Hearing tongue loops: Perceptual sensitivity to acoustic signatures of articulatory dynamics

Hosung Nam[a)]
Haskins Laboratories, 300 George Street, New Haven, Connecticut 06511

Christine Mooshammer
Institut für deutsche Sprache und Linguistik, Phonetik/Phonologie Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany

Khalil Iskarous
Department of Linguistics, University of Southern California, 3601 Watt Way Grace, Ford Salvatori Hall 301, Los Angeles, California 90089

D. H. Whalen[b)]
Department of Speech-Language-Hearing Sciences, City University of New York, Fifth Avenue, The Graduate Center 365, Room 7107, New York, New York 10016

(Received 11 June 2013; revised 10 September 2013; accepted 19 September 2013)

Previous work has shown that velar stops are produced with a forward movement during closure, forming a forward (anterior) loop for a VCV sequence, when the preceding vowels are back or mid. Are listeners aware of this aspect of articulatory dynamics? The current study used articulatory synthesis to examine how such kinematic patterns are reflected in the acoustics, and whether those acoustic patterns elicit different goodness ratings. In Experiment I, the size and direction of loops was modulated in articulatory synthesis. The resulting stimuli were presented to listeners for a naturalness judgment. Results show that listeners rate forward loops as more natural than backward loops, in agreement with typical productions. Acoustic analysis of the synthetic stimuli shows that forward loops exhibit shorter and shallower VC transitions than CV transitions. In Experiment II, three acoustic parameters were employed incorporating F3-F2 distance, transition slope, and transition length to systematically modulate the magnitude of VC and CV transitions. Listeners rated the naturalness in accord with those of Experiment I. This study reveals that there is sufficient information in the acoustic signature of "velar loops" to affect perceptual preference. Similarity to typical productions seemed to determine preferences, not acoustic distinctiveness.
 [http://dx.doi.org/10.1121/1.4824161]

PACS number(s): 43.70.Mn, 43.71.Es, 43.70.Aj [BRM] Pages: 3808–3817

I. INTRODUCTION

Perception is often sensitive to the dynamics of speech production (e.g., Fowler, 2005; Iskarous, 2010; Liberman *et al.*, 1967; Liberman and Whalen, 2000), but it is not always the case that what is typically produced is the preferred perceptual pattern: Recent work by Iskarous *et al.* (2010) examined vowel sequences created with articulatory synthesis that had an articulatory "pivot point" (Iskarous, 2005), or a greater or lesser degree of approximation of the tongue to the palate. The pivot pattern is that most typically produced, and it was found to be preferred over trajectories in which the tongue glides upward along the palate. However, listeners preferred more extreme departures from the palate even more. Thus the most commonly encountered pattern was not the most preferred one perceptually.

One frequently found, but still only partially explained articulatory pattern is the "loop" formed by the tongue when making a velar closure for a stop or nasal following low and/or back vowels. First described by Houde (1968), this pattern is that of a forward motion into and during the closure and a backward motion out of it depending on the vowel context. This pattern could be attributed to a coarticulatory adjustment, for which forward sliding during closure assists target achievement for the following vowel. However, since no backward loops were found following front vowels this idea was rejected (see Houde, 1968). As an alternative, Ohala (1983) suggested that the forward sliding during the velar closure is a cavity enlargement strategy in order to sustain voicing during /g/. However, Mooshammer *et al.* (1995) found more pronounced forward loops for /k/ than for /g/. Another early explanation posited the build-up of air pressure behind the closure as the cause (Coker, 1976; Houde, 1968; Kent and Moll, 1972). The extent of the forward movement during closure could therefore be modulated by pressure built up behind the closure. However, the existence of loops with velar nasals made the aerodynamic explanation only partially tenable (Mooshammer *et al.*, 1995). Modeling suggests that the organization of muscles may explain most of the motion (Perrier *et al.*, 2003). Iskarous (2005) points out that flesh points on the tongue are not likely to be

[a)]Author to whom correspondence should be addressed. Electronic mail: nam@haskins.yale.edu
[b)]Also at: Haskins Laboratories, 300 George Street, New Haven, Connecticut 06511.

controlled directly, so a higher level explanation is called for; he suggests that motion through a pivot point results in a stable articulation that has complex consequences for (non-controlled) flesh points. Up to now no conclusive explanation has been provided but despite this fact, the movement itself has been consistently found in a large variety of studies with different methods and for many languages: e.g., Australian languages: Butcher and Tabain (2004) (Electropalatography); Catalan: Recasens and Espinosa (2010) [Electromagnetometry (EMA)]; English: Houde (1968) (X-ray), Perkell (1969) (X-ray), Lofqvist and Gracco (2002) (EMA); German: Mooshammer *et al.* (1995) (EMA); Hungarian: Geng (2009) (EMA); Korean: Brunner *et al.* (2011) (EMA).

In the present study, we used both articulatory and formant synthesis to determine whether articulatory loops affect perception. Because loops are consistently found, it is not feasible to manipulate natural tokens to avoid them. With articulatory synthesis, on the other hand, we can force the (two-dimensional representation of the) tongue to take a linear path or, indeed, a backward-to-forward loop. With a velar closure, all the resulting consonants should sound like a /g/, but we will look for preferences within those categories. Then, we will be able to test specific aspects of those acoustic realizations for a chance to see how much each acoustic aspect contributes to the perception. We hypothesized that listeners would be sensitive to the acoustic signatures of the movement trajectories rather than focusing only on target locations.

II. EXPERIMENT I: (ARTICULATORY SYNTHESIS)

A. Method

1. Stimuli

In this study, we used Haskins laboratories' configurable articulatory synthesizer (CASY; Rubin *et al.*, 1996) to generate /aga/ sequences with various articulatory trajectories. CASY is based on Mermelstein's (1973) articulatory model and allows one to parametrically control major articulators (jaw, tongue body, tongue tip, lips, etc.). These points are defined by anatomy and geometry using crucial articulatory points and the model articulator variables [see Nam *et al.* (2013) for details]. The position of an articulator point is defined as a polar coordinate relative to another articulator point. For example, tongue body center (TBC) is defined by the two articulator variables, CL and CA, which are length and angle, respectively, with respect to the mandibular condyle. Tongue tip position (T) is defined by TL and TA with respect to tongue blade. In the model, TBC can be employed to model velar consonants and vowels.

