Cues to perception of reduced flaps

Natasha Warner

(Dept. of Linguistics, Univ. of Arizona, Tucson, and

Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands)

Amy Fountain

(Dept.of Linguistics, Univ. of Arizona, Tucson)

Benjamin V. Tucker

(Dept. of Linguistics, Univ. of Alberta, Edmonton, Canada)

Running Title: Cues to perception of reduced flaps

Corresponding Author:

Natasha Warner
PO Box 210028
Department of Linguistics
University of Arizona
Tucson, AZ 85721-0028
U.S.A.

Until 7/2008:
MPI
Box 310
6500 AH Nijmegen
the Netherlands

Phone: +31-24-352-1382
FAX: +31-24-352-1213
E-mail:nwarner@u.arizona.edu

ABSTRACT

Natural, spontaneous speech (and even quite careful speech) often shows extreme reduction of many speech segments, even resulting in apparent deletion of consonants. Where the flap ([ɾ]) allophone of /t/ and /d/ is expected in American English, one frequently sees an approximant-like or even vocalic pattern, rather than a clear flap. Still, the /t/ or /d/ is usually perceived, suggesting the acoustic characteristics of a reduced flap are sufficient for perception of a consonant. This paper identifies several acoustic characteristics of reduced flaps based on previous acoustic research (size of intensity dip, consonant duration, and F4 valley), and presents phonetic identification data for continua that manipulate these acoustic characteristics of reduction. The results indicate that the most obvious types of acoustic variability seen in natural flaps do affect listeners' percept of a consonant, but not sufficiently to completely account for the percept. Listeners are affected by the acoustic characteristics of consonant reduction, but they are also very skilled at evaluating variability along the acoustic dimensions that realize reduction.

PACS Number: 43.71.Es

I. INTRODUCTION

A quick look at any corpus of spontaneous speech shows that speakers do not produce every segment of a word, and do not produce sounds as one would expect (Greenberg, 1997, 1999; Johnson, 2004; Pluymaekers *et al.,* 2005a, 2005b). For example, one of our recordings includes an American English utterance [bɹ ʌ ʒ ləɪ ʔ], "but I was like," in which the speaker deleted some segments, shifted the qualities of others, and inserted r-coloration. Still, native listeners understand the utterance easily. Which ways that speech sounds vary during reduction are important for speech recognition? How acceptable is it to a listener if the speaker changes the manner of articulation of a consonant, weakening it to an approximant, vs. if the speaker shortens a consonant, or fails to produce a drop in intensity for the consonant, nearly deleting it? In this article, we focus on how reduction affects American English intervocalic /t/ and /d/ in flapping ([ɾ]) position (e.g. 'pretty, prejudice'). Previous work (e.g. Koopmans-van Beinum, 1980; Arai, 1999; Ernestus et al., 2002) shows that reduced words and sounds are quite difficult to perceive when removed from their context, although they are perceived well in context. This finding leads to the question of whether specific acoustic characteristics of reduction hinder perception.

Acoustically, prototypical flaps are characterized by a very brief closure, resembling a voiced stop closure except for its brevity (Port, 1977; Zue & Laferriere, 1979). Zue & Laferriere (1979) find an average duration of just 26-27 msec. for flaps, as compared to 75 and 129 msec. for pre-stress /d/ and /t/. They also find that the duration of the consonant does not differ for flap derived from /t/ vs. /d/, although the duration of the preceding vowel does differ slightly. Fisher & Hirsh (1976) find some differences in duration and intensity between flapped /t, d/. Flaps are not expected to have a burst, as they are expected to be so short that air pressure cannot build up behind the closure. However, Zue & Laferriere (1979) also find flap-like tokens with

surprisingly long closures more similar to a [d] but no burst, as well as tokens with short, flap-like closures with a clear burst. Horna (1998) confirms that several variants, ranging from more stop-like to more approximant-like, are possible, with the approximant-like variants more common in conversation than in read speech. De Jong (1998) and Fukaya & Byrd (2005) both give articulatory data on flaps, and argue that gradient gestural differences lead to the percept of flap as an acoustically different sound. Son (2008) shows that even if a clear tongue tip gesture occurs, there may be little or no acoustic sign of an expected flap in Korean. As for perception of flaps, Port (1977) finds that short consonantal duration is such a strong cue that the percept of "rabbit" shifts entirely to "ratted" ([b] to [ɾ]) if the [b] is made short enough, despite the conflicting place cues. McLennan *et al.* (2003, 2005) and Connine (2004) have investigated the perception of flapped /t, d/ in comparison to non-flapped stop [t] or [d] in words where flap would be expected (e.g. 'atom, Adam'). However, since flap rather than stop is the normal pronunciation in this environment, this investigates processing of allophonic variation rather than of speech reduction.

In our previous work (Warner, 2005; Warner & Tucker, 2007) on /t/ and /d/ in flapping position, we encountered many tokens with approximant-like /t, d/ (Figure 1), in carefully read speech as well as conversation. Out of more than 1900 measurable tokens of /t/ or /d/ in flapping environment, produced by seven speakers, the second and/or third formants continued at least faintly throughout the consonant in 88% of tokens. In 56% of tokens, they continued strongly throughout the consonant, as one would expect for an approximant but not a true flap. During the /t/ or /d/, intensity dipped by an average of only 10.11 dB relative to the peaks of surrounding vowels, whereas /p/ and /k/ in comparable environments showed an average dip of 32.07 dB. Of the /t, d/ tokens in our study, 70% lacked a burst, 97.5% had voicing throughout the consonant,

and many lacked a clear onset and offset. The average duration for the /t, d/ was 32 msec. An additional 78 tokens had the /t, d/ so thoroughly deleted that we could not locate any acoustic trace of it to measure. Furthermore, some tokens had a large valley in the fourth formant timed to the /t, d/ (Figure 1A), even if the consonant was extremely reduced (Dungan *et al.,* 2007; Warner & Tucker, 2008). This valley in the F4 was particularly common following /r/ (e.g. 'party, quarter'), visible in 46% of tokens (out of 120 read speech tokens with preceding /r/, from six speakers). It occurred in only 2% of tokens before the vowel /i/ and not after /r/ (e.g. 'city').

INSERT FIGURE 1 ABOUT HERE

Most reduced tokens in our previous work, despite resembling an approximant in the spectrogram, sounded clearly like a /t/ or /d/. Lexical and phonotactic expectations might contribute: 'forty' is unlikely to be misperceived as a non-word /foɹ i/, or 'status' as phonotactically impossible /stæəs/. However, even when deletion of the flap would form a real word (e.g. 'powder/power, needle/kneel'), very reduced tokens rarely sound ambiguous. Thus, even a very small acoustic cue may be sufficient for listeners to perceive the consonant.

This study investigates whether the acoustic dimensions that vary in natural productions influence listeners' percept of a /t, d/ consonant. Previous research on reduced speech has shown that listeners need surrounding context in order to recognize reduced segments or words well (Bard *et al.,* 1988; Arai, 1999; Ernestus *et al.,* 2002). Listeners also take account of how often segments reduce in various environments when compensating for reduction (Mitterer & Ernestus, 2006). Here, we turn to specific acoustic dimensions that vary during reduction of a particular segment, the flap. We investigate the effects of degree of intensity dip (size in dB)

(Experiment 1), duration of the consonant (Experiment 2), and size of F4 valley (Experiment 3) on perception of real-word pairs such as 'powder/power.' Intensity dip and duration of the consonant were the most reliable and variable measures in our previous production data. We choose to manipulate the F4 valley as well because such large, clear F4 valleys were a surprising finding in the production study. Furthermore, F4 valleys sometimes occurred even in otherwise very reduced /t, d/ tokens, so we wish to determine whether this acoustic characteristic could be a perceptual cue. We use re-synthesized continua in a phonetic identification task. Thus, when listeners hear a reduced /t, d/ as in Figure 1A, do they attend to the intensity dip (Experiment 1), the duration of the consonant (Experiment 2), and/or the valley in F4 (Experiment 3) when reconstructing the sound? The over-arching question, then, is what listeners attend to within the extreme variability of natural speech. We investigate flaps as one case of this variability, and in each experiment, we test one dimension observed in natural speech variability.

II. EXPERIMENT 1: MANIPULATING DEGREE OF INTENSITY DIP

The first experiment manipulates the degree of dip in intensity at the /t, d/. We have previously observed a wide range of intensity dips for flaps, from a large intensity dip indicative of tongue closure, even with cessation of voicing, through small intensity dips, to a few tokens with no intensity dip (Figure 2, and cf. Figure 1) (Warner, 2005; Warner & Tucker, 2007). Experiment 1 uses Praat resynthesis (Boersma & Weenink, 2008) to create two continua with a range of intensity dips, based on one word with a flap ('needle') and a matched word without ('kneel'). We predict that listeners will be less likely to hear a /t/ or /d/ in stimuli with only a small dip in intensity at the consonant. However, because we see many tokens with only minor intensity dips in natural productions, we also predict that listeners will still be able to perceive an intended /t/

or /d/ relatively often even if the intensity dip is small. That is, we predict an effect of degree of intensity dip on consonant perception, but not a large shift such as from 0% to 100%.

INSERT FIGURE 2 ABOUT HERE

A. Methods

1. Materials

A female speaker of American English produced multiple tokens of flap (VDV) and no-flap (VV) pairs, such as 'powder/power, title/tile.' This recording provided the tokens from which to create the stimuli for this experiment (intensity manipulation) as well as Experiments 2 and 3 (duration and F4 manipulations), which are reported below. Both underlying /t/ and /d/ words were included, but the speaker's dialect is that of Southern California, so she did not have a clear distinction between pairs such as 'writer/rider' based on Canadian vowel raising or vowel length. She was recorded in a sound-attenuated booth at 44.1 kHz directly to hard drive, using a high-quality stand-mounted microphone. The word list consisted of twelve pairs of words, more than were planned for inclusion as stimuli, so that appropriate items for resynthesis could be selected from among them. The recording list used a pseudo-random order, with the members of a pair not contiguous. The speaker, a linguist but not a phonetician, did not know the topic was reduction. She was asked to produce the words several times, varying her speech from "careful but natural" to "sloppy." She produced varied VDV (flap) tokens, none unnaturally careful (e.g. not [tʰ aɪ̯ɾʰl̩] for 'title'), but all within a range from flap to near-deletion, similar to the tokens produced by non-linguist subjects in our acoustic study (Warner, 2005; Warner & Tucker, 2007).

