

強調形に現れる促音と有声性

or

How do you make a voicing
distinction when the constriction
is soooooooooo long?

慶應義塾大学

言語文化研究所

川原繁人

Background

- Maintaining the glottal vibration during stop closure presents an aerodynamic challenge (Ohala 1983 *et seq*).
- Therefore, Tokyo Japanese did not use to have lexical voiced geminates.
- Voiced geminates appear in loanwords, but they nevertheless show “semi-devoicing” (Kawahara 2006, Lg).

Background 2

- But recent studies on non-Tokyo dialects show that:
 - (i) some dialects do show lexical contrasts between voiced and voiceless geminates, and that
 - (ii) some voiced geminates can be fully voiced.
- Recall the talks today by Matsuura and Takada.

Emphatic gemination

- This project investigates another type of geminates which have been understudied.
- Emphatic geminates:
 - su**gggoi**
 - hiddddoi
- Its durational properties have been studied by Kawahara and Braver (2014, JIPA).
- Some speakers can make six-way durational distinctions.

A brief look at Kawahara and Braver (2014)

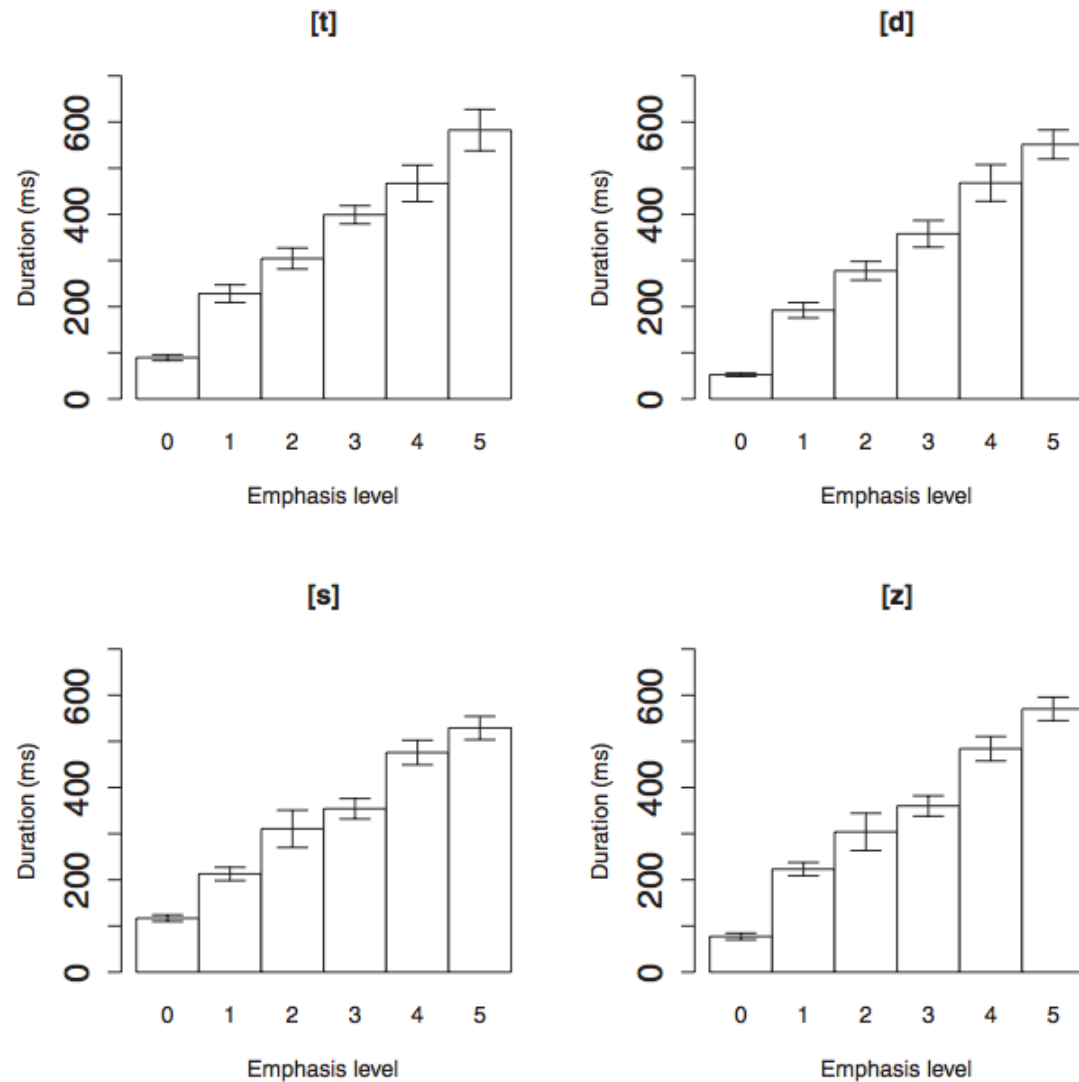


Figure 3 The average durations of each emphasis level with 95% confidence intervals: Speaker FR.

Research question

- How do Japanese speakers make a voicing contrast when the constriction is very long?
- In the longest consonant (ca. 600ms), it is probably *physically* impossible to maintain enough transglottal airpressure drop to maintain voicing.
- But the voicing contrast is still audible.



Method

- A subset of data recorded by K&B (2014).
- Only a stop pair. Five levels of emphasis, plus no emphasis.

– hidoi	ひどい
– hiddoi	ひっどい
– hidddoi	ひっどい
– hidddoi	ひっどい
– hiddddoid	ひっどい
– hiddddoid	ひっどい

The stimuli and the speakers

- The speakers were female undergraduate students, recorded at International Christian University.
- The pairs analyzed (10 reps each):

hidoi	katai
kudoι	itai
- The best three speakers who distinguished durational differences between different emphasis levels.

Measurement and predictions

	Voiceless	Voiced
Closure duration	longer	shorter
Preceding vowel duration	shorter	longer
Closure voicing duration	shorter	longer
F0 at preceding vowel	higher	lower
F0 at the following vowel	higher	lower

Multiplicity of voicing cues (Kawahara 2006; Kingston and Diehl 1994 (Lg); Lisker 1986 (Lg&Sp), a.o.).

(For F1, look at the next slide).

The case of Japanese: the geminate data from Kawahara (2006)

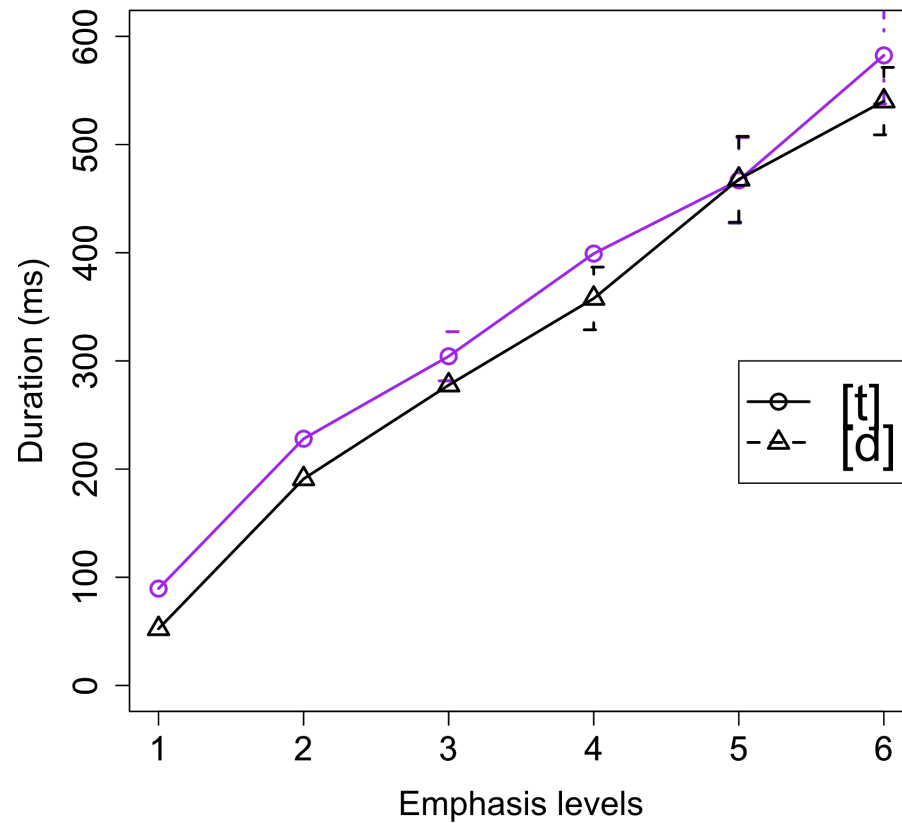
	Voiceless	Voiced
Closure duration	longer (130ms)	shorter (110ms)
Preceding vowel duration	shorter (50ms)	longer (70ms)
Closure voicing duration	shorter (10ms)	longer (40ms)
F0 at preceding vowel	higher (280Hz)	lower (260Hz)
F0 at the following vowel	higher (300Hz)	lower (290Hz)

Measurements automatically extracted using a Praat script (available upon request). F1 not measured because vowel quality is not controlled.