To simulate /aga/ sequences with various tongue body trajectories for a perception experiment, we first identified Cartesian coordinates of TBC for /g/ and /a/. As shown in Fig. 1(a), seven different trajectories of the tongue body connecting these two points (/a/ and /g/) were generated by modulating the direction and the size of loop created by /aga/ sequence. As shown in Fig. 1(b), quadratic Bézier curves using three points were used to modulate the width of loop and the direction. Two fixed points from /a/ and /g/ were used for the ends of curves. The other point was used to determine the shape of a curve (either upward or downward trajectory of a loop) in terms of symmetry and maximal curvature. For example, if the point is orthogonal to the connecting line of the two fixed points (/a/ and /g/) at the center of the line, the curve will be symmetrical between above and below the point. Otherwise, it will be asymmetrical. In addition, if the point is farther from the connecting line, the curve will be curvier. As shown in Fig. 1(b), the points were set on both right and left sides at 5 mm (loop 1 and 7), 3.33 mm (loop 2 and 6), 1.67 mm (loop 3 and 5), and 0 mm (loop 4) from the connecting line, resulting in three sizes of forward (back-to-front) loop, 1 linear loop and 3 sizes of backward (front-to-back) loops. Note that we set all the points orthogonal to 8 mm up from the center of the two fixed points to ensure smooth sliding movement during the production of /g/.

For each loop, 62 points were created equidistantly and each point synthesized an acoustic signal with a duration of 1/F0 (F0 = 150 Hz, sampling rate = 10 000 Hz), resulting in acoustic signals of 413 ms in duration. The vocalic segment, V1, varied in duration from 167–184 ms, and V2 complementarily varied from 180–163 ms. The exact ratio depended on how quickly the loop type attained closure. The closure itself always lasted 66 ms.

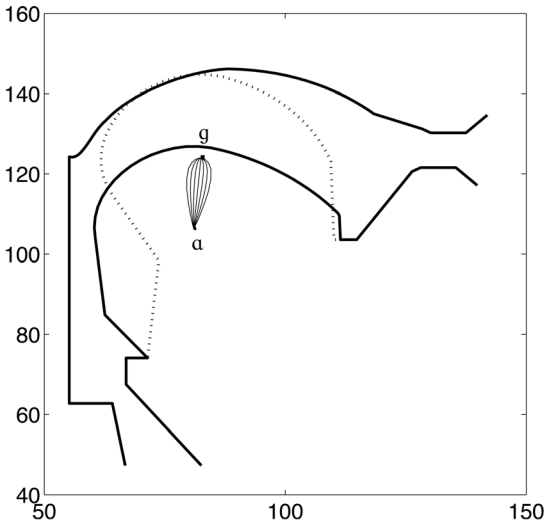
2. Participants

Thirteen native speakers of American English (6 males and 7 females) participated in the perception experiments (discrimination and goodness rating) after providing informed consent. They had no reported history of speech or hearing problems.

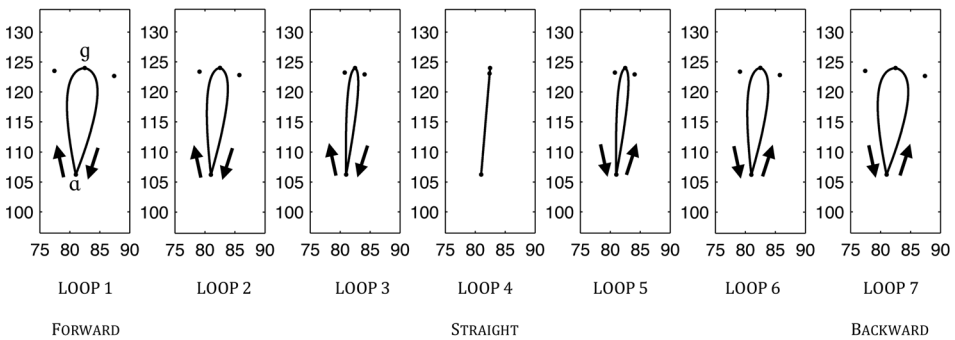
3. Procedure

A discrimination task involving all the stimuli was followed by a binary choice goodness decision task. All instructions were provided in written form. For the discrimination task, 5 two-step pairs (1–3, 2–4, 3–5, 4–6, 5–7), 4 three-step pairs (1–4, 2–5, 3–6, 4–7), 3 four-step pairs (1–5, 2–6, 3–7), 2 five-step pairs (1–6, 2–7), and 1 six-step pair (1–7) of triads were selected from the 7 different loops in the scale (1 to 7). Each pair was used for four types of AXB sequences: AAB, BBA, BAA, and ABB; hence, 15 pairs produced 60 stimuli in total. The 60 stimuli were randomized to form one block. Each participant was given 8 blocks, with a rest after every 2 blocks, for a total of 480 stimuli or 32 responses per pair. They were instructed to answer whether X is identical to A or B in AXB sequence by pressing one of two keys on a computer keyboard.

After the discrimination task was completed, the participants judged stimuli for goodness. 13 blocks of randomized presentations of the 7 loops were presented, for 91 stimuli in total. At this point, the listeners had heard the sounds many times through the discrimination task. They were told to evaluate whether the token of "aga" they heard was a good rendition or not; this was a binary forced choice. (See the full instructions in Appendix A.)



(a) 7 different trajectories of tongue body for /agɑ/ generated in CASY by using Bézier curves and varying the width of loop and the direction.



(b) 3 sizes of forward / 1 linear loop /3 sizes of backward loops

FIG. 1. Tongue body trajectories for /aga/. The pharynx is at the lower left, and the lips at the upper right.

B. Results

Figure 2 shows the results from the discrimination task aligned with the goodness ratings for the articulatorily synthesized stimuli. The horizontal axis is the index for loops as illustrated in Fig. 1. The lower the index is, the more forward the loop is. The discrimination results are averaged across listeners and connected in dashed lines for a given step. Note that the number of steps is the distance between a pair of tokens compared. For each step width, the result data is plotted at the center of the indices of the two compared tokens. For example, AXB sequences for step width 2 compare 1–3, 2–4, 3–5, 4–6, and 5–7 and listeners' discrimination accuracy percentages are plotted at 2, 3, 4, 5, and 6, which are the pairs' mid values. The discrimination from all step widths is above 60%, which is above chance, although listeners better discriminate tokens with more steps in between. The accuracy percentages are above 90% for step width 4, 5, and 6 while they are below 90% for step width 2 and 3. In particular, for step width 2 and 3, discrimination is observed to slightly increase around the straight line stimulus (index 4) which also shows a change in goodness ratings.