Experiment 1 used two continua, one based on a token of 'needle' and the other based on a token of 'kneel.' Tokens of other words from the recording were used for the other experiments, as described below. The base token of 'needle' (for the VDV continuum) had a moderate intensity dip (4.3 dB relative to average of surrounding vowel peaks, approximately 49 msec) and clear second and third formants through the consonant. The speaker pronounced 'kneel' as bisyllabic ([nijl]), but a token with minimal intensity dip (1.1 dB relative to average of neighboring vowel peaks, approximately 42 msec) was selected for the VV continuum. Both tokens were resynthesized using Praat's intensity editor (8 msec. steps) to flatten the intensity contour throughout the natural dip, while the rest of the signal was multiplied by a constant to maintain the shape of the intensity contour outside that area. (This was done to maintain the natural onset and offset of the word.) This token, flattened throughout the consonant, was used as step 1 of each continuum (which should sound least like it contains a /t, d/), and it was used as the base from which to synthesize steps 2-8. For each subsequent step, we decreased the intensity values for 8 time points (at 8 msec. intervals) at the time of the original, natural dip. The third through the fifth time points of the dip were reduced in increments of 3 dB per step (e.g. 3 dB for step 2, 6 dB for step 3, 21 dB for step 8). Thus, the original signal was not used as any continuum step. 21 dB was the most extreme, because a larger decrease sounded like a computer glitch rather than a consonant. (The continuum range was chosen not based on our previous production results, but as being the largest practical range. In our production data, the 5^{th} and 95^{th} percentiles for /t, d/ intensity dips fall at 2.82 and 18.66 dB, and the average at 10.11 dB, so the synthesized continuum is slightly larger than the typical natural range.) In order to ensure that extreme drops in intensity did not de facto generate longer durations of perceived intensity dip, the time points other than the most extreme part of the dip (points 1, 2 and 6

through 8) were set to 1-2dB less than the surrounding consonant, and were held constant for all continuum steps. Thus, only the degree (in dB) of the consonant-like dip varied, not its duration. In total, a 16 msec. stretch was at the lowest amplitude, and the entire dip including the gradual ramp covered 72 msec. Figure 3 illustrates several resulting stimuli for the 'needle' continuum. This procedure produced stimuli at the large-dip end of the continuum with the appearance of a brief closure and a slight ramping of intensity going into and out of it, as one often sees in natural productions that have clear flaps.

INSERT FIGURE 3 ABOUT HERE

2. Participants and procedures

Thirty-four native speakers of American English, students in introductory linguistics courses at the University of Arizona, participated in the experiment. None reported any speech or hearing disorders. Data from additional participants who did not identify themselves as monolingual English speakers were not analyzed. Participants received a small amount of course credit.

Participants sat in a sound-attenuated booth and heard the stimuli over headphones. The task was two-alternative forced choice, with a button box for responses. Six repetitions of each stimulus (normalized for overall amplitude) were presented, in a single session randomized with the stimuli for Experiments 2 and 3 below. They were not blocked by continuum or experiment, as that might induce listeners to focus on the acoustic dimension manipulated in each continuum. For each stimulus, the two words of the pair were presented on a computer screen visible through the booth window. For half the subjects, the VDV (flap) item appeared consistently on the left and the VV item on the right, with the reverse for the other half. Subjects responded which word

they had heard using the button box. For a few word pairs where deletion of the flap could lead to two alternative real words (e.g. 'waiter vs. weigher/wear'), both no-flap alternatives appeared.

Subjects first read a list of the target words, to be sure they would have all response options in mind. They then performed a practice test on similar materials, followed by the real test, with one break. For each item, the response options appeared on the screen, and one second later, the auditory stimulus began. Subjects had a 3-second window from onset of the auditory stimulus in which to respond. After a response, there was a one second pause before the following visual stimulus appeared. The EPrime software (Psychology Software Tools, Inc.) controlled the experiment and recorded responses. The experiment took approximately 20 minutes. Subjects answered questions about their language and dialect background after the experiment.

B. Results

Proportion of VDV responses (Figure 4) and reaction times (RTs), averaged over the six presentations of each stimulus, were analyzed using within-subjects ANOVAs. RTs will not be presented, however, as they generally supported the patterns in the proportion VDV data, adding little information. The factors were Step (1-8) and Continuum (VDV 'needle' vs. VV 'kneel'). A between-subjects control factor (VDV presented on left or right of screen) was included in the statistical analyses to remove variance. No subjects were outliers, so none were excluded.

INSERT FIGURE 4 ABOUT HERE

The mean proportion of VDV responses showed significance for both main effects and their interaction (Continuum: $F(1,32)=331.82$; Step: $F(7,224)=71.75$; Interaction:

F(7,224)=7.35; all p's <.001). (The proportion of VV responses is always the inverse of VDV responses.) Both continua showed more VDV responses with larger intensity dips, with significant simple effects (VDV ('needle') continuum: F(7,224)=30.85; VV ('kneel') continuum: F(7,224)=36.68, both p's <.001). Still, listeners perceive both continua predominantly as the word from which the continuum was formed: the VV-base continuum shifts from approximately 0% to 50% VDV responses, while the VDV-base continuum covers the range from 60-100%.

The significant interaction shows that the shape of the identification curve differs for the two continua. The VV-base continuum shows lesser slope at the small dip (VV percept, low step number) end of the continuum, and the VDV-base continuum shows a flattening of the curve at the opposite end of the continuum. Thus, the two continua seem to represent separate parts of a categorical perception curve, with neither continuum alone achieving a complete shift. To test this, we used interaction comparisons over restricted step ranges. An interaction comparison with the factors Step (steps 1-2 only) and Continuum showed significance for both main effects (Continuum: F(1,32)=252.16, p<.001; Step: F(1,32)=12.90, p<.005) and their interaction (F(1,32)=15.61, p<.001). The simple effect of continuum step (steps 1-2 only) was significant for the VDV-base continuum (F(1,32)=14.96, p<.005), but not the VV-base continuum (F(1,32)=1.00, p>.05). A second interaction comparison of steps 5 vs. 8 also showed significance for both main effects and their interaction (Continuum: F(1,32)=86.34, p<.001; Step: F(1,32)=6.24, p<.02; Interaction: F(1,32)=7.39, p<.02), but this time, the simple effect of Step (5 vs. 8) was significant for the VV-base (F(1,32)=7.50, p<.02) but not the VDV-base (F<1) continuum. This shows that the VV-base continuum is flat at low steps (little or no intensity dip), while the proportion of VDV judgments is already increasing for the VDV continuum. For the large-dip steps (5 vs. 8), the VDV continuum has already reached ceiling, but the VV

continuum is still increasing. Both continua show parts of a categorical perception curve, as both have a plateau and a range showing increase. However, they show opposite parts of the complete curve, even though they cover the same range of intensity dips: 0 - 21 dB decrease.

C. Discussion

These data verify that the size of the intensity dip is a cue to the perception of a /t, d/ consonant. A large dip in intensity increases perception of a flap. Thus, when we see variability in natural speech in how deeply intensity drops during a word with flapped /t/ or /d/, this variability is indeed along an acoustic dimension that influences how consonantal the token sounds.

However, the effect of base continuum in this experiment is also quite large, and the two continua cover different parts of the categorical perception curve despite their equal acoustic range. Thus, there must be other acoustic cues. Even a large intensity dip (21 dB) does not make 'kneel' sound as if it contained a /d/ more than about half the time. This is not a failure to include a sufficient continuum range: a larger dip in pilot stimuli sounded like a non-speech sound, and the 'needle' continuum verifies that a larger dip is not necessary to reach 100% VDV judgments. In the other direction, even deletion of the intensity dip fails to make 'needle' into 'kneel' more than 50% of the time. This matches with our observation from the production study that even tokens with little intensity dip often have a clear consonantal percept. In Experiment 2, we turn to another acoustic dimension: duration, rather than degree, of the intensity dip.

III. EXPERIMENT 2: MANIPULATING THE CONSONANT DURATION

In natural speech data, the /t, d/ varies greatly in duration (Warner, 2005; Warner & Tucker, 2007). Even a clear flap closure is short, since this sound is defined by its quick "flap" of the

tongue against the roof of the mouth, but there is still a considerable range of consonant duration. Our production data for /t, d/ had the 5th and 95th percentiles of consonant duration at 15 and 56 msec., respectively, with 32 ms as the average. Although reduced /t, d/ can be difficult to measure, consonant duration correlated well with degree of intensity dip, with clearer flaps being longer. We posited that duration might also be a significant cue. Experiment 2 manipulates duration of the consonant independently in order to determine whether duration is itself a salient cue to the /t, d/. Because of the correlation of duration with other measures of reduction (shorter durations for more reduced tokens), we predict that listeners will be less likely to perceive a /t, d/ with shorter duration. However, since the flap is inherently a very short sound, even an extremely short flap may be rather perceptible, at least if it has a clear intensity dip or gap in formant structure.

A. Methods

Three tokens of VDV words (Figure 5) were chosen from the same recording as in Experiment 1, from which to resynthesize the duration continua. One token of 'needle' was a clear flap realization, with a substantial intensity dip (11.5 dB relative to surrounding vowel peaks), sudden onset and offset of the tongue closure as evidenced by a gap in formants, and voicing throughout. One token of 'waiter' was an approximant-like realization, with a smaller dip in intensity (7.3 dB) and no sudden change in formants indicative of closure. Finally, one token of 'title' represented an unusually clear flap realization: this token had at most extremely low amplitude voicing during the consonant (16.0 dB intensity dip), and could perhaps be considered voiceless.

INSERT FIGURE 5 ABOUT HERE

For this experiment, no VV word was used as a base form, because it would not be clear what portion of the signal to lengthen or shorten to manipulate consonant duration. However, the use of three VDV base words probes whether the effect of consonant duration differs for various realizations of the consonant. Because all of the base tokens were intended as VDV words, they were used as the sixth step of the eight-step continuum (near the VDV end).