Voicing automatically detected from the “voice report” function in Praat, which itself is based on the presence of pulse.

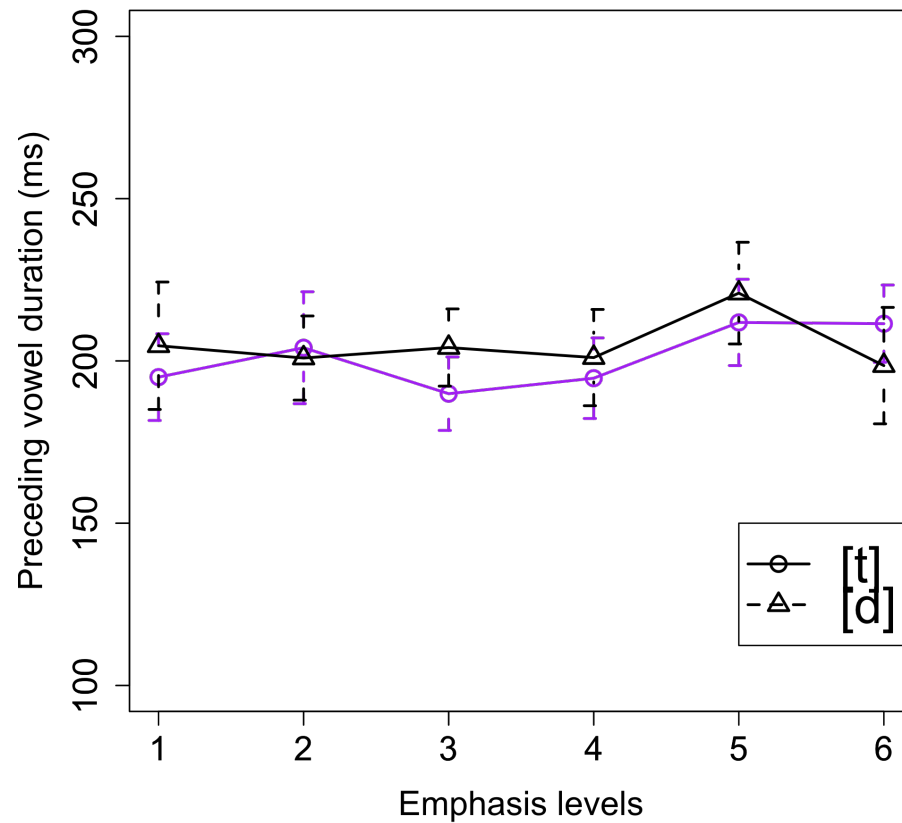
Speaker FR

Closure duration



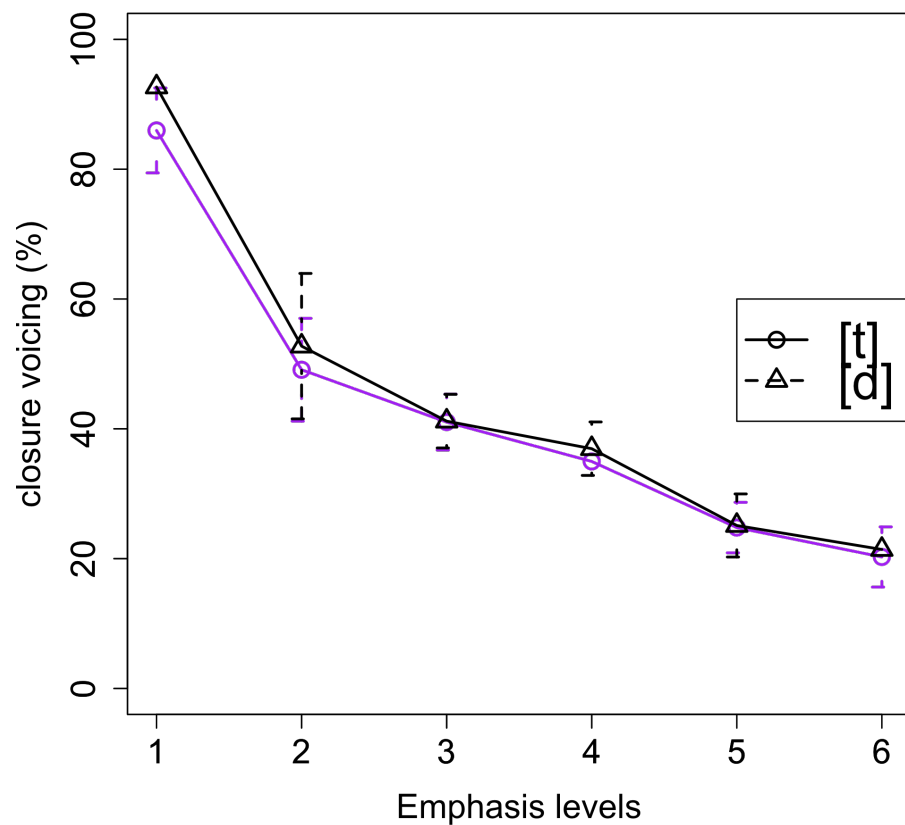
[t] is longer than [d]
up until emphasis
level 4.

V1 duration



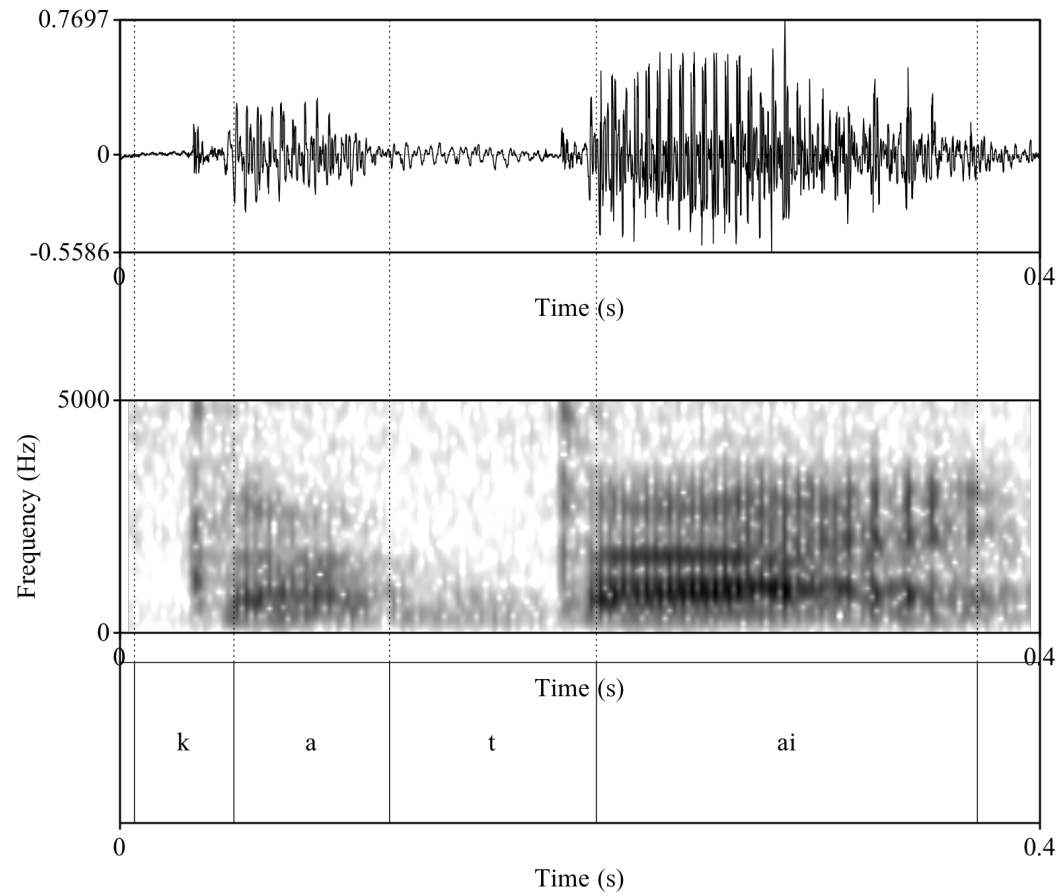
V1 duration is slightly longer before [d] than before [t], but not in all conditions.