Results for the goodness rating are presented in means and standard deviations for each loop. The rating is

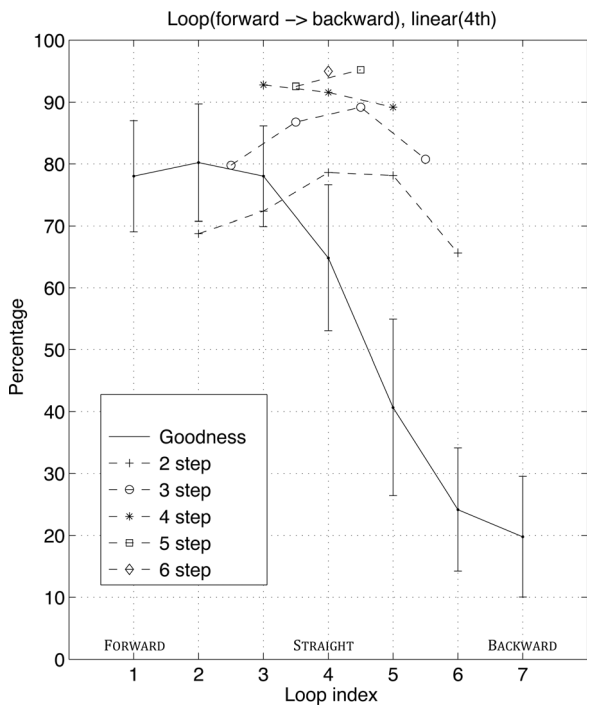


FIG. 2. Results of perception Experiment I as a function of loop index (see Fig. 1): Goodness rating (solid), discrimination accuracy (dashed).

TABLE I. Pairwise comparisons of goodness ratings using Tukey's *t*-test for Experiment I (*$p < 0.05$).

Loop 1	Loop 2	Loop 3	Loop 4	Loop 5	Loop 6	Loop 7
			*			

approximately 80% for the forward loops (loop index 1, 2, and 3), drops to 64% for the linear loop (loop index 4) and further drops between 40% and 20% for the backward loops (loop index 5, 6, and 7). We performed a one way repeated measures analysis of variance (ANOVA) with loop index as the independent variable to examine if the mean ratings differ within subjects [$F(6, 72) = 14.82$, $p < 0.0001$]. It showed that at least one pair of loops significantly differ from each other. A pairwise *t*-test (using Tukey's multiple comparison test adjusted by Bonferroni method) was then run to examine where the difference occurs. As in Table I, the mean ratings of loop 6 and 7 significantly differ from those of loop 1, 2, 3, and 4. This implies that listeners rate the forward loops (1, 2, 3) and the linear loop (4) significantly more natural than the backward loops (6, 7).

C. Acoustic analysis

In order to explore what acoustic aspects correlate with the type of loop and the goodness judgment results we performed acoustic analysis of the articulatorily modeled (CASY) stimuli. To quantify the formant change during the vowel transition during the velar stop, we measured F2 values (Hz) at the transition beginnings (for CV) and endings (for VC) and the transition durations. F3 was static with no substantial vowel transition (1650 Hz). Note that higher F2 indicates larger velar pinch, i.e., F2 and F3 coming together (Lamel, 1988; Olive *et al.*, 1993). Figure 3 presents measured F2 [3(a)] and duration [3(b)] as a function of the trajectory shape. Forward loops (steps 1–3) exhibit less acoustic distinctiveness in the VC transition (lower F2 and shorter duration) and more in the CV transition (higher F2 and longer duration). This corresponds to a typical production pattern that greater salience is found in CV transitions than VC transitions (Ohala and Kawasaki, 1984). On the other hand, backward loops (steps 5–7) exhibit more acoustic distinctiveness in the VC transition (higher F2 and longer duration) and less in the CV transition (lower F2 and shorter duration). The spectrograms for the two extreme cases [loop 1 (forward) and loop 7 (backward)] are compared in Fig. 4.

D. Pivot analysis

Our stimuli consisted of two transitions between speech goals, /ag/ and /ga/, and they were thus amenable to analysis for an articulatory pivot (Iskarous, 2005) in the movements of the (synthetic) articulators. These are shown in Fig. 5, for /ag/ (panel a) and /ga/ (panel b). The point at which the beginning configuration (red line) and the ending (blue line) cross is where we would expect to see a pivot. The degree to which all the intervening configurations cross at the same point defines whether a pivot exists: If all of them coincide, then there is a pivot, while a substantial number of non-agreement indicates a lack of a pivot. In the forward loops (1–3), there is no pivot in the /ag/ portion (panel a) but pivots in the /ga/ portion (panel b). The reverse is true for the backward loops (5–7). Changes in degree of pivoting were tested in an earlier study [Iskarous *et al.* (2010)] with a vowel sequence /ai/. There, it was found that pivoted trajectories received positive goodness ratings, but non-pivoted trajectories that avoided the palate were even more highly rated. Here, higher ratings were elicited by the opening gestures that had pivots; the pivots in the closing gestures were not enough to offset the lack of pivots in the opening gestures.

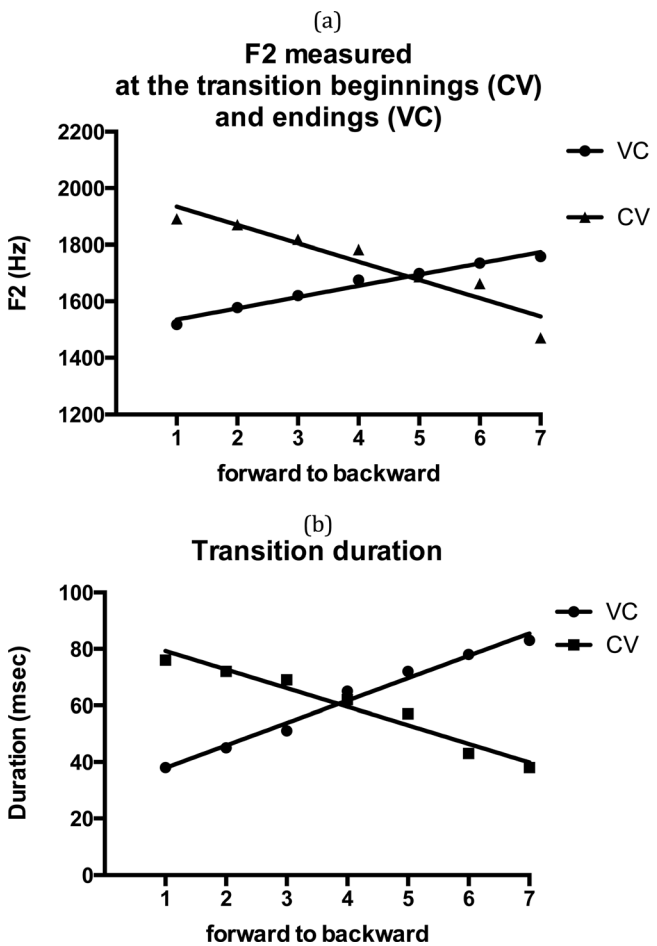


FIG. 3. (a) Measured F2 values (Hz) at the transition beginnings (for CV) and endings (for VC) and (b) the transition durations as a function of the level of forwardness for the articulatorily modulated stimuli.