We used PSOLA resynthesis within Praat to manipulate consonant durations (Figure 6). A region covering the intensity dip was located for each base form. This region extended to near the intensity maxima for the surrounding vowels, and was thus larger than what would be measured as the consonant duration. PSOLA was then used to lengthen this region to 1.2 and 1.4 times its original duration for continuum steps 7 and 8, and to shorten it to .8, .6, .4, .2, and 0 times its duration for steps five through one. Since shortening the region of the intensity dip also tends to lessen the degree of the dip (in dB), we then located the lowest amplitude glottal period in the original signal and spliced it into the resynthesized forms for steps 2-5 and 7-8, replacing the lowest amplitude period of each one. This guaranteed that each stimulus above step 1 did drop to the same extent (in dB) as it originally did, however briefly. This was not done for step 6 (the original item), or for step 1, which had no intensity dip. This single low-amplitude period had a duration very similar to the period it replaced, usually within 1 msec, so that this manipulation did not affect the step-wise manipulation of duration. The duration range for the continuum was thus chosen not based on our production data, but on the base tokens used. For step 2 (the shortest dip without deletion of it, .2 of original duration), the resulting duration of the manipulated portion of the signal (more than the consonant itself) was 10-11 msec. for each continuum. Despite the necessity of rapid intensity changes, splicing at zero-crossings at

consistent points of the glottal pulse, and the use of PSOLA, prevented the introduction of spurious burst-like noises. For step 8 (the longest dip, 1.4 of original duration), the resulting duration of the manipulated portion was 66 msec. for 'needle' and 'waiter,' and 74 msec. for 'title.'

INSERT FIGURE 6 ABOUT HERE

The subjects and procedures were identical to those for Experiment 1. As described above, the stimuli for all three experiments were presented in a single session, in random order.

B. Results

The proportion VDV results (Figure 7) was analyzed using an ANOVA with the within-subjects factors Continuum (near-voiceless closure, voiced closure, approximant) and Step (1-8), and the same between-subjects control factor (response side of screen) as above. Both main effects and the interaction were significant (Continuum: $F(2,64)=15.94$; Step: $F(7,224)=65.99$; Interaction: $F(14,448)=17.69$; all p's $<.001$). The simple effect of Step showed significantly more VDV responses at longer consonant durations for all three continua (near-voiceless: $F(7,224)=8.77$; voiced closure: $F(7,224)=70.21$; approximant: $F(7,224)=21.29$; all p's $<.001$). However, this includes step 1, which lacks an intensity dip entirely.

INSERT FIGURE 7 ABOUT HERE

To be sure that duration of the consonant's dip, rather than just its presence, affects the percept, we used an interaction comparison of only steps 2-5, the region containing ambiguity

outside the no-dip first step. Both main effects as well as the interaction were significant (Continuum: $F(2,64)=5.10$, $p<.01$; Step: $F(3,96)=12.54$, $p<.001$; Interaction: $F(6,192)=3.94$, $p<.005$), and the simple effects showed an increase across this duration range for the voiced closure continuum ($F(3,96)=14.19$, $p<.001$) and the approximant continuum ($F(3,96)=4.11$, $p<.01$), but not the near-voiceless continuum ($F(3,96)=1.21$, $p>.05$). Thus, if the /t, d/ is realized as a less obstruent-like consonant, longer duration makes it sound more consonantal and shorter duration makes it sound more deleted. However, if its intensity dips very low, then even an extremely short dip of effectively one glottal period is sufficient to make the /t, d/ quite perceptible (VDV response at ceiling). Furthermore, even though two continua do show effects of consonant duration without step 1, both already receive more than 80% VDV judgments by step 2. Thus, even an approximant-like realization does not need a long duration to be perceived most often as containing a /t/ or /d/. It appears that almost any dip in intensity, no matter how small in degree or duration, can be perceived as a realization of /t, d/.

Even with extremely long consonant duration, the approximant continuum in Figure 7 never reaches the high level of VDV responses the other two continua do. In an interaction comparison of steps 4-8, only the main effect of Continuum was significant (Continuum: $F(2,64)=7.52$, $p<.005$; Step: $F<1$; Interaction: $F(8,256)=1.01$, $p>.05$). The voiced closure continuum had more VDV responses than the approximant continuum (main effect of Continuum for just these two: $F(1,32)=10.29$, $p<.005$), but it did not differ from the near-voiceless continuum ($F(1,32)=1.15$, $p>.05$). Thus, the voiced closure continuum patterns with the near-voiceless continuum in being at ceiling for longer durations, while the approximant continuum is not at ceiling. If a /t, d/ is realized as an approximant, rather than as a clear flap, even durations up to 1.4 times natural duration do not render it unambiguously consonantal.

These comparisons (steps 2-5 and 4-8) together show that the voiced closure continuum patterns with the approximant continuum at short consonant durations, but with the near-voiceless continuum at long consonant durations. If the intensity dip is short, it has to be a very clear dip to definitively sound like a realization of /t, d/, but if it is long, a slight dip is sufficient.

Because step 1 lacks the consonantal dip entirely, comparing steps 1 and 2 shows how much the presence vs. absence of even a very short dip affects the percept. (This is different from Experiment 1 above, which lacked short intensity dips.) In an interaction comparison of just steps 1 and 2, the estimated effect size of the main effect of Step was .71 (partial eta-squared), whereas in an interaction comparison of step 2 to step 8, it was .34. The majority of the increase in VDV responses happens between steps 1 and 2, not at longer durations (Figure 7). Thus, the presence of any intensity dip at all has more impact on the percept of a consonant than even a large difference in the duration of that dip (from .2 to 1.4 times the original duration).

C. Discussion

These results clarify several points. First, consonant duration is clearly a perceptual cue to /t, d/ in flapping environment. However, in the continuum we tested with a very low-intensity consonant (near-voiceless closure), any duration of consonant at all is sufficient to cue its presence. In the continua with a fully voiced closure or a reduced, approximated consonant, the change from short to moderate duration increases perception of the consonant. This data also demonstrates that the presence vs. absence of any intensity dip at all has a far greater effect on perception of a /t, d/ than even a large difference in consonantal duration. An extremely short dip, even one lacking consonantal closure, still contributes greatly to listeners' percept of a /t, d/.

Finally, the continuum with the middle degree of closure (voiced closure) patterns with the approximant continuum at short durations, but with the near-voiceless one for long consonants. If the consonant is short, only the particularly strong closure suffices for it to be definitively perceived. However, if the consonant is long, the weaker closures we tested still lead listeners to perceive the consonant. The approximant we tested, however, is not fully perceived as a realization of /t, d/, and lengthening it does not make it any more like a /t, d/. Thus, even though natural speech very often has /t, d/ realized as approximants, they are not fully accepted by the listener. However, this decrement for long approximated consonants is small: the /t, d/ is perceived in over 90% of such stimuli.

IV. EXPERIMENT 3: MANIPULATING THE F4 VALLEY

In our previous production work, we noticed a surprisingly large change in F4 in some tokens where a flap would be expected (Dungan *et al.,* 2007; Warner & Tucker, 2008). This valley in F4 (Figure 1A) can traverse a range of 1000 Hz, and it is timed to the flap consonant, not to a neighboring segment. Some speech sounds do have systematic effects on F4, including retroflexes, American English /r/, and taps or flaps in some other languages (Espy-Wilson *et al.,* 2000; Avelino & Kim, 2002; Hamann, 2003; Zhou *et al.,* 2007). However, systematic effects on F4 are rare. In our previous work, we found that this F4 valley is most common after /r/ (e.g. 'hurdle, fertilizer,' 46% of tokens), and least common before /i/ (e.g. 'beauty,' 2% of tokens). It does not occur in the majority of tokens (visibly in 18% of tokens overall), and F4 is not always clear, but when the valley does occur, it is often a striking visual effect. Furthermore, even some tokens with extremely reduced /t, d/ have a clear F4 valley. Since such tokens appear to offer few other cues to the consonant, we wondered whether the F4 valley was a perceptual cue.

Because the F4 valley can be so striking, and can be the only apparent acoustic realization of the consonant, we predict that listeners can make some use of an F4 valley to detect a /t, d/. However, listeners may have little reason to attend to F4 in the language overall, and F4 has low amplitude. For this reason, we predict that any effect of the F4 valley will be small. To test this, we manipulated F4, using LPC resynthesis in Praat.

A. Methods

1. Materials

Three tokens, two of 'quarter' (VDV) and one of 'core' (VV), were chosen from the same recording used for Experiment 1. The speaker pronounced 'quarter' with onset /k/ matching 'core,' not a /kw/ cluster. One token of 'quarter' we selected had a clear F4 valley and a relatively large intensity dip. F4 was clearly visible in the spectrogram throughout, a requirement for this experiment. A second token of 'quarter' had a clear F4 valley but minimal intensity dip. That is, the /t/ in this token was nearly deleted, but the F4 valley remained. Finally, a token of 'core' with clearly visible F4 but no valley in it was chosen as a matched VV item. The use of two VDV items allows for comparison of a continuum where the F4 valley might be the primary cue to the /t/ to one where other cues are obvious. All three tokens were downsampled to 11,000 Hz, and formants were located through LPC analysis, using a prediction order of 10 for 'core' and of 12 for both tokens of 'quarter.' (The higher prediction order was necessary for accurate tracking of F4, in order to resynthesize without leaving traces of the original F4.) We then inverse filtered the tokens, using the LPC analysis, to obtain estimated glottal source functions. After manipulating the formant values (5 msec time step) as described below, we resynthesized to create the stimuli. Figure 8 shows the resynthesized extreme steps for all three continua.

INSERT FIGURE 8 ABOUT HERE

The formants were manipulated as follows. The time range of interest, lasting 85-100 msec, was located. It was from the beginning of rapid decrease to the end of rapid increase in F4 in the two tokens of 'quarter,' and for a range timed similarly relative to the onset of the preceding vowel for 'core' (which had no F4 valley). Any outlier points in the F4 LPC track during that time range were manually corrected. For the two VDV ('quarter') continua, the F4 at step 8 of the continuum (maximal F4 valley) used the original values, except for such outlier correction. For step 1, the F4 was measured at the time points immediately outside this time period, and the F4 was interpolated (in Barks) from the preceding to the following value over the time range. For steps 2-7, the difference between the F4 for step 8 (natural) and step 1 (interpolated, straight F4) was divided into perceptually equal steps, as measured in Barks. The range F4 valley size was thus determined by the base token, not by overall averages from production results. Because F4 is not clear in all naturally produced tokens and the valley does not always occur, we chose to base the continuum range on the naturally produced valley in a token with a clear F4 valley, rather than on production data averaged across tokens.