Closure voicing %

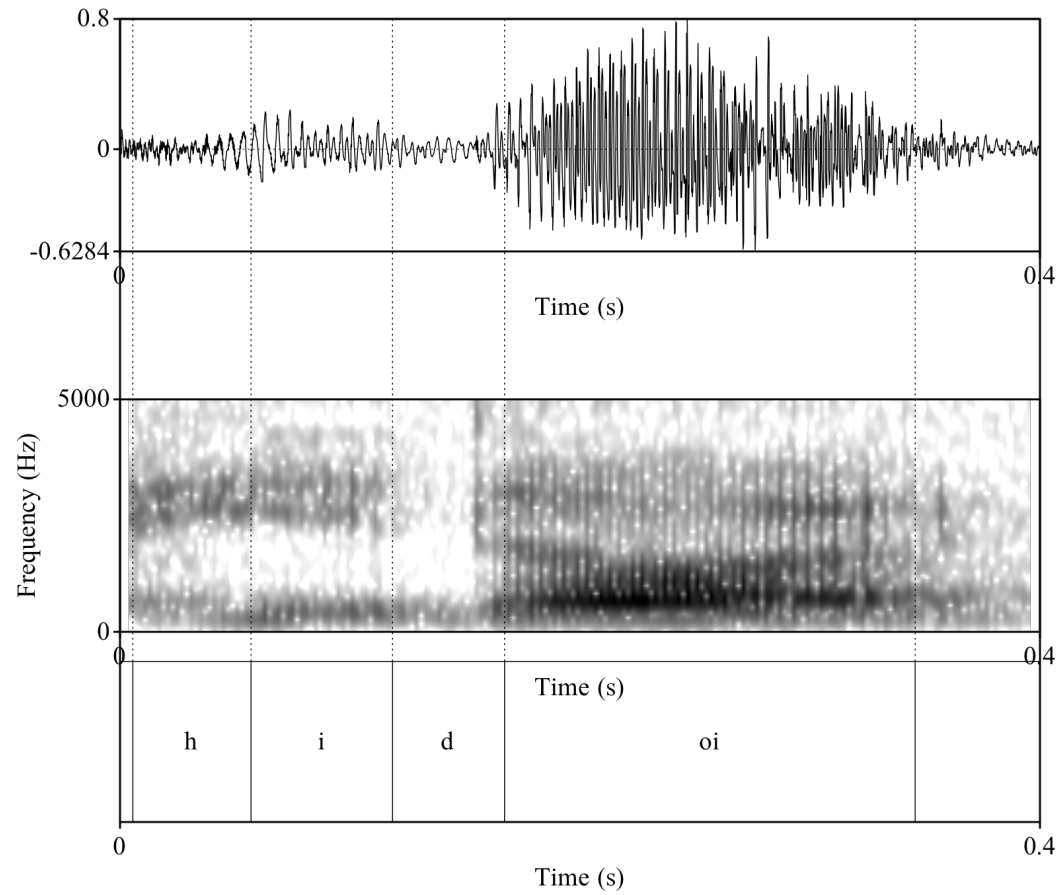


[t] and [d] do not differ much in terms of closure voicing percentages in all the conditions. (???)

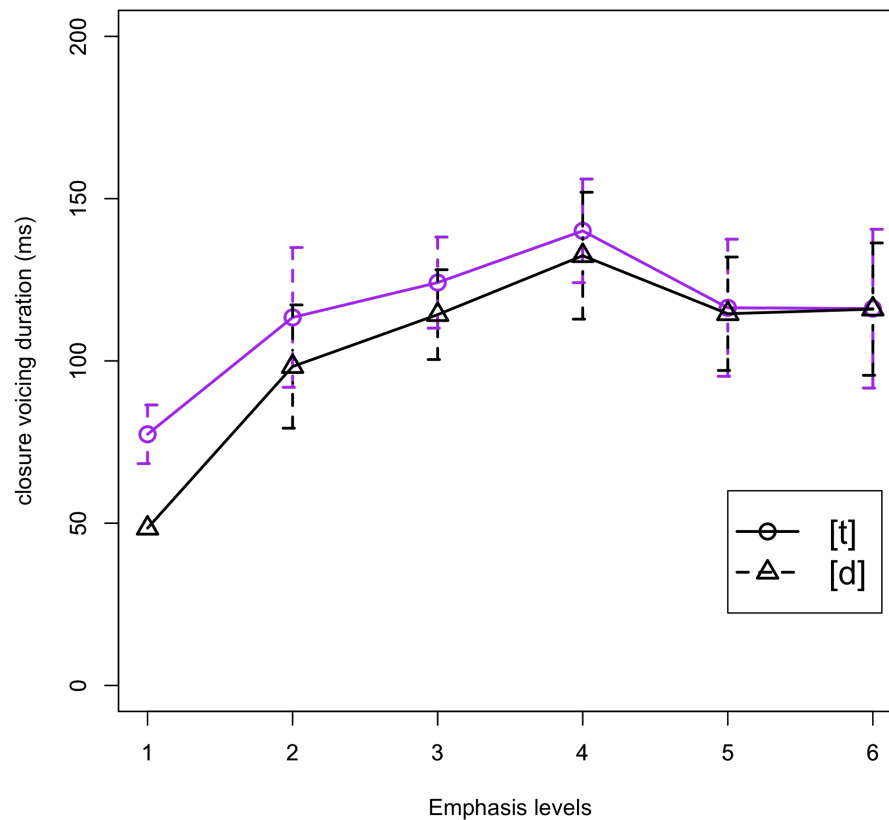
[t] with full voicing (?)



[d] for comparison



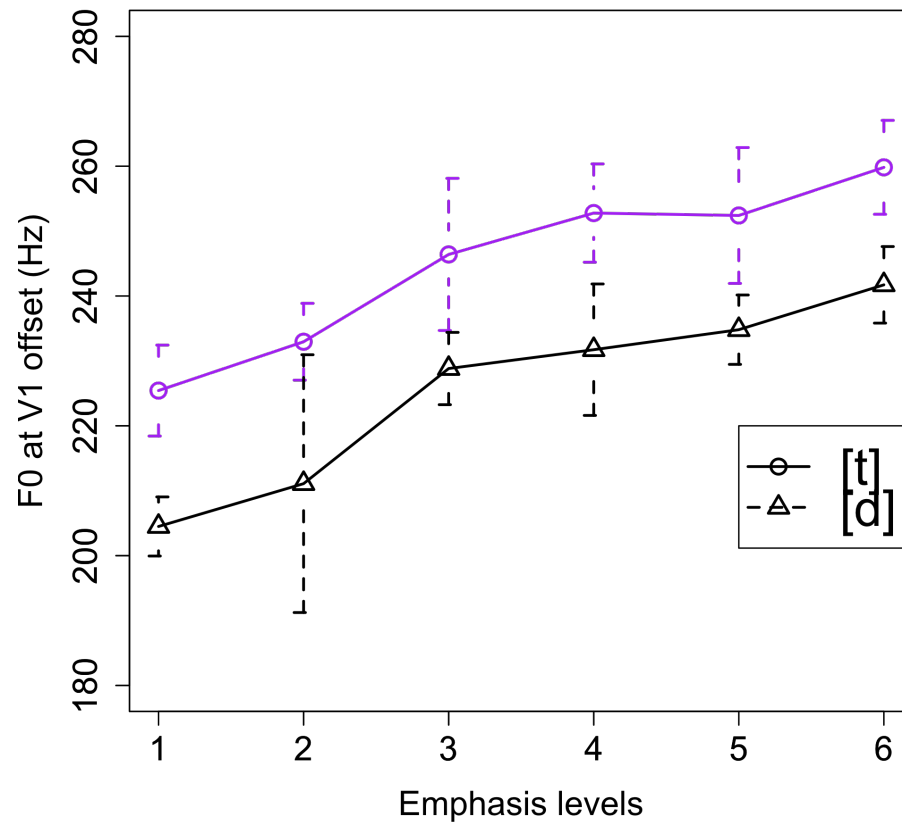
Absolute closure voicing duration



More voicing during [t] except for the last two conditions. (???)