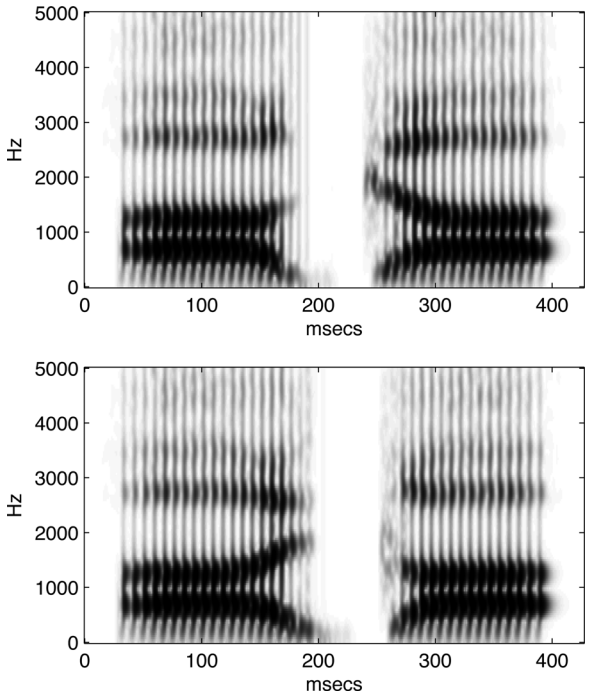


FIG. 4. Spectrogram of /aga/: loop 1 (forward) (top) and loop 7 (backward) (bottom).

E. Discussion

The results show that listeners are sensitive to the acoustic consequences of articulatory "loops" even though all items were accepted, and rated, as the intended /aga/ sequence and therefore belong to the same phonological category. This is similar to the findings of Iskarous *et al.* (2010) for vowel sequences (/ai/, in their case). It is interesting that there is still a similar amount of acoustic differentiation in loops of similar magnitude (say, numbers 1 and 7), yet their acceptability differs greatly (Fig. 2).

The particular acoustic patterns can be manipulated independently of precise articulatory relations. It is often the case that manipulations that are straightforward in the acoustic domain have an indeterminate relationship to any potential articulation (Atal *et al.*, 1978). Nonetheless, it is of interest to see whether such acoustically motivated patterns can also affect perception. Therefore, we performed an additional perception experiment using stimuli generated by systematically modulating acoustic parameters for the VC and CV transitions of the velar stop. In the first experiment, pivots in CV correlated with the higher goodness ratings, while pivots in VC did not. Because earlier results indicated that pivoted gestures would be expected to be preferred (Iskarous *et al.*, 2010), this leads us to expect that CV is more responsible for the goodness judgments than the VC. This prediction was tested from the acoustic point of view in Experiment 2.

III. EXPERIMENT II (ACOUSTIC SYNTHESIS)

A. Method

1. Stimuli

In this experiment, we systematically modulated the strengths of the vowel transition for the velar stop in the acoustic domain. Three parameters were chosen for the acoustic modulation: transition slope, transition duration, and F2-F1 distance. We employed HLsyn (Hanson and Stevens, 2002) to generate /aga/ stimuli with the modulations. For the static region of the vowel /a/, F1, F2, F3, and F4 are set to 645, 1180, 2670, and 3300 Hz. Three parameters were considered to modulate the magnitude of VC and CV transition for velar consonants, i.e., a velar pinch: F3 to F2 *distance,* transition *slope,* transition *duration*. However, any of these parameters could not be modulated independently but one is always varied along with another. We thus grouped each two into a combination parameter, resulting in three combination parameters: "DistSlope" (distance-slope), "SlopeDur" (slope-duration), and "DurDist" (duration-distance) as illustrated in Fig. 6. For "DistSlope," we fixed the transition duration for F3 and F2 to 100 ms and varied the F3-F2 distance [400–300–200 Hz] and the transition slope at the same time. For "SlopeDur," we fixed the F3-F2 distance to 300 Hz and varied the transition slope and the duration [70–100–130 ms] at the same time. For "DurDist," we fixed the transition slope and varied both the duration [70–100–130 ms] and the distance [400–300–200 Hz]. Each combination parameter was varied in three levels: strong, medium, and weak in VC and CV; hence, each combination parameter creates nine stimuli (VC: strong/med/weak × CV: strong/med/weak) and 27 stimuli in total.

2. Participants

Eleven native speakers of American English (6 males and 5 females) participated in this experiment. Informed consent was obtained before the experiment. Participants had no reported history of speech or hearing problems. None of the participants from the Experiment I participated in this experiment.

3. Procedure

All instructions were provided in written form. For the goodness rating task, the 27 acoustically manipulated stimuli were randomized within each block. Each participant was given 32 blocks with a rest after every four blocks and therefore heard each stimulus 32 times. They were instructed to report whether or not the token they hear is a good version of the final VCV in "Lady Gaga" or not (see the full instructions in the Appendixes). "Yes" responses were coded as 1, "No," as 0.

B. Results

Figure 7 shows the results from the goodness rating tasks for the acoustically manipulated stimuli. For each combination parameter, the x axis is CV transition magnitude in three levels (1: weak, 2: mid, 3: strong), the y axis is VC transition magnitude, and the z axis is goodness rating. Each combination parameter shows nine rating scores (black circles) averaged across listeners and the standard deviations (vertical lines). The nine points are superimposed on a meshed grid connecting one another smoothly. The color is proportional to surface height highlighting the highest (red)