For the VV ('core') continuum, there was no original F4 valley, so one had to be added to create step 8. Furthermore, the speaker's F3 dropped for the /r/ considerably later in 'core' than in 'quarter,' and this meant that if F3 were not manipulated, the added F4 valley would cross over the F3. Therefore, for 'core,' the F3 was lowered to 504 Hz below the F4 value of step 8 (lowest F4) for the descending portion of the F4 valley. After reaching the minimum, the F3 remained at that value until the natural F3 dropped below that value later in the word, at which point the

natural F3 values were allowed to resume. (We did not wish to create a drop-rise pattern in F3 as well as F4.) This 504 Hz separation was based on the average for the vowel up until the manipulated time period. The same lowering of F3 was applied to all steps of the 'core' continuum, so that only F4 would vary by continuum step. Once F3 was thus moved out of the way of F4, a somewhat parabola-shaped F4 valley of 1000 Hz was created for step 8, modeled on the properties of the natural F4 valleys in the VDV continua. For the VV continuum, the natural F4 values were used as step 1, and the difference between steps 1 (natural) and step 8 (added F4 valley) was calculated and divided among the other steps as for the VDV continua.

After resynthesis, the low-step stimuli lacked any F4 valley, showing steady F4. The high-step stimuli had a clear F4 valley resembling that of natural tokens. Those steps with original F4 values were also resynthesized, so that they would be equally degraded by LPC resynthesis. The stimuli for this experiment did contain LPC clicking noises that the other experiments' stimuli did not. However, they were still clear realizations of either 'quarter' or 'core.'

2. Subjects and procedures

Subjects and procedures were identical to Experiments 1 and 2. All three experiments were presented together, so the LPC-degraded stimuli were randomized among the higher quality PSOLA and intensity resynthesis items. The instructions mentioned that some items sounded like computer speech, and the practice items included some created through LPC resynthesis.

B. Results

ANOVAs were used to analyze the data (Figure 9) with the within-subjects factors Continuum (VDV-intensity dip, VDV-F4 valley only, VV) and Step (1-8) and the usual between-subjects

control factor (response side of screen). Only the main effect of Continuum was significant, with both VDV-base continua perceived as VDV far more often than the VV continuum was (Continuum: $F(2,64)=27.12$, $p<.001$; Step: $F(7,224)=1.72$, $p>.05$; Interaction: $F<1$).

INSERT FIGURE 9 ABOUT HERE

One VDV continuum (F4 valley only) might show some effect of Step, despite the lack of an interaction. In order to check for any possible effect of the F4 valley, we performed post hoc comparisons of steps 1 and 2 to step 8 for this continuum. Step 8 was identified as VDV significantly more often than either step 1 ($F(1,32)=5.31$, $p<.03$) or step 2 ($F(1,32)=6.13$, $p<.02$).

C. Discussion

The stimuli in this experiment were nearly always perceived as the base word from which they were formed, regardless of how F4 was manipulated. The presence of an F4 valley does affect perception, but the effect is extremely small, and is limited to the VDV-base continuum that lacked a strong dip in intensity. Based on what is known about perceptual cue-trading (e.g. Repp, 1983), it is not surprising that any perceptual effect of the F4 valley is limited to the continuum with the highest chance of ambiguity, where the /t/ was very approximant-like, with intensity nearly as great as that of the surrounding vocalic sounds. The formants continued strongly throughout the consonant, and there was certainly no consonant closure. The valley in F4 might be important, because it is the only visually clear trace of the /t/ remaining. If listeners ever attend to the F4 valley, it would be in an approximated token such as this one. However, the results show minimal use of the F4 valley. Furthermore, listeners perceived a /t/ in even this

continuum at nearly ceiling, in more than 94% of tokens even for step 1 (with flattened F4). Thus, there must be ample perceptual cues to the /t/ aside from F4 valley.

Still, the fact that there is any perceptual effect of the F4 at all is noteworthy. The fourth formant does not provide important cues to other segmental distinctions in English, as far as we know. Although American English /r/ may affect F4 (Zhou *et al.,* 2007), its low F3 is a far more likely cue (Best & Strange, 1992). The literature shows L2 listeners have difficulty learning to attend to acoustic dimensions not used for L1 distinctions (Best & Strange, 1992; Iverson *et al.,* 2003; Wagner *et al.*, 2006). Thus, English listeners should not be very good at using F4 variation as a cue. Furthermore, because we randomized the stimuli of all three experiments (total of 8 continua, varied on 3 dimensions), listeners probably could not learn to attend to the F4 valley over the course of the experiment. In addition to F4 not being a perceptual cue otherwise, it is also acoustically weak, with lesser amplitude than the lower formants. These factors seem to outweigh the fact that a drop of 1000 Hz in the F4 is a striking acoustic characteristic. The result is a minimal, but present, perceptual effect of the F4 valley.

V. GENERAL DISCUSSION

These experiments show a large effect of the degree of intensity dip on perception of a /t, d/ in flapping environment, a relatively large effect of the presence vs. absence of any intensity dip at all, a smaller effect of consonant duration, and an extremely small effect of the F4 valley. Larger and longer intensity dips are perceived as more consonantal, but even a very short intensity dip can be enough to cue the presence of /t, d/. Furthermore, there must be other cues to the reduced consonant, as none of the continua span the 0-100% response range. It is possible that the dimensions we manipulated would show a greater shift in the absence of other cues (cue-trading,

cf. Repp, 1983). However, we used a variety of base forms, making multiple continua that differed in the presence and strength of alternative cues, to allow for this possibility. Experiment 2, manipulating duration, exemplifies this with the use of three base tokens for resynthesis, differing widely in the other cues to the consonant. Because we did not systematically manipulate two or more cues at once through an entire range, we cannot be sure of potential cue trading. It could be that the F4 valley would show a larger effect with other cues more ambiguous, and it could be that duration might cause a larger effect if intensity were manipulated simultaneously. However, the use of several base tokens that vary widely on the most likely other cues does provide information about the probable range of effect sizes for each cue.

Experiment 3 manipulated F4, adding or removing a large drop-rise pattern. This was based on our observation of a striking valley in F4, often traversing 1000 Hz from surrounding vowels, in some tokens where flapped /t, d/ is expected. Experiment 3 showed that despite the large acoustic change, the F4 valley had minimal effect on listeners, and even that only when there were no other obvious cues to the consonant. This suggests that the valley observed in our production work is probably an articulatory artifact, having to do with a constriction the tongue moves through between targets for surrounding segments. This artifact may appear striking on a spectrogram, but not be so to listeners. However, listeners are slightly able to use this fine phonetic detail of the speech signal, despite the reasons for them to ignore it (i.e. low amplitude of F4 and lack of importance of F4 for other distinctions). In future research, it would be possible to test the F4 valley while manipulating duration of the pre- and post-valley portions and degree of intensity dip, to determine whether the F4 valley might play a larger role in a cue-trading relationship. However, we suspect that F4 is simply not used as a major perceptual cue.

Duration of the consonant clearly is a cue to its presence. However, Experiment 2 showed that the presence of any intensity dip in the signal at all, no matter how short it is, is more important than even a very large difference in consonant duration. Because the flap is normally a very brief consonant, usually with an intensity dip but otherwise with quite a bit of variability, listeners may categorize almost any brief intensity dip as a realization of /t, d/ most of the time.

Experiment 1 shows that the degree of the intensity dip (its size in decibels rather than its duration) is a relatively strong cue to the /t, d/. However, even these continua fail to traverse the 0-100% range, despite covering the largest practical acoustic range. Furthermore, the continua based on a VDV vs. a VV token represent different parts of the categorical perception curve. Thus, even in this case, there must be other perceptual cues.

In typical categorical perception experiments, the continuum should ideally cover the range from 0-100% responses. However, the fact that this did not happen here is not a failure of the experiments. Since we tested acoustic dimensions that vary in reduced vs. careful realizations, if any of these dimensions led to a VDV-base stimulus receiving 0% VDV responses, it would mean that listeners were unable to understand reduced realizations of the word. We are testing for whether the dimensions that vary in natural productions of one category (VDV) affect perception of that category. They do, but not to the extent of fully turning tokens into the other category (VV). This makes sense: if a smaller intensity dip made 'title' into 'tile,' then the amount of variability present in natural speech would mean that there was a merger in progress, rather than simply synchronic variability in what seem to be stable categories.

The three experiments together indicate that there must be additional cues to the presence vs. absence of a 'flapped' /t, d/, in addition to the ones tested here. Some possible cues are the duration of the preceding and following segments (vowels, /r/, or syllabic /l/), and the timing of

formant transitions between surrounding segments. We do not attempt to identify and test every potential cue to the VV/VDV distinction in this paper. Rather, we set out to determine whether the acoustic dimensions that vary when speakers reduce their flaps affect how /t, d/-like the resulting sound is. Since reduction sometimes resembles deletion, how do the dimensions of reduction affect the listener's percept? The results show that the dimensions that differ between clear vs. reduced realizations do affect how likely listeners are to perceive a consonant.

However, approximated forms are accepted as realizations of /t, d/ almost, but not quite, at ceiling rates. A highly reduced flap does not entirely count as a flap, but it is very acceptable. Listeners are very tolerant of the kinds of variation that happen in natural reduced speech. They are sensitive to this variation, and our other work shows they use reduction in deciding what realizations to expect in upcoming speech, adjusting for reduction in the context (Tucker, 2007). However, these same dimensions that cue reduction are also cues to the presence or absence of the consonant. Listeners are very skilled at combining cues to segmental content with their knowledge of variability and reduction to perceive the intended word.

