A Phonological and Grammatical Study. Bloomington: Indiana University.

Pickett, J.M., H. Timothy Bunnell, and Sally G. Revoile. 1995. Phonetics of Intervocalic Consonant Perception: Retrospect and Prospect. Phonetica 52:1-40.

Smith, N.V. (1967). The Phonology of Nupe. Journal of African Languages 6.2: 153-169