Let's bear in mind though that [t]'s closure is longer.

F0 at the offset of the preceding vowel

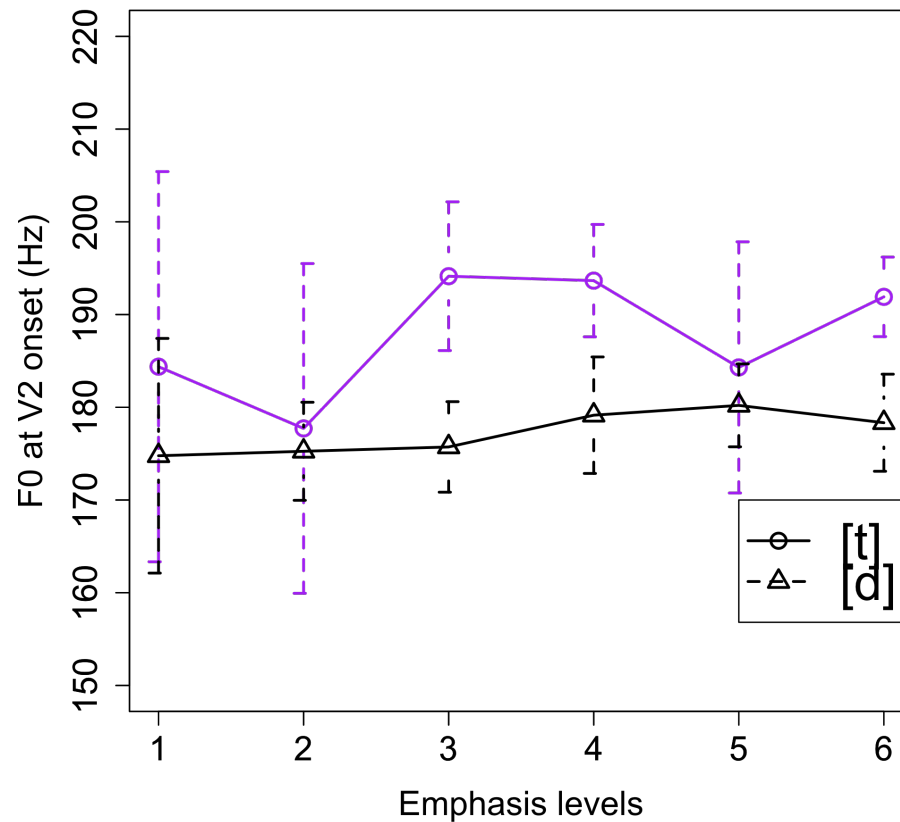


F0 is systematically higher before [t] than before [d] by ca. 20Hz.

F0 also becomes higher as the emphasis levels goes higher.

[i, u] before [d], and [i, a] before [t]—we would expect the former to have higher F0 due to the intrinsic vowel effect (Whalen & Levitt 1995, JPhon).

F0 at the onset of the following vowel



F0 is higher after [t] than after [d].

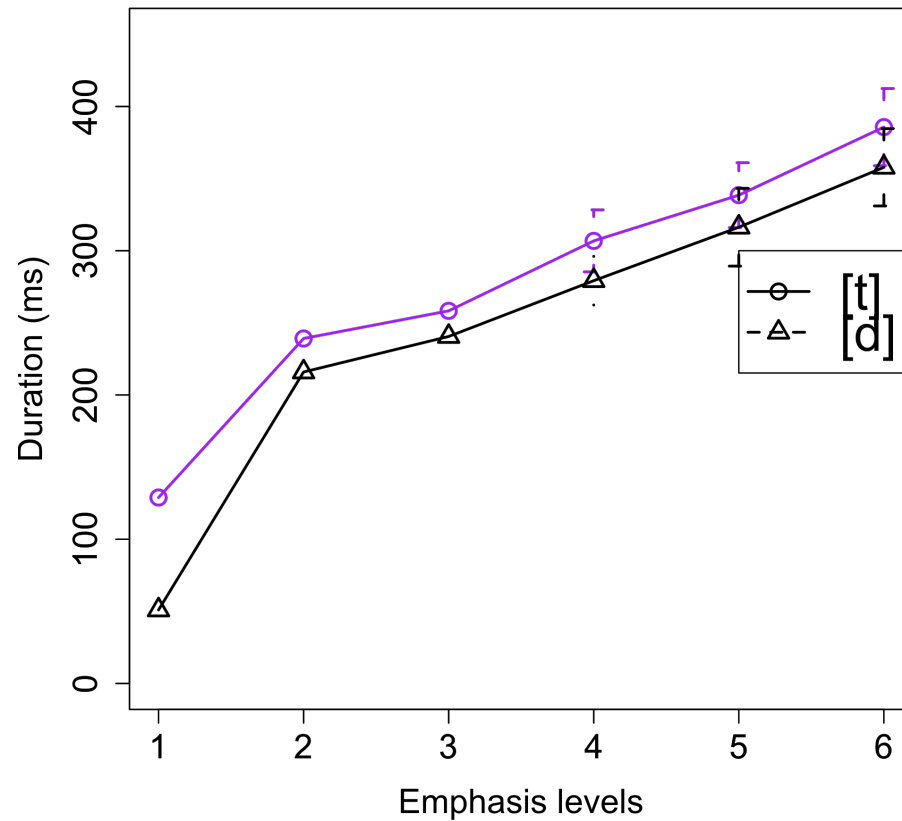
[o] after [d] and [a] after [t], so we would expect the former to have higher f0 due to the intrinsic vowel effect.

Summary of Speaker FR

- For this speaker, F0 in the surrounding vowels seems to be the most consistent acoustic correlate for all the vowels.
- There was even a reversal in terms of absolute closure voicing duration.
- Strength, not mere its duration, of closure voicing? What really is the voicing during [t]?