The vast majority of speech perception research investigates acoustic cues to the distinction between one sound and another--the presence of one sound vs. the presence of another. There is a long history of research on place of articulation, voicing, or other distinctions such as fricatives vs. affricates (Raphael, 2005). There has also been work on cues to the presence vs. absence of some segments, e.g. in the "say/stay" and "slit/split" distinctions (Repp, 1983; Raphael, 2005). The current study is similar in investigating the presence vs. absence of intervocalic /t, d/, as in 'waiter/weigher, needle/kneel,' etc. However, it is about more than cues to the presence of a consonant. It is also about perceptual use of the cues that differentiate a clear production of a segment from a reduced one. Reduction is rampant in natural speech

(Greenberg, 1997, 1999; Pluymaekers *et al.,* 2005a, 2005b; Johnson, 2004), and surprisingly common even in careful speech (Warner & Tucker, 2007). Thus, to understand how listeners comprehend speech, we need to study acoustic characteristics that vary with reduction.

As a phonetician, when one records 'quarter' or 'title' and sees an approximant or a vocalic sequence rather than a flap, one is often surprised. We noticed many tokens in our production study (Warner, 2005; Warner & Tucker, 2007) that were articulatorily not flaps (no consonantal closure), and that was part of the motivation for this study: how do listeners react to variations in whether a consonant has a closure? Are approximated, non-canonical realizations acceptable where a flap is expected? Does reduction make the consonant less consonantal? Although we as phoneticians may be surprised by such non-flap-like realizations in a spectrogram, listeners are in fact less surprised. Reduction does affect how clearly listeners perceive a /t, d/, but listeners are very much able to cope with such variability. This is exactly what listeners should do: the acoustic dimensions of reduction tested here are dimensions that vary during natural speech. Listeners are tolerant of highly variable speech that contains both nearly deleted versions of "flaps" and clear versions of them, because such variability is typical in the spontaneous, natural language listeners hear most often. Therefore, listeners use the information available in the duration of the consonant and the degree of its intensity dip, but they also evaluate these cues relative to the wide range of realizations of 'flapped' /t, d/ they normally hear.

Acknowledgments

We would like to thank Holger Mitterer, James McQueen, Alexandra Jesse, and Anne Cutler, as well as three anonymous reviewers for their helpful comments. We also thank Jesse Tucker and Kayla McDaniel for assistance with the work. Any errors are, of course, our own.

REFERENCES

Arai, T. (**1999**). "A case study of spontaneous speech in Japanese," Proceedings of the International Congress of Phonetic Sciences (ICPhS), Vol. 1, pp. 615-618, San Francisco.

Avelino, H. and Kim, S. (**2002**). "An articulatory and acoustic study of Pima coronals," J. Acoust. Soc. Am. **112**, 2419.

Bard, E. G., Shillcock, R. C., and Altmann, G. T. M. (**1988**). "The recognition of words after their acoustic offsets in spontaneous speech: Effects of subsequent context," Percept. Psychophys. **44**, 395-408.

Best, C. T., and Strange, W. (**1992**). "Effects of phonological and phonetic factors on cross-language perception of approximants," J. Phonetics **20**, 305-330.

Boersma, P. and Weenink, D. (**2008**). Praat: doing phonetics by computer (Version 5.0.08) [Computer program]. Retrieved February 11, 2008, from http://www.praat.org/.

Connine, C. M. (**2004**). "It's not what you hear but how often you hear it: On the neglected role of phonological variant frequency in auditory word recognition." Psychonomic Bull. and Review **11,** 1084-1089.

de Jong, K. (**1998**). "Stress-related variation in the articulation of coda alveolar stops: flapping revisited," J. Phon. **26,** 283-310.

Dungan, M., Morian, K., Tucker, B. V., and Warner, N. (**2007**). "Fourth formant dip as a correlate of American English flaps," J. Acoust. Soc. Am. **121**, 3167.

Ernestus, M., Baayen, R. H., and Schreuder, R. (**2002**). "The recognition of reduced word forms," Brain Lang. **81**, 162-173.

Espy-Wilson, C. Y., Boyce, S. E., Jackson, M., Narayanan, S., and Alwan, A. (**2000**). "Acoustic modeling of American English /r/," J. Acoust. Soc. Am. **108**, 343–356.

Fisher, W. M., and Hirsh, I. J. (**1976**). "Intervocalic flapping in English," Papers from the Regional Meetings, Chicago Linguistic Society, 1976, pp. 183-198.

Fukaya, T., and Byrd, D. (**2005**). "An articulatory examination of word-final flapping at phrase edges and interiors," J. Int'l. Phon. Assoc. **35,** 45-58.

Greenberg, S. (**1997**). "On the origins of speech intelligibility in the real world," Proceedings of the ESCA Workshop on Robust Speech Recognition for Unknown Communication Channels, Pont-a-Mousson, France, pp. 23-32.

Greenberg, S. (**1999**). "Speaking in shorthand - A syllable-centric perspective for understanding pronunciation variation," Speech Commun. **29**, 159-176.

Hamann, S. (**2003**). *The Phonetics and Phonology of Retroflexes* (LOT, the Netherlands).

Horna, J. E. (**1998**). *An Investigation into the Acoustics of American English Flaps, with a Secondary Emphasis on Spanish Flaps, in Fluent Speech*. Ph.D. dissertation, New York Univ.

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., and Siebert, C. (**2003**). "A perceptual interference account of acquisition difficulties for non-native phonemes," Cognition **87**, B47-B57.

Johnson, K. (**2004**). "Massive reduction in conversational American English," In *Spontaneous Speech: Data and Analysis. Proceedings of the 1st Session of the 10th International Symposium*, edited by K. Yoneyama and K. Maekawa (The National International Institute for Japanese Language, Tokyo, Japan), pp. 29-54.

Koopmans-van Beinum, F. J. (**1980**). *Vowel Contrast Reduction: An Acoustic and Perceptual Study of Dutch Vowels in Various Speech Conditions*. Ph.D. dissertation, Univ. Amsterdam.

McLennan, C. T., Luce, P.A., and Charles-Luce, J. (**2003**). "Representation of lexical form," *J.* Exper. Psych.: Learning, Memory, and Cognition **29**, 539-553.

McLennan, C. T., Luce, P. A., and Charles-Luce, J. (**2005**). "Representation of lexical form: Evidence from studies of sublexical ambiguity," J. Exper. Psych.: Human Perception and Performance **31**, 1308-1314.

Mitterer, H., and Ernestus, M. (**2006**). "Listeners recover /t/s that speakers reduce: evidence from /t/-lenition in Dutch," J. Phonetics **34**, 73-103.

Pluymaekers, M., Ernestus, M., and Baayen, R. H. (**2005a**). "Lexical frequency and acoustic reduction in spoken Dutch," J. Acoust. Soc. Am. **118**, 2561-2569.

Pluymaekers, M., Ernestus, M., and Baayen, R. H. (**2005b**). "Articulatory planning is continuous and sensitive to informational redundancy," Phonetica **62**, 146-159.

Port, R. F. (**1977**). "The influence of tempo on stop closure duration as a cue for voicing and place," Haskins Labs Status Report on Speech Res. **SR-51/52,** 59-73.

Raphael, L. J. (**2005**). "Acoustic cues to the perception of segmental phonemes," In *The Handbook of Speech Perception,* edited by D.B. Pisoni and R.E. Remez (Blackwell), pp. 182-206.

Repp, B. H. (**1983**). "Trading relations among acoustic cues in speech perception are largely a result of phonetic categorization," Speech Commun. **2**, 341-361.

Son, M. (**2008**). "Pitfalls of spectrogram readings of flaps," J. Acoust. Soc. Am. **123,** 3079.

Tucker, B. V. (**2007**). *Spoken Word Recognition of the Reduced American English Flap.* Ph.D. dissertation, Univ. Arizona.

Wagner, A., Ernestus, M., and Cutler, A. (**2006**). "Formant transitions in fricative identification: the role of native fricative inventory," J. Acoust. Soc. Am. **120**, 2267-2277.

Warner, N. (**2005**). "Reduction of flaps: speech style, phonological environment, and variability," J. Acoust. Soc. Am. **118**, 2035.

Warner, N., and Tucker, B. V. (**2007**). "Categorical and gradient variability in intervocalic stops," presented at the Linguistic Society of America Annual Meeting, Anaheim, California.

Warner, N. and Tucker, B. V. (**2008**). "Fourth formant drop as a correlate of American English flaps," presented at the Linguistic Society of America Annual Meeting, Chicago, Illinois.

Zhou, X., Espy-Wilson, C., Tiede, M., and Boyce, S. (**2007**). "Acoustic cues of 'retroflex' and 'bunched' American English rhotic sound," *J. Acoust. Soc. Am.* **121**, 3168.

Zue, V. W., and Laferriere, M. (**1979**). "Acoustic study of medial /t,d/ in American English," *J. Acoust. Soc. Am.* **66**, 1039-1050.

FIGURE CAPTIONS

Figure 1. Spectrograms and waveforms of a reduced, approximant-like consonant in 'fertilizer' (A) and a clear flap in 'spider' (B) for intervocalic /t/ or /d/.

Figure 2. Waveforms, spectrograms, and overlaid intensity curves of /t, d/ realized with differing degrees of intensity dip. A: Large intensity dip indicative of tongue closure, with voicing ending by the burst, in 'needy.' Consonant duration is marked by vertical lines (defined by F2 offset/onset). B: No visible intensity dip for the consonant, in '...with it if...'. Consonant duration cannot be measured.

Figure 3. Example stimuli for the intensity VDV-base continuum ('needle'). Waveform, spectrogram, calculated intensity contour (overlaid on spectrogram), and stipulated intensity contour (multiplied by identical base token), for steps 1 (flat), 3, and 8 (largest dip). A, B, and C are steps 1, 3, and 8 respectively. Intensity is displayed over the same range for all three stimuli.

Figure 4. Responses (proportion VDV) for Experiment 1, size of intensity dip continua. Low step numbers have flat intensity through the consonant; high dip numbers have a large intensity dip.

Figure 5. Waveforms of the three base tokens used to make the duration continua. Vertical lines delimit the consonant (defined by onset/offset of clear F2), and the duration of the consonant is shown. A: 'Needle,' a typical clear flap. B: 'Waiter,' an approximant realization. C: 'Title,' a very clear flap with extremely low amplitude and near-voicelessness during the closure.