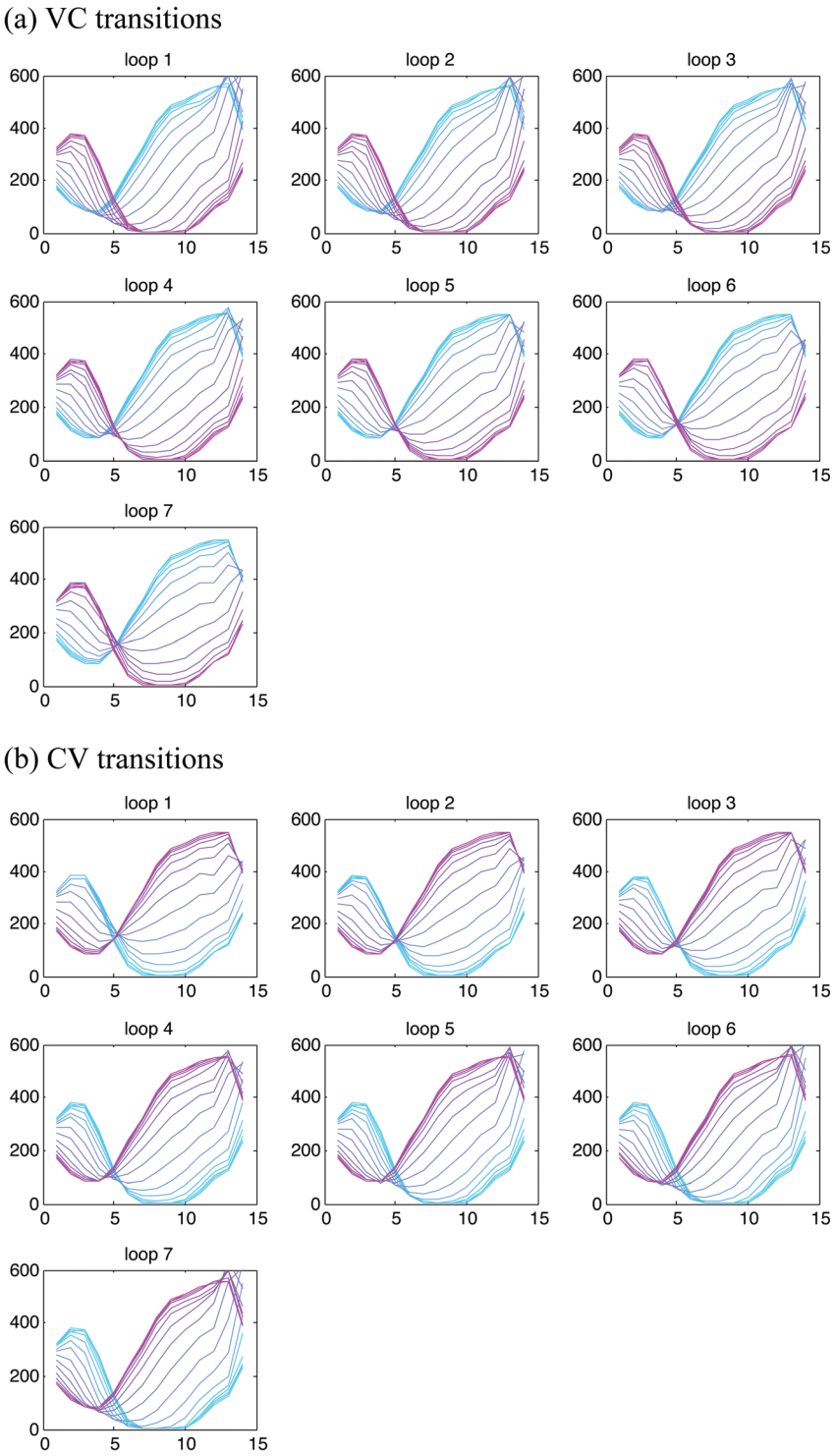


FIG. 5. Time-varying vocal tract functions for (a) VC transitions and (b) CV transitions. Time ranges from beginning (purple line) to end (blue line). X-axis shows location in the vocal tract (1 = pharynx, 14 = lips). Y-axis shows area at each point in the vocal tract (arbitrary units; higher is more open). Pivots can be seen at location 5 on the x axis when all lines converge (loops 5–7 for VC and 1–3 for CV); lack of pivoting is seen in the lack of convergences there (loops 1–3 for VC and 5–7 for CV).

and lowest points (blue). For all parameters (DistSlope, SlopeDur, DurDist), listeners rated stronger magnitude in CV transition, and smaller magnitude in VC transition more natural. For each parameter, the best rating (red in the meshed grid) is observed when VC transition is weak and CV transition is strong and the worst rating (blue in the meshed grid) is observed when VC transition is strong and CV transition is weak. This result corresponds to the acoustic analysis of articulatorily manipulated stimuli (Sec. II C). That is, the forward loop exhibited a strong transition in onset but a weak transition in coda whereas a weak onset transition and a strong coda transition were found for the backward loop.

We first performed a repeated measures three-way ANOVA with the factors VC transition magnitude ("weak," "mid," "strong"), CV transition magnitude ("weak," "mid," "strong"), and modulation parameter ("DistSlope," "SlopeDur," "DurDist") to show whether listeners' goodness ratings differ as either of the factors is varied. There were significant effects for all three factors [VC transition magnitude: $F(2, 208) = 23.37$, $p < 0.001$, means of 0.55, 0.51, 0.34 (w/m/s); CV transition magnitude: $F(2, 208) = 32.16$,

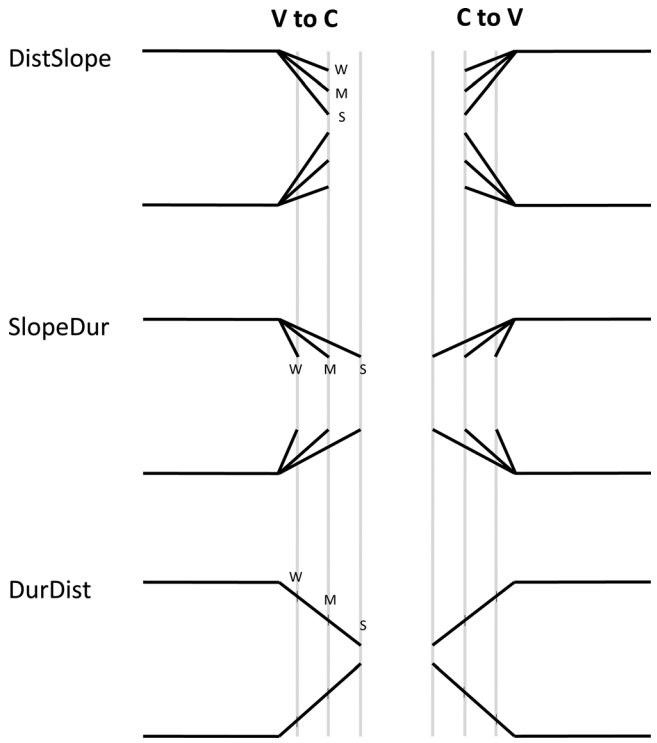


FIG. 6. Illustration of three types of acoustic manipulation for the vowel transition of a velar stop in /aga/. Three levels (strong, mid, weak) of the transition are modeled for each parameter.

$p < 0.001$, means of 0.31, 0.50, 0.58 (w/m/s); modulation parameter: $F(2, 208) = 7.39$, $p < 0.01$, means of 0.50, 0.46, 0.44 (DistSlope / SlopeDur / DurDist)]. For each modulation parameter, a pairwise *t*-test (using Tukey's multiple comparison test corrected by Bonferroni method) was then done for each of the factors (VC and CV transition) in order to identify the level of the difference. Table II shows that for all three parameters, VC-weak and VC-mid elicited significantly better ratings than VC-strong, and CV-strong and CV-mid elicited better ratings than CV-weak. Adjusted *p*-values are provided for the pairs showing significant differences.