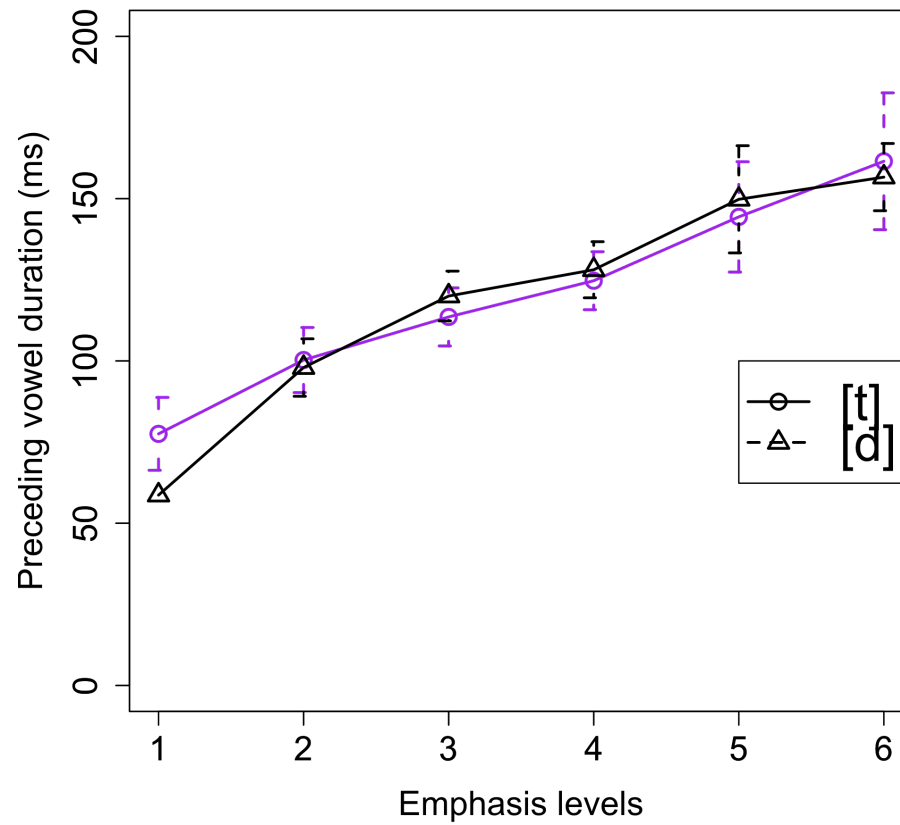
Speaker EL

Closure duration



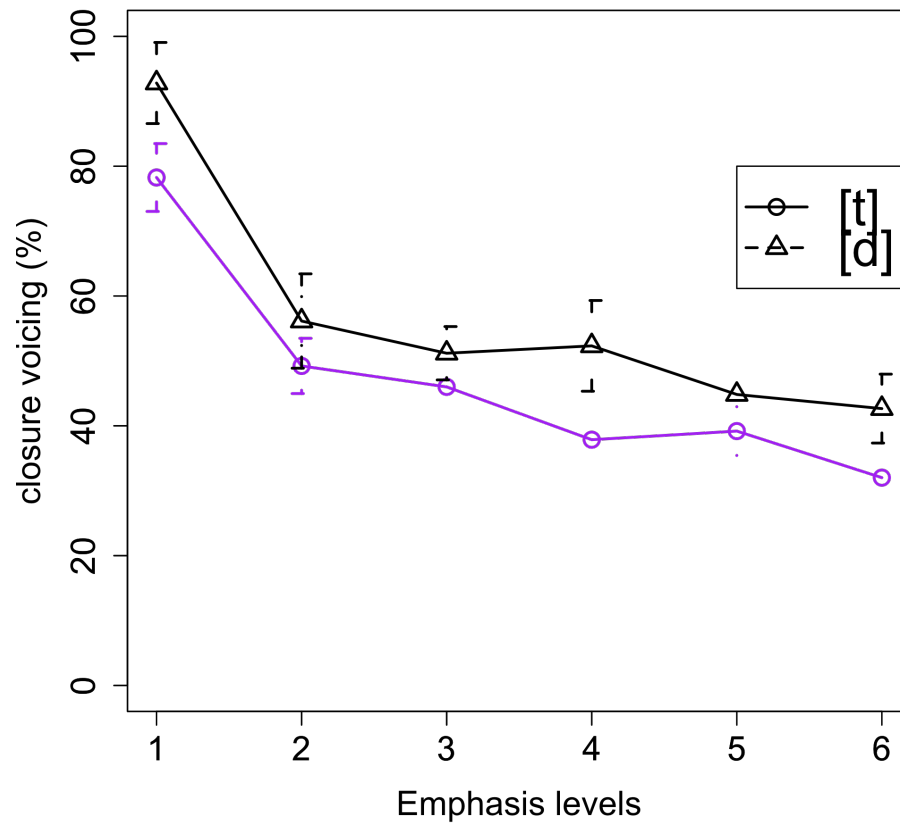
[t] is longer than [d]
for all the emphasis
levels.

V1 duration



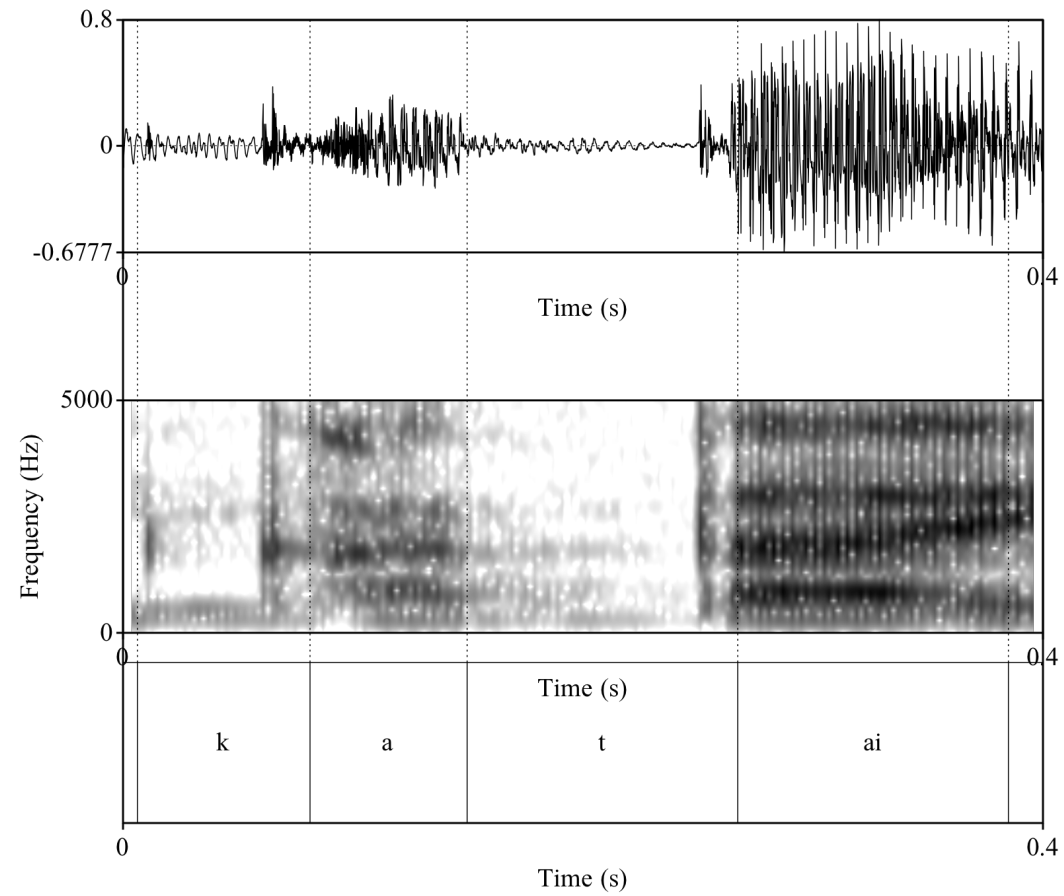
No clear distinction,
except at the non-
emphatic condition.

Closure voicing %

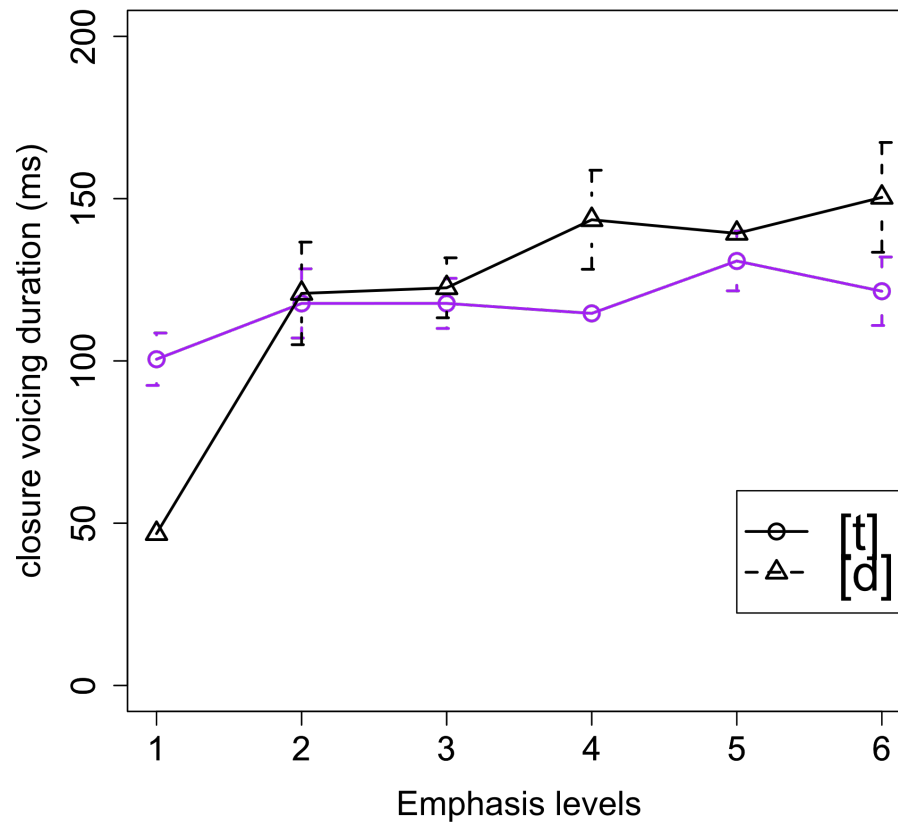


Much clearer
difference in terms
of closure voicing %
at all levels,
compared to
Speaker FR.