Figure 6. Waveforms of steps 1, 2, and 8 of the 'title' duration continuum.

Figure 7. Responses (proportion VDV) for duration continua. Low step numbers have shortened or removed consonantal duration; high step numbers have lengthened consonant duration. The original productions are at step 6.

Figure 8. Spectrograms of step 1 (no F4 valley, on left) and step 8 (maximal F4 valley, on right) stimuli, for A: 'quarter' (VDV-intensity dip), B: 'quarter' (VDV-F4 valley only), and C: 'core' (VV). Overlaid dashed lines trace F4.

Figure 9. Responses (proportion VDV) for the F4 continua. Low step numbers have flat F4 (natural for VV); high step numbers have a large F4 valley (natural for VDV).

Fig. 1

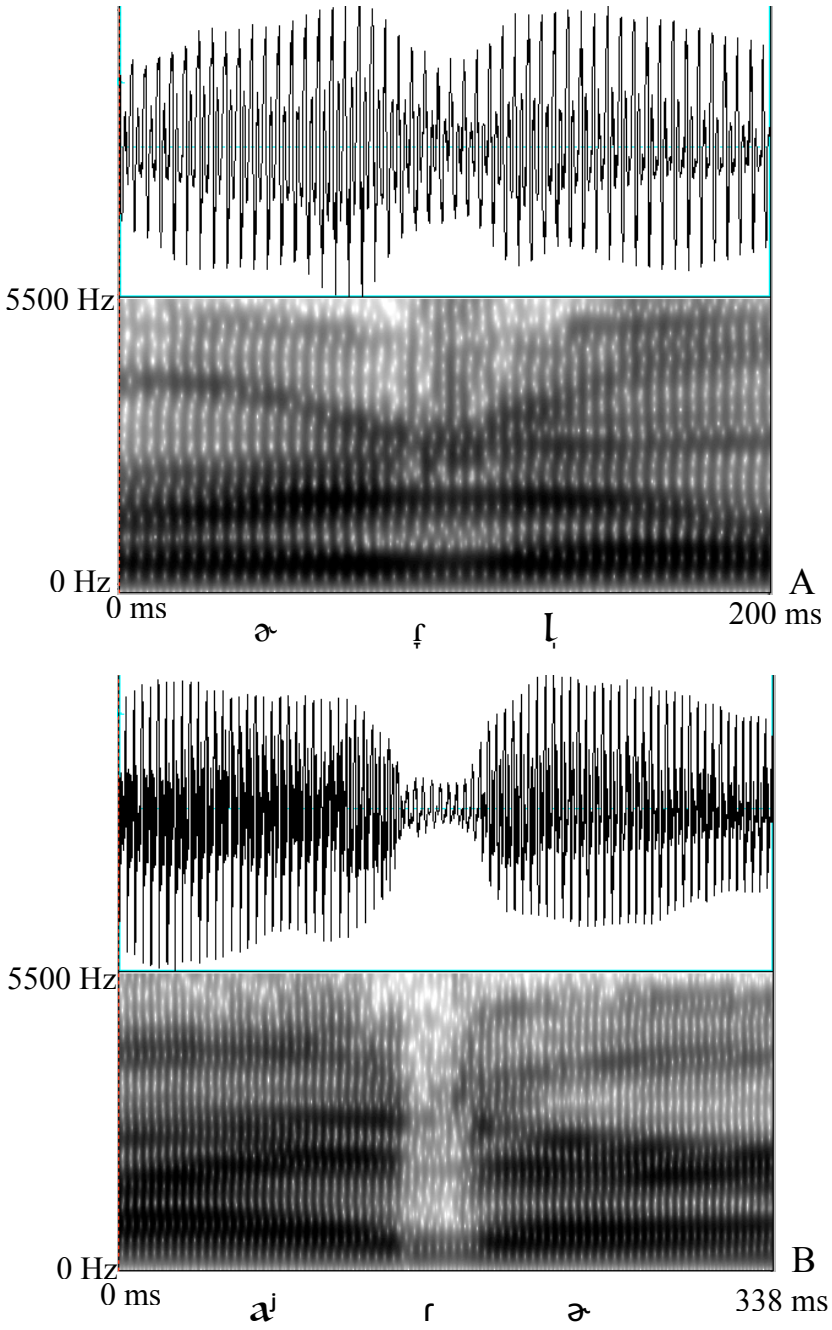


Fig. 2

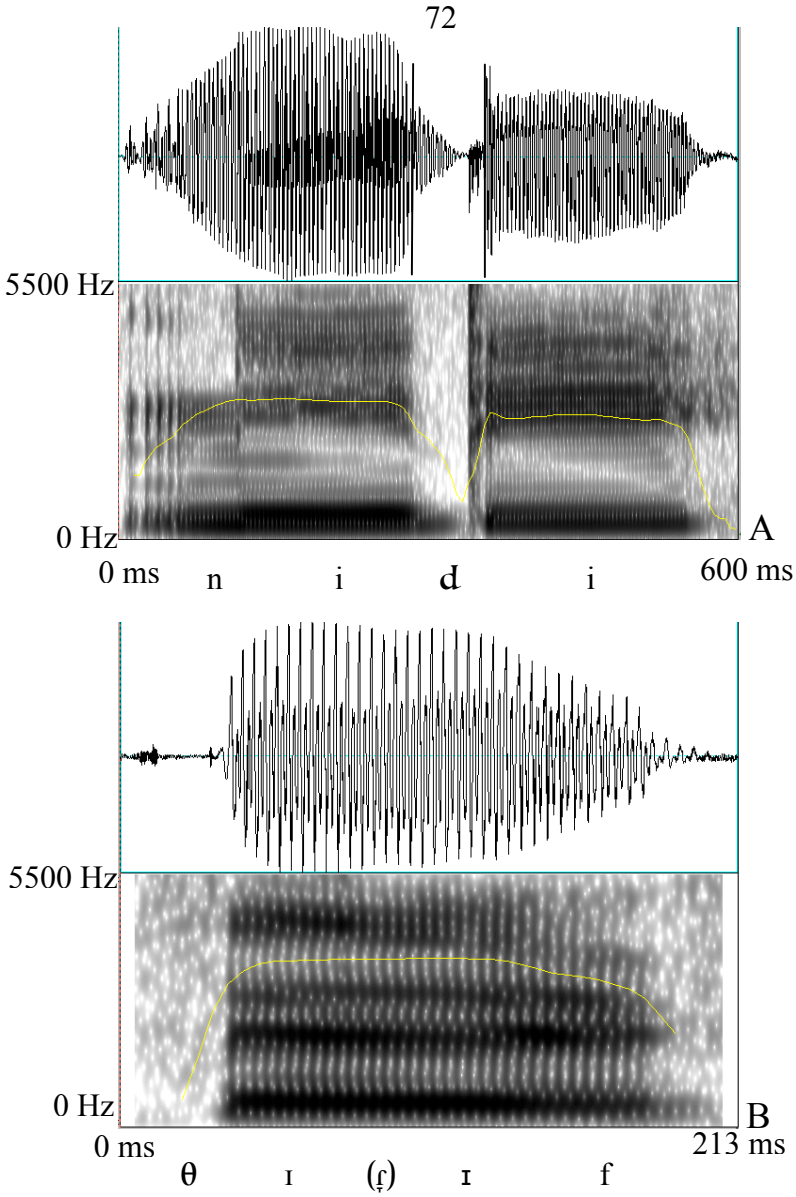


Fig. 3

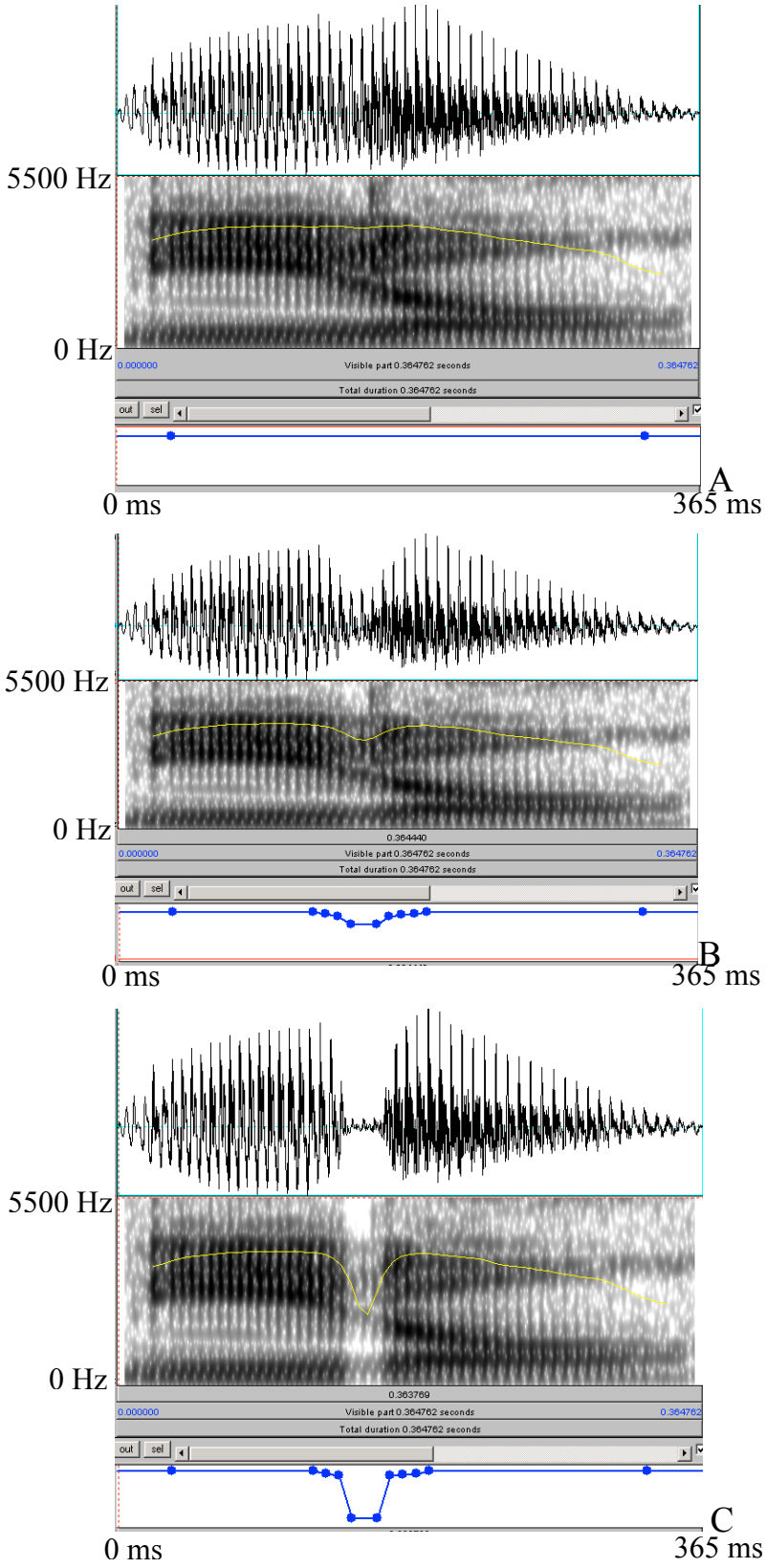