C. Discussion

Acoustic manipulation of the strength of cues into and out of the silent portion signaling a stop closure resulted in perceptual patterns very similar to the articulatory synthesis of the first experiment. Whether the change was in the transitions' duration, frequency or both, strong CV and/or weak VC stimuli were preferred. Although it is possible to conceive of an experiment in which VC and CV would come from different parameters, the number of stimuli that would result was too large to justify testing them here. Overall, when CV transitions were strong or mid and VC transitions were weak or mid, ratings were fairly uniformly high, ranging from 0.54 to 0.70. Other combinations grew progressively worse, until the uniformly disfavored combination of weak CV and strong VC transitions. The pseudo-spectrograms of the best and worst cases are shown in Fig. 8. Because all transitions independently supported a velar stop percept, it appears that these preferences were based on the success of individual productions rather than on the resolution of mismatches (as in Repp, 1978).

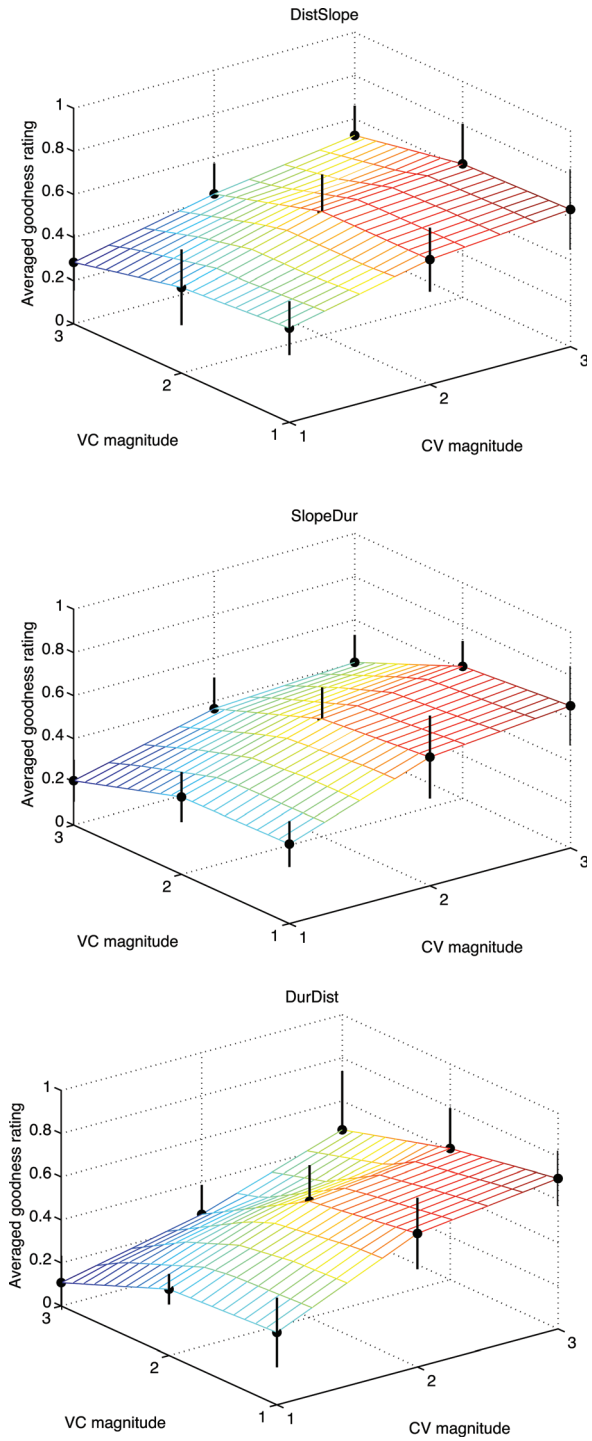


FIG. 7. Results of Experiment II with acoustically modulated stimuli for all combination parameters: "DistSlope," "SlopeDur," and "DurDist."

IV. GENERAL DISCUSSION

Perception of synthesized velar stop closures, whether generated via articulatory or acoustic synthesis, showed that listeners prefer the result that reflects a path of tongue motion that is consistently present in typical productions and yet not apparently of direct usefulness. In physiological terms, a forward "loop" of the tongue is preferred to a

TABLE II. Pairwise comparisons of goodness ratings using Tukey's *t*-test for Experiment II (a) VC transition and (b) CV transition.

VC transition	Weak	Mid	Strong
DistSlope	p = 0.006	p = 0.024	
SlopeDur	p = 5.7e-06	p = 1.6e-04	
DurDist	p = 3.9e-06	p = 4.8e-04	

CV transition	Weak	Mid	Strong
DistSlope		p = 8.4e-04	p = 1.4e-06
SlopeDur		p = 3.4e-05	p = 1.0e-07
DurDist		p = 8.2e-05	p = 2.3e-09

backward one. In acoustic terms, strong CV transitions and weak VC transitions are preferred. It is not surprising, perhaps, that the most common pattern is preferred [though this was not the case for Iskarous *et al.* (2010)], but it is surprising that this preference could be found in stimuli that all achieved their intended phonetic target ([aga]). Listeners appear to be extremely sensitive to the dynamical pattern of their language, picking up coarticulatory information as soon as it is available (e.g., Beddor *et al.*, 2013).

All of the stimuli were synthesized, so, in some sense, none of them had an articulatory pattern underlying them. The principles of articulatory synthesis, however, result in fairly close matches between the modeled articulation and natural productions of similar sequences. To the extent that we succeeded in producing acoustically realistic results, the results of the first experiment indicate that articulation, even this relatively minor component, is important to perception. The acoustic manipulations of the second experiment show that listeners are willing to give a phonetic interpretation to patterns that may not match any particular articulation exactly, as with the current backward loops for the /a/ context. This has been known since the early days of speech synthesis (e.g., Cooper *et al.*, 1952), in which relatively straight-line formant transitions were successful in eliciting stop judgments even though they were, like ours, physiologically somewhat unrealistic (Liberman and Whalen, 2000). While the flexibility of the speech perception system is widely documented (e.g., Green *et al.*, 1991; Jenkins *et al.*, 1994; Remez *et al.*, 1994), most responses in such experiments have been categorized only at the level of distinctive sounds (phonemes). The present results, like some others (e.g., Volaitis and Miller, 1992), indicate that listeners can judge some of the recovered representations as being better realizations than others. This suggests that there is much more to be learned about how the acoustic manipulations are treated in speech perception.