[t] with closure voicing again

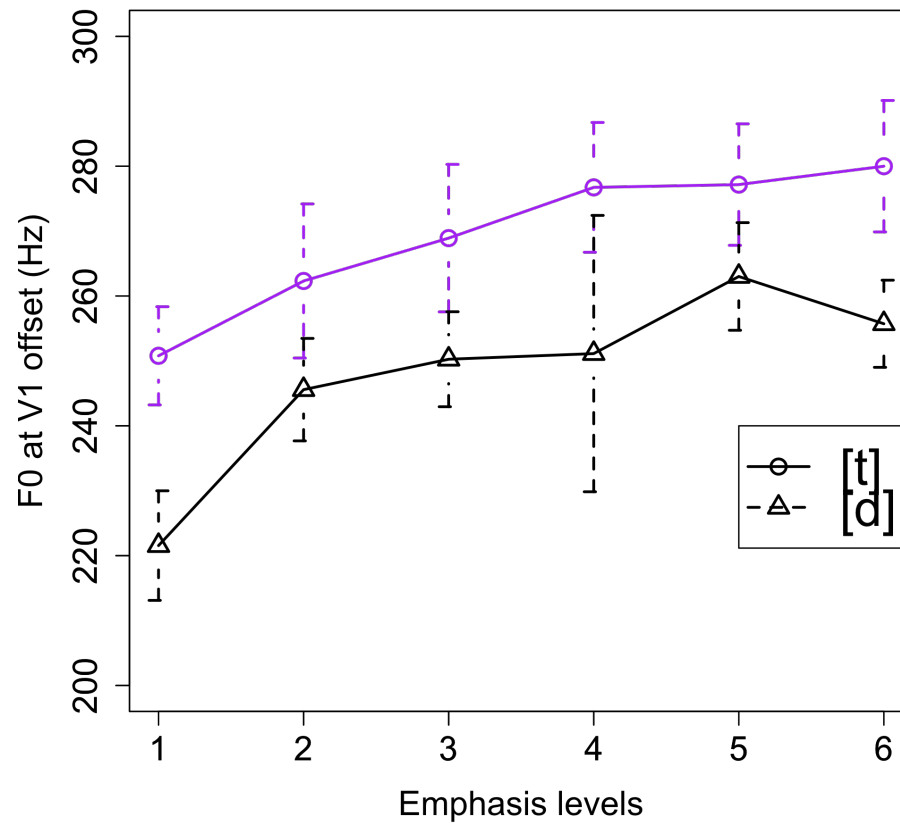


Absolute closure voicing duration



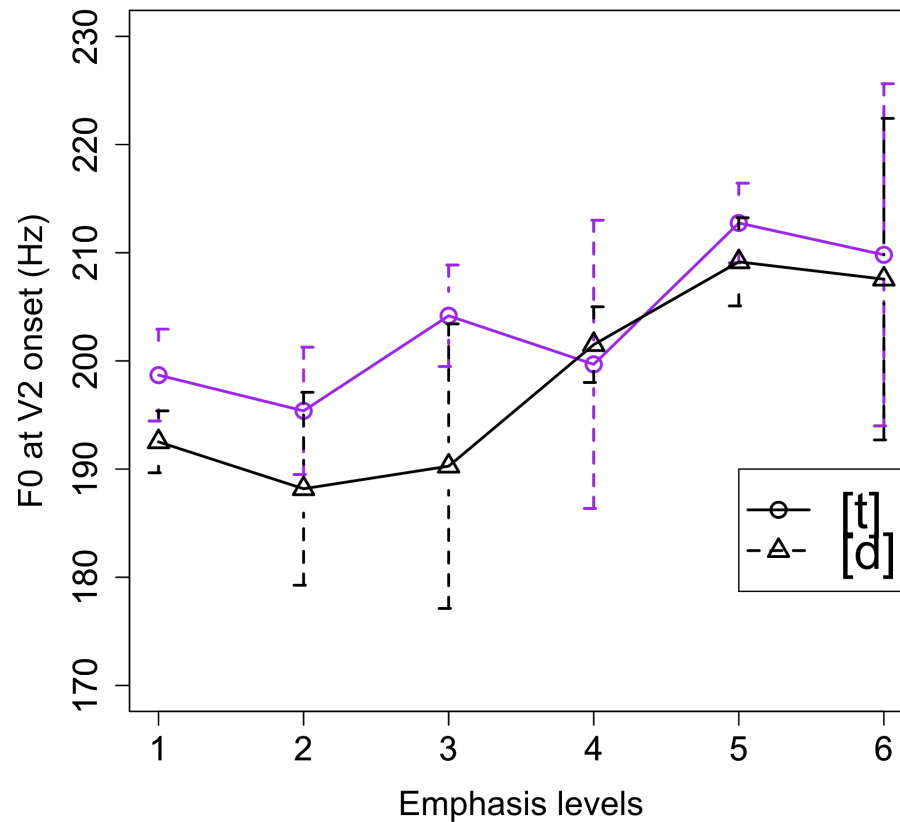
More voicing during [d] except for the non-emphatic condition (but recall that [t] is longer in the first place).

F0 at the offset of the preceding vowel



F0 is higher before [t] than before [d]. The separation between the two conditions is very clear.

F0 at the onset of the following vowel



F0 is generally higher after [t] than after [d].

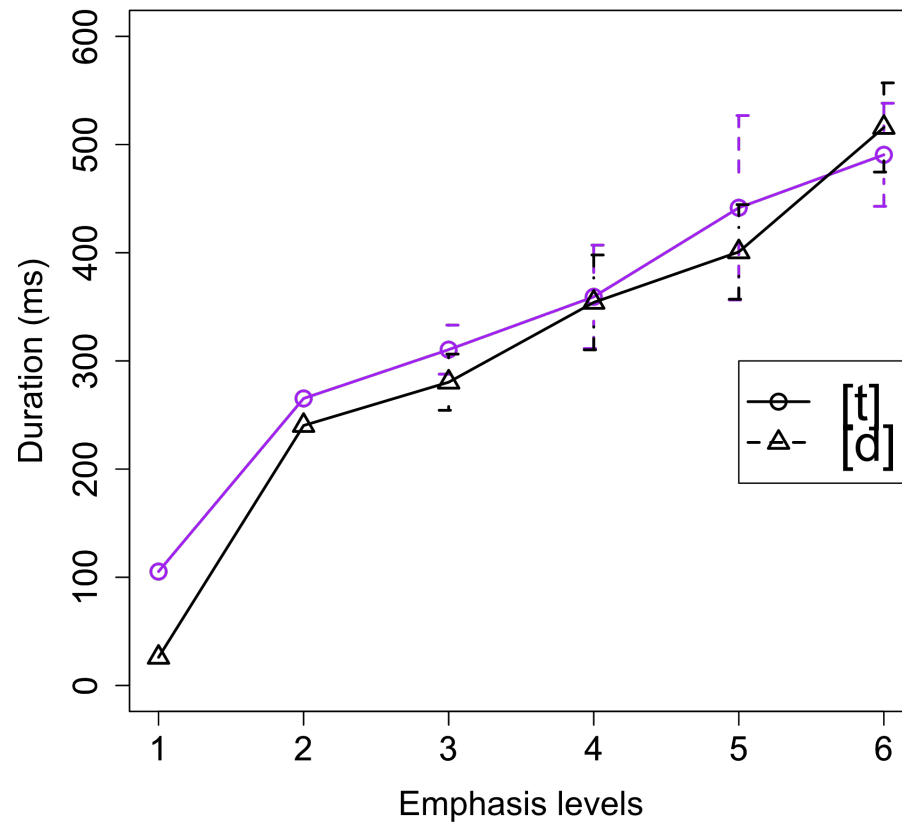
F0 also generally becomes higher as the emphasis levels goes higher.

Summary of Speaker EL

- Clear differences in terms of closure voicing duration, closure duration, and the following F0.
- The two speakers discussed so far make use of a different (yet overlapping) set of acoustic cues to convey a voicing contrast in this experiment on emphatic forms.