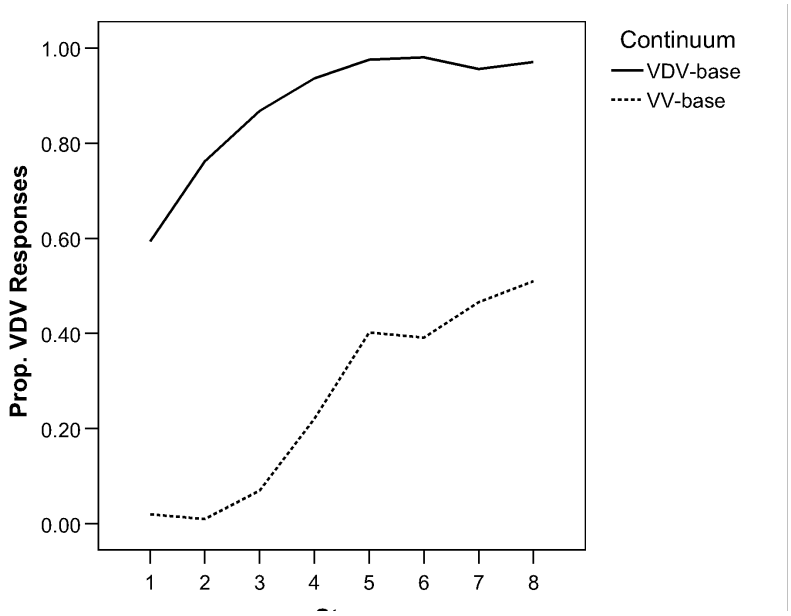
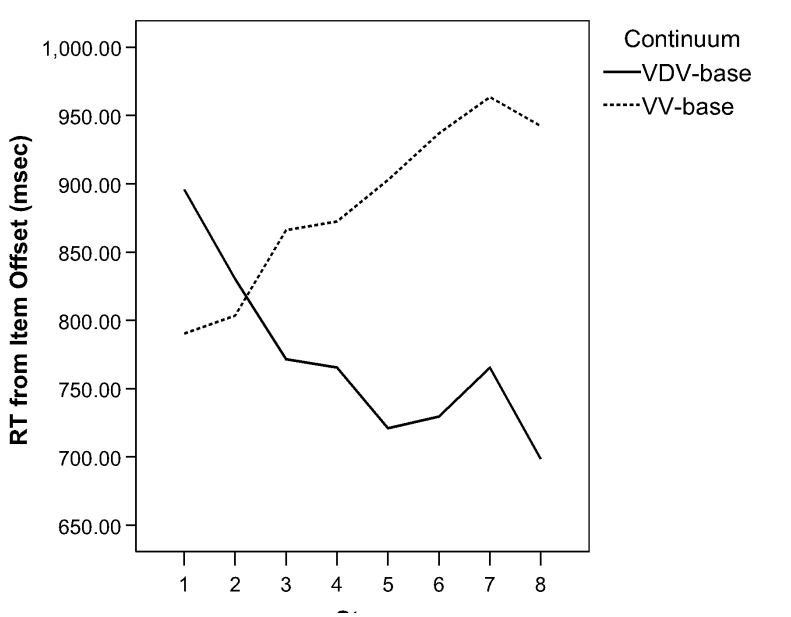

5500 Hz
0 Hz
Visible part 0.364762 seconds
Total duration 0.364762 seconds
out
sel
A
0 ms
365 ms
5500 Hz
0 Hz
0.364440
Visible part 0.364762 seconds
Total duration 0.364762 seconds
out
sel
B
0 ms
365 ms
5500 Hz
0 Hz
0.363769
Visible part 0.364762 seconds
Total duration 0.364762 seconds
out
sel
C
0 ms
365 ms

Fig. 4


Continuum
VDV-base
VV-base
Prop. VDV Responses
1.00
0.80
0.60
0.40
0.20
0.00
1
2
3
4
5
6
7
8

A


Continuum
VDV-base
VV-base
RT from Item Offset (msec)
1,000.00
950.00
900.00
850.00
800.00
750.00
700.00
650.00
1
2
3
4
5
6
7
8

B

Fig. 5

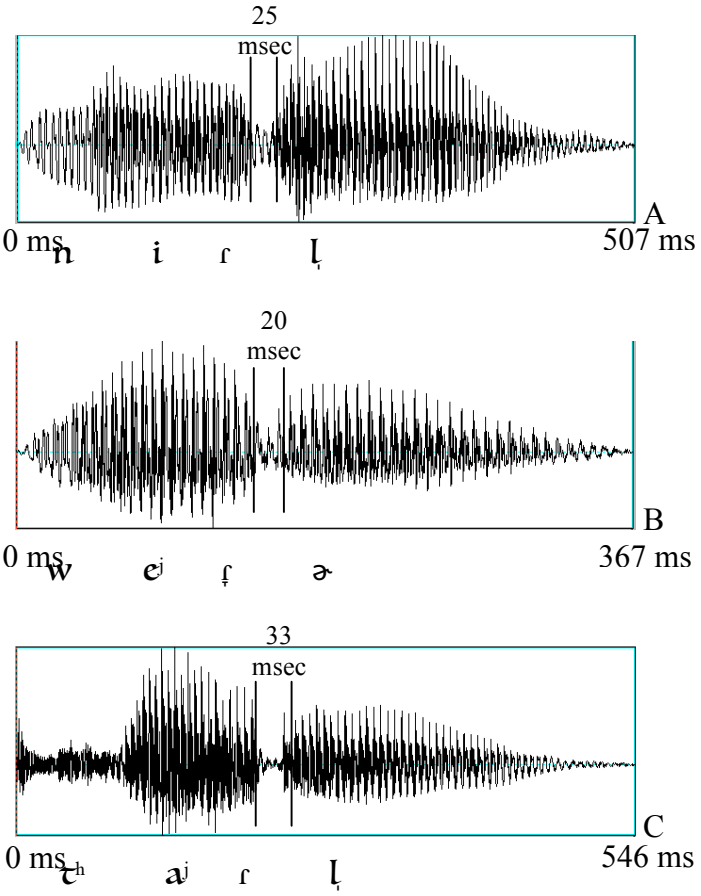


Fig. 6

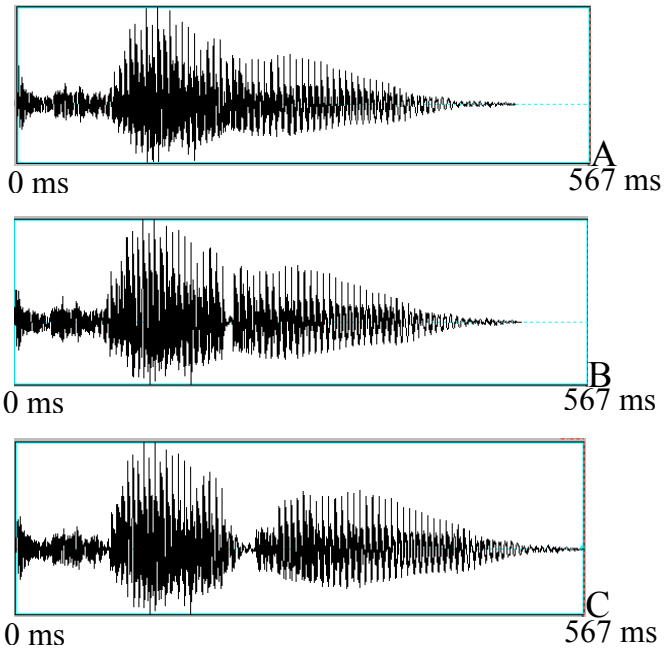
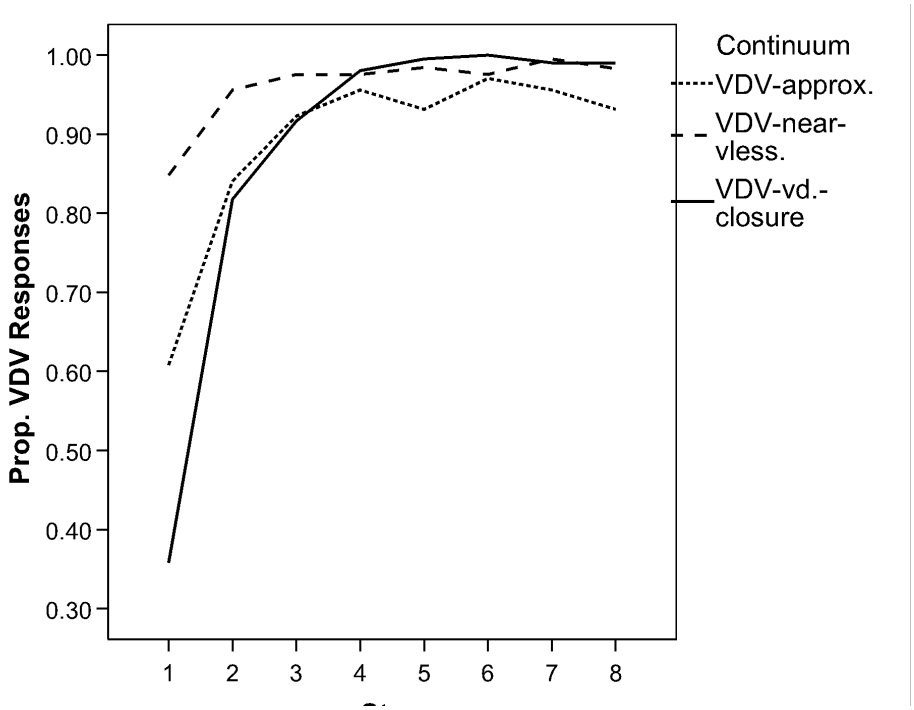
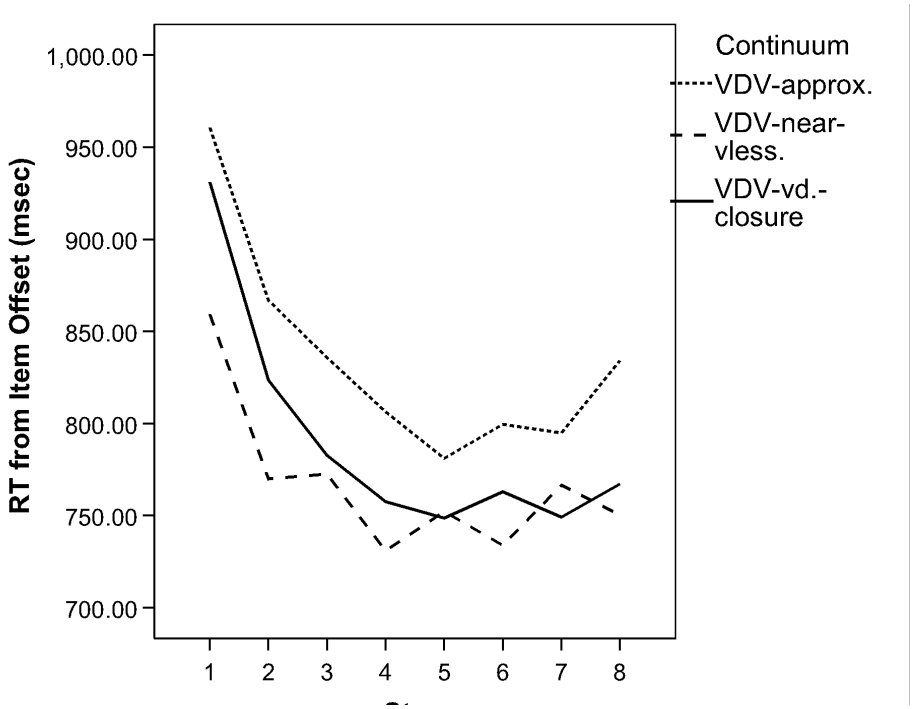