The articulatory loops are also well-documented, but the movement is not immediately suggestive of an acoustic output. With the results in hand, we can postulate that the greater time spent close to the palate in the forward part of the loop will generate shallower transitions than the more rapid departure from the palate in the release (see Fig. 1). While this may be straightforward, it does not make a straightforward perceptual prediction because the shallower transitions might result in reducing the information in VC. One would think that strengthening cues would, in general, strengthen the percept. That is not the case in our stimuli, where strong VC transitions received the worst perceptual ratings. There are three possible explanations. First, listeners might prefer patterns that conform to the typical productions, where CV transitions are more saliently produced than VC ones (Ohala and Kawasaki, 1984). Alternatively, listeners might prefer patterns that conform to an asymmetry existing in the human perception mechanism. Results from the experiments using cross-spliced stops have shown that CV transitions are given more perceptual weight than VC ones (Repp, 1978). When the place information in VC and CV transitions is contradictory, CV transitions dominate the VC ones. For example, listeners hear /aba/ when VC from /aga/ and CV from /aba/ are spliced. However, this asymmetry is observed in synthetic stimuli where VC and CV transitions are symmetrical in magnitude (Dorman *et al.*, 1975) and in natural speech, where CV is expected to be more salient than

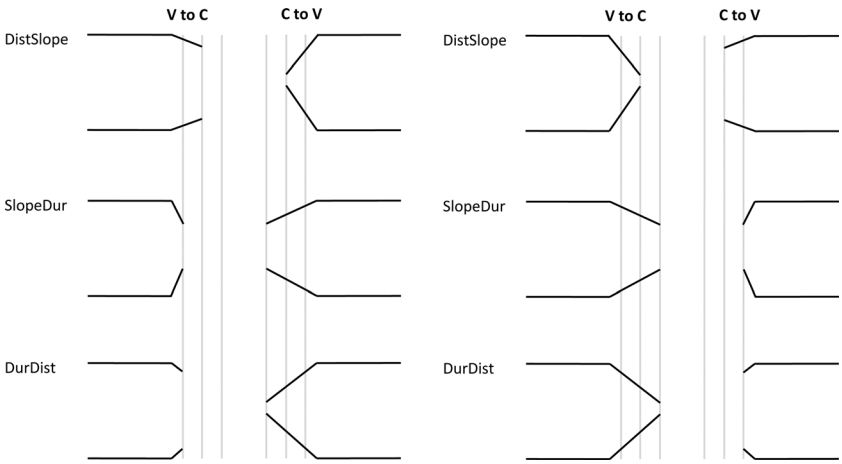


FIG. 8. Combinations showing the highest (left) and worst ratings for each modulation parameter.

VC, for both playback directions (Fujimura *et al.*, 1978). Hence, CV's perceptual dominance in these studies cannot be attributed to the acoustic salience typically observed in CV transitions. It can be postulated that listeners hear CV transitions with greater weight regardless of the actual production patterns, i.e., the size/amount of the acoustic cues. Unlike this account, Steriade's perceptual-map is based on the actual production patterns (Steriade, 2001). For example, the CV dominance in perception does not hold for retroflex consonants. The phones [t] and [ʈ] are more distinguishable in VC than in CV because for [ʈ] because the tongue tip slides forward during the closure and releases at a similar site to [t], which makes CV transitions less distinguishable than VC transitions (Steriade, 2001). Finally, it might be that strong VC transitions more likely introduce a bit of additional vowel-like sound whereas strong CV transitions are only used to cue /g/.

The loop pattern of the velar consonants does not seem to be founded entirely on aerodynamics (Mooshammer *et al.*, 1995), even though the extent of the loops probably does (Hoole *et al.*, 1998). This characteristic forward trajectory may stem from physiological constraints (Perrier *et al.*, 2003). However, the preference for weak VC and strong CV transitions is also found in alveolar stops, in which a smaller articulatory loop is observed (Hoole *et al.*, 1998). This raises the possibility that the perceptual preference is the cause of velar loops, not the other way around. Perhaps the only way to produce the asymmetrical pattern in VC vs CV transitions in velars is with loops, while there are other articulatory strategies available to the other places of articulation. This bears further investigation, both in synthesis and articulatory measurements.

ACKNOWLEDGMENTS

This material is based upon work supported by NIH Grant No. DC-002717 to the Haskins Laboratories and NSF Grant No. 1246750 to the University of Southern California. We thank Will Grathwohl and Shabnam Elahi for help with the experiments.

APPENDIX A: INSTRUCTIONS FOR EXPERIMENT I

Thank you for participating in our perception experiment. If at any point you are uncomfortable or have any question, please tell us. And at any time during the experiment you can stop. There are two types of tasks in this experiment: Task 1 consists of four sessions and Task 2 includes one session. Estimated time for each session is 10 min (5 min for Task 2). Note all the sounds you will hear are slightly deviant from one single targeted V-C-V sequence, where the two vowels surrounding the consonant are supposedly identical. Task 1: You will press "Start" and hear three sounds. Your task is to determine if the second one sounds more like the first one or the third one. That is, are the first two sounds more similar or are the second two sounds more similar. If the second sounds like the first, please press the left button. If it sounds like the third one, please press the right button. For instance, if you hear "A A B," you would press the left button. Whereas if you hear "A B B," you would press the right button. You will not hear "A B A." Please wait until you hear all three sounds before you make a decision. Once you press the left or right button, you will hear the next set of sounds. The sounds are very similar to each other. Please do your best, despite the difficulty of the task. If you feel like you did not make the correct decision on one set of sounds, just do your best on the next set. And it is natural to find the degree of difficulty varying. Please continue being attentive to the task, even if you feel like you are not performing well. In order to familiarize you with the sounds you will hear, we will do 10 familiarization trials, before the experiment begins. We will be there to answer questions you may have about the task between the familiarization period and the experiment. This task will be repeated four times. You can take a break between these sessions, if you are tired. At the end of each session, you will see "Start" button again and press it to move on to the next session. At the end of this task (after four sessions), when you will see "RATE" and "Start" button on the screen, please alert the experimenter that you are done. Task 2: At this point you will have heard the sounds in the experiment many times. In the second task, you will be presented with a single sound, once you press "Start." If it sounds natural, please press the left button. If not, and deviant of the target sound, please press the right button.

APPENDIX B: INSTRUCTIONS FOR EXPERIMENT II

Lady Gaga is a famous popular singer. In this experiment, you will hear synthetic sounds that attempt to match the last part of her name, what is after the first G in "Gaga." We ask you to indicate after you hear each sound whether this is a good rendition of the last part of her name or not. If it is a good rendition, please hit the left key on the keyboard. If it is not, then please hit the right key.

Atal, B. S., Chang, J. J., Mathews, M. V., and Tukey, J. W. (**1978**). "Inversion of articulatory-to-acoustic transformation in the vocal tract by a computer-sorting technique," J. Acoust. Soc. Am. **65**, 1535–1555.