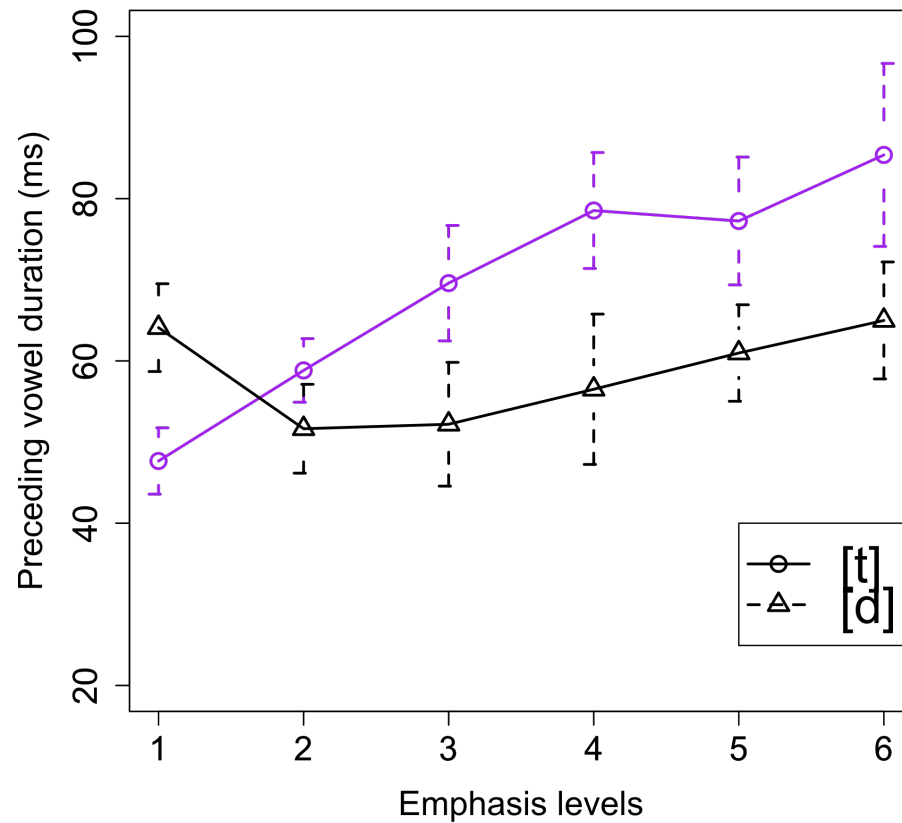
Speaker TW

Closure duration



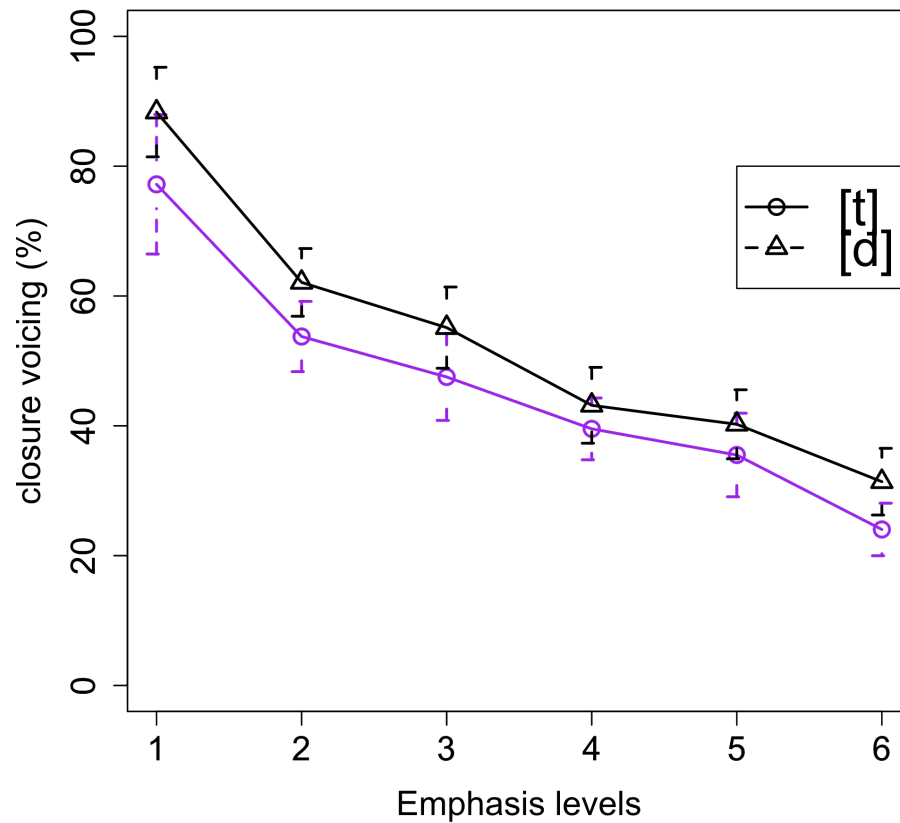
[t] is longer than [d]
for the non-
emphatic and the
first two emphatic
levels.

V1 duration



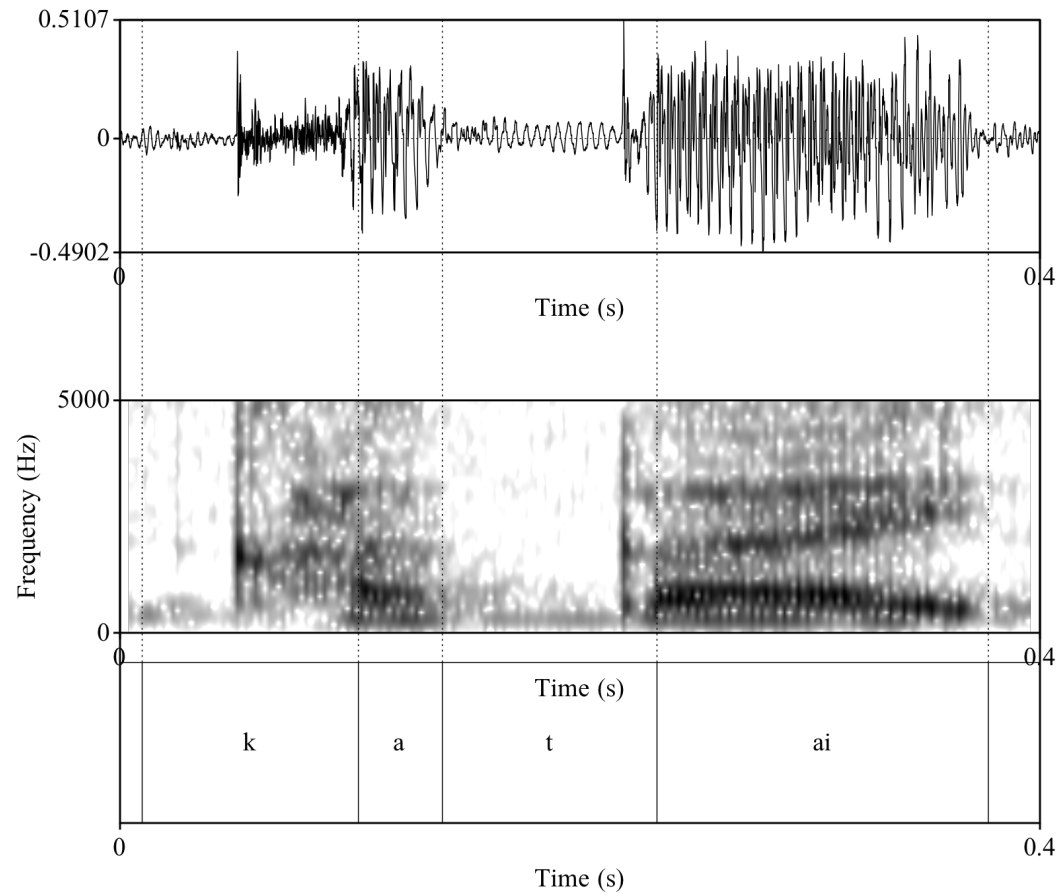
Reversal: V1 is longer before [t] than before [d] (!); expected relation in the non-emphatic condition.