Fig. 7


Prop. VDV Responses
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
1
2
3
4
5
6
7
8
Continuum
VDV-approx.
VDV-near-vless.
VDV-vd.-closure

A


RT from Item Offset (msec)
1,000.00
950.00
900.00
850.00
800.00
750.00
700.00
1
2
3
4
5
6
7
8
Continuum
VDV-approx.
VDV-near-vless.
VDV-vd.-closure

B

Fig. 8

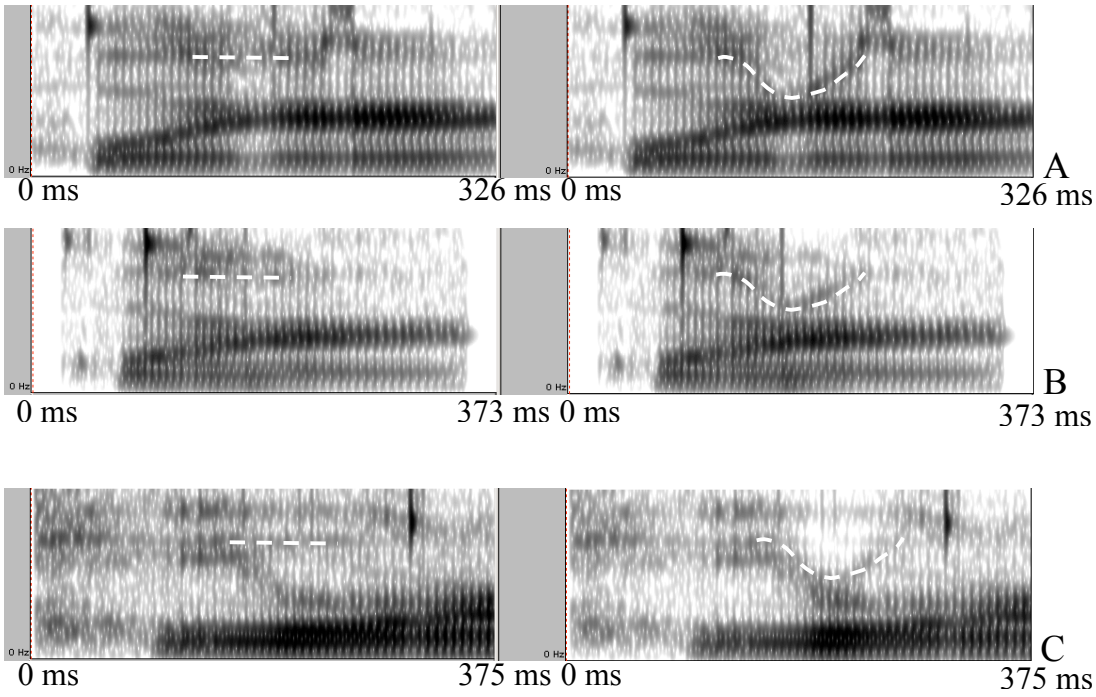
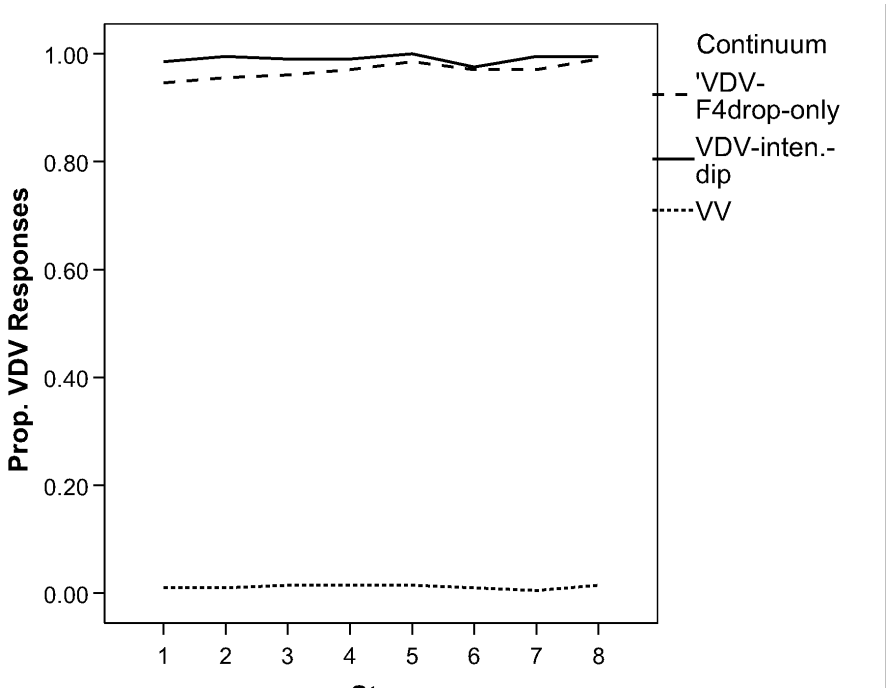
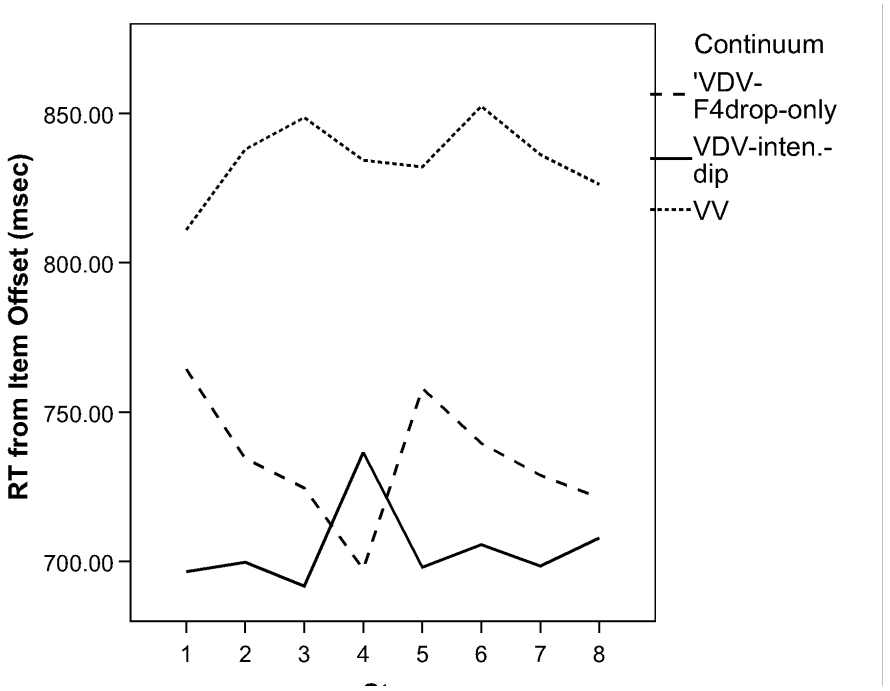

0 ms
326 ms 0 ms
326 ms
A
0 ms
373 ms 0 ms
373 ms
B
0 ms
375 ms 0 ms
375 ms
C

Fig. 9


Prop. VDV Responses
1.00
0.80
0.60
0.40
0.20
0.00
1
2
3
4
5
6
7
8
Continuum
'VDV-F4drop-only
VDV-inten.-dip
VV

A


RT from Item Offset (msec)
850.00
800.00
750.00
700.00
1
2
3
4
5
6
7
8
Continuum
'VDV-F4drop-only
VDV-inten.-dip
VV

B