Beddor, P. S., McGowan, K. B., Boland, J. E., Coetzee, A. W., and Brasher, A. (**2013**). "The time course of perception of coarticulation," J. Acoust. Soc. Am. **133**, 2350–2366.

Brunner, J., Fuchs, S., and Perrier, P. (**2011**). "Supralaryngeal control in Korean velar stops," J. Phonetics **39**(2), 178–195.

Butcher, A., and Tabain, M. (**2004**). "On the back of the tongue: Dorsal sounds in Australian languages," Phonetica **61**, 22–52.

Coker, C. H. (**1976**). "A model of articulatory dynamics and control," Proc. IEEE **64**, 452–460.

Cooper, F. S., Delattre, P. C., Liberman, A. M., Borst, J. M., and Gerstman, L. J. (**1952**). "Some experiments on the perception of synthetic speech sounds," J. Acoust. Soc. Am. **24**, 597–606.

Dorman, M. F., Raphael, L. J., Liberman, A. M., and Repp, B. (**1975**). "Some maskinglike phenomena in speech perception," Haskins Laboratories Status Report on Speech Research, SR-42/43, pp. 265–276.

Fowler, C. A. (**2005**). "Parsing coarticulated speech in perception: effects of coarticulation resistance," J. Phonetics **33**, 199–213.

Fujimura, O., Macchi, M. J., and Streeter, L. A. (**1978**). "Perception of stop consonants with conflicting transitional cues: A cross-linguistic study," Lang. Speech **21**(4), 337–346.

Geng, C. (**2009**). *A Cross-Linguistic Study on the Phonetics of Dorsal Obstruents: Experiment Investigations* (Südwestdeutscher Verlag, Saarbrücken, Germany), p. 292.

Green, K. P., Kuhl, P. K., Meltzoff, A. N., and Stevens, E. B. (**1991**). "Integrating speech information across talkers, gender, and sensory modality: Female faces and male voices in the McGurk effect," Percept. Psychophys. **50**, 524–536.

Hanson, H. M., and Stevens, K. N. (**2002**). "A quasiarticulatory approach to controlling acoustic source parameters in a Klatt-type formant synthesizer using HLsyn," J. Acoust. Soc. Am. **112**, 1158–1182.
Hoole, P., Munhall, K. G., and Mooshammer, C. (**1998**). "Do airstream mechanisms influence tongue movement paths?," Phonetica **55**, 131–146.
Houde, R. (**1968**). *A Study of Tongue Body Motion during Selected Speech Sounds* (Speech Communications Research Laboratory, Santa Barbara, CA), p. 167.
Iskarous, K. (**2005**). "Patterns of tongue movement," J. Phonetics **33**, 363–381.
Iskarous, K. (**2010**). "Vowel constrictions are recoverable from formants," J. Phonetics **38**, 375–387.
Iskarous, K., Nam, H., and Whalen, D. H. (**2010**). "Perception of articulatory dynamics from acoustic signatures," J. Acoust. Soc. Am. **127**, 3717–3728.
Jenkins, J. J., Strange, W., and Miranda, S. (**1994**). "Vowel identification in mixed-speaker silent-center syllables," J. Acoust. Soc. Am. **95**, 1030–1043.
Kent, R. D., and Moll, K. L. (**1972**). "Cinefluorographic analyses of selected lingual consonants," J. Speech Hear. Res. **15**, 453–473.
Lamel, L. F. (**1988**). "Formalizing knowledge used in spectrogram reading: Acoustic and perceptual evidence from stops," Ph.D. thesis, Department of Electrical and Computer Engineering, Massachusetts Institute of Technology, Cambridge, May 1988, p. 378.
Liberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. (**1967**). "Perception of the speech code," Psychol. Rev. **74**, 431–461.
Liberman, A. M., and Whalen, D. H. (**2000**). "On the relation of speech to language," Trends Cogn. Sci. **4**, 187–196.
Lofqvist, A., and Gracco, V. L. (**2002**). "Control of oral closure in lingual stop consonant production," J. Acoust. Soc. Am. **111**, 2811–2827.
Mermelstein, P. (**1973**). "Articulatory model for the study of speech production." J. Acoust. Soc. Am. **53**, 1070–1082.
Mooshammer, C. M., Hoole, P., and Kühnert, B. (**1995**). "On loops," J. Phonetics **23**, 3–21.
Nam, H., Giulivi, S., Goldstein, L. M., Levitt, A. G., and Whalen, D. H. (**2013**). "Computational simulation of CV combination preferences in babbling," J. Phonetics **41**, 63–77.
Ohala, J. J. (**1983**). "The origin of sound patterns in vocal tract constraints," in *The Production of Speech*, edited by P. F. MacNeilage (Springer-Verlag, New York), pp. 189–216.
Ohala, J. J., and Kawasaki, H. (**1984**). "Prosodic phonology and phonetics," Phonology **1**, 113–127.
Olive, J. P., Greenwood, A., and Coleman, J. (**1993**). *Acoustics of American English Speech: A Dynamic Approach* (Springer Verlag, New York), p. 396.
Perkell, J. S. (**1969**). *Physiology of Speech Production: Results and Implications of a Quantitative Cineradiographic Study* (MIT Press, Cambridge, MA), p. 104.
Perrier, P., Payan, Y., Zandipour, M., and Perkell, J. (**2003**). "Influences of tongue biomechanics on speech movements during the production of velar stop consonants: A modeling study," J. Acoust. Soc. Am. **114**, 1582–1599.
Recasens, D., and Espinosa, A. (**2010**). "Lingual kinematics and coarticulation for alveolopalatal and velar consonants in Catalan," J. Acoust. Soc. Am. **127**, 3154–3165.
Remez, R. E., Rubin, P. E., Berns, S. M., Pardo, J. S., and Lang, J. M. (**1994**). "On the perceptual organization of speech," Psychol. Rev. **101**, 129–156.
Repp, B. H. (**1978**). "Perceptual integration and differentiation of spectral cues for intervocalic stop consonants," Percept. Psychophys. **24**, 471–485.
Rubin, P. E., Saltzman, E., Goldstein, L. M., McGowan, R. S., Tiede, M. K., and Browman, C. P. (**1996**). "CASY and extensions to the task-dynamic model," in *Proceedings of the 1st ESCA ETRW on Speech Production Modeling and 4th Speech Production Seminar*, pp. 125–128.
Steriade, D. (**2001**). "Directional asymmetries in place assimilation: A perceptual account," in *The Role of Speech Perception in Phonology*, edited by E. Hume and K. Johnson (Academic, San Diego, CA), pp. 219–250.
Volaitis, L. E., and Miller, J. L. (**1992**). "Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories," J. Acoust. Soc. Am. **92**, 723–735.