Closure voicing %

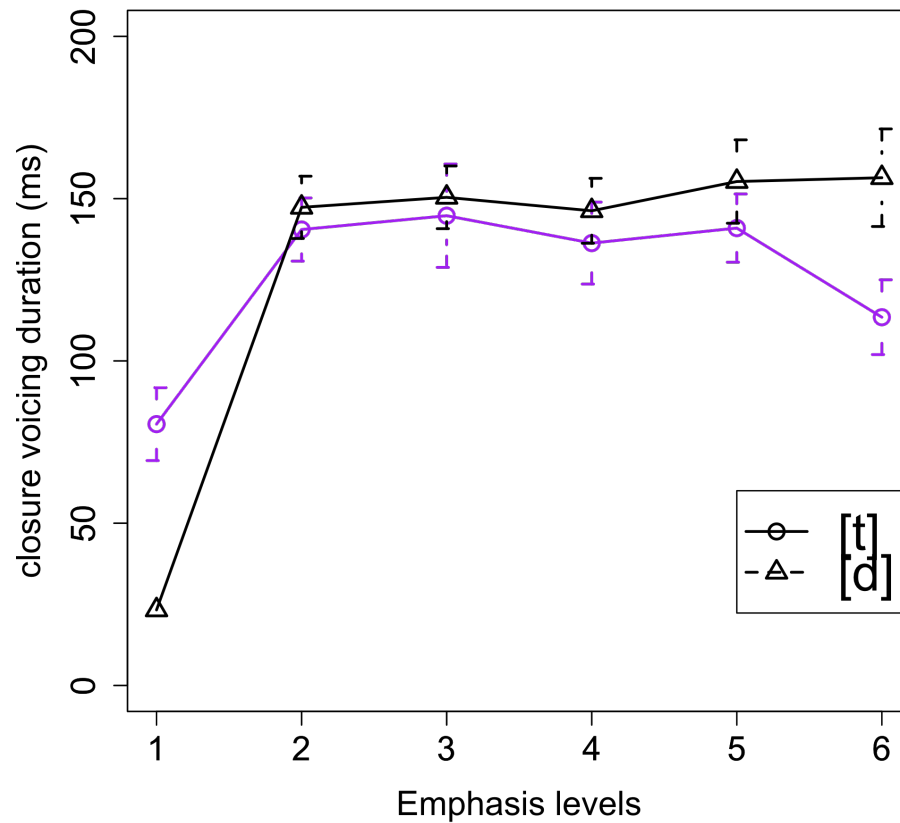


Much clearer
difference in terms
of closure voicing %
at all levels, than
Speaker FR.

[t] with voicing once again

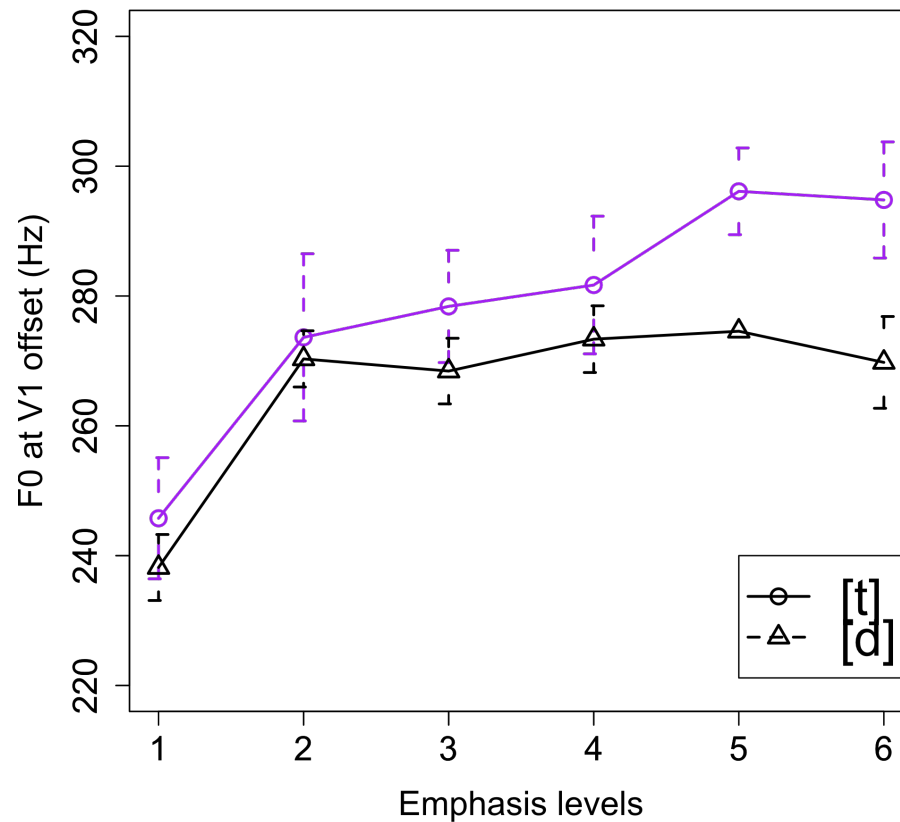


Absolute closure voicing duration



More voicing during [d] except for the non-emphatic condition.

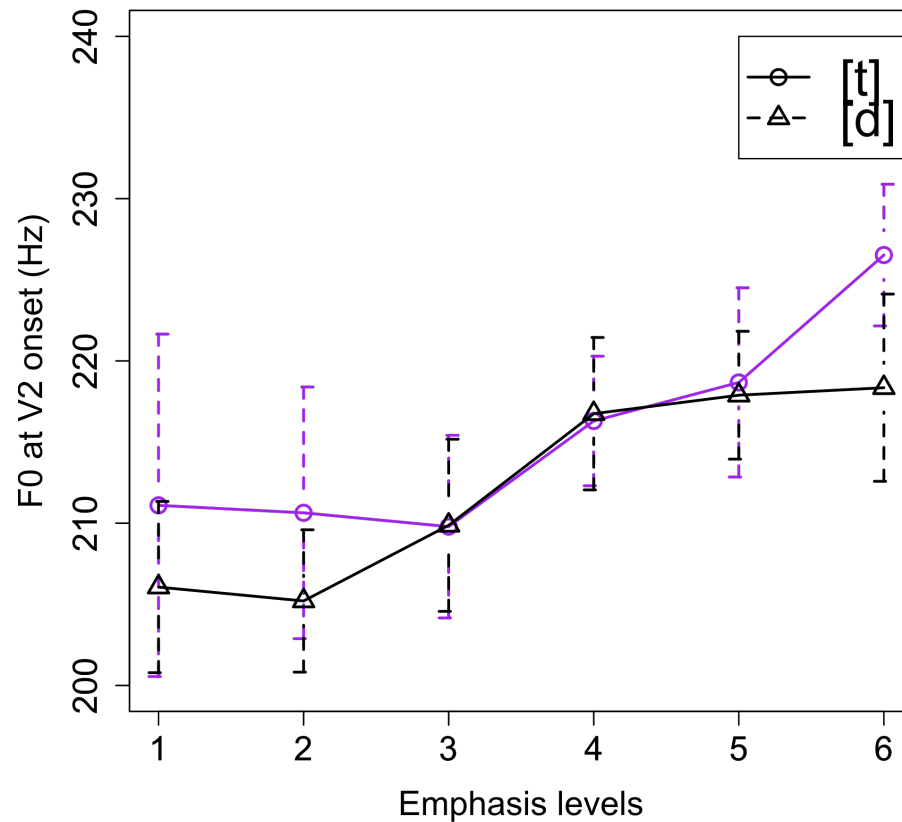
F0 at the offset of the preceding vowel



F0 is higher before [t] than before [d].

Very large differences in the last two emphatic conditions.

F0 at the onset of the following vowel



NB: There were some outliers.

F0 is generally higher after [t] than after [d] in the first two conditions.

F0 also generally becomes higher as the emphasis levels goes higher.

Summary

	Speaker FR	Speaker EL	Speaker TW
closure duration	✓	✓	(✓)
preceding vowel duration	(✓)	X	(✓)
closure voicing (%)	X	✓	✓
closure voicing (absolute)	X	(✓)	✓
F0 at the preceding vowel offset	✓	✓	✓
F0 at the following vowel	✓	(✓)	(✓)

Discussion

- A voicing contrast is realized in various acoustic dimensions (Kawahara 2006; Kingston and Diehl 1994; Lisker 1986).
- The current analysis supports this view.
- Moreover, different speakers seem to deploy different subsets of cues (cf. English /r/).

- Why was [t] so “voiced”? (emphasis-specific phonetics? voicing continuation via reverberation?)
- Speaker FR does not show clear distinctions in terms of closure voicing duration.
- Its perceptual impact? (see Matui’s talk today)
- The role of F0 in voicing perception (Fujimura 1971; Holt et al. 2001, JASA?)

Discussion

- F0 differences at V2 onset are particularly clear and consistent across the three speakers.
- F0 differences due to [voice] differences are not automatic, but instead controlled (see esp. Speaker TW).
- Speakers consciously control their articulation, with explicit “phonetic knowledge” (Kingston & Diehl 1994; Keating 1986, 1988)

An addendum with a bit of self-promotion

- English speakers, who do not have a length contrast in their native language, can make similar durational distinctions.
- The full poster presented at Phonology 2014 at MIT:

<http://user.keio.ac.jp/~kawahara/pdf/EngLengtheningAMP2014.pdf>

A brief look

