

# Top-down effects on compensation for coarticulation are not replicable

Holger Mitterer

Max-Planck-Institut für Psycholinguistik Nijmegen

holger (dot) mitterer (at) mpi (dot) nl

## Abstract

Listeners use lexical knowledge to judge what speech sounds they heard. I investigated whether such lexical influences are truly top-down or just reflect a merging of perceptual and lexical constraints. This is achieved by testing whether the lexically determined identity of a phone exerts the appropriate context effects on surrounding phones. The current investigations focuses on compensation for coarticulation in vowel-fricative sequences, where the presence of a rounded vowel (/y/ rather than /i/) leads fricatives to be perceived as /s/ rather than /ʃ/. This results was consistently found in all three experiments. A vowel was also more likely to be perceived as rounded /y/ if that lead listeners to be perceive words rather than nonwords (Dutch: *meny*, English *id.* vs. *meni* nonword). This lexical influence on the perception of the vowel had, however, no consistent influence on the perception of following fricative.

**Index Terms:** human speech perception

## 1. Introduction

One of the fundamental issues in cognitive science is how perception is influenced by conceptual knowledge. In the field of human speech perception, the debate focused on the relation of lexical and phonological—perceptual—processing.

There are several clear demonstration that the lexicon influences how a given phone is perceived. Warren [19] showed that listeners restore noise-masked phonemes in long words such as *legi\*lature*, in which an effectively inaudible /s/ is nevertheless perceived as present. Ganong [2] showed a lexical influence in a categorization task. If participants have to judge whether an initial stop is either /t/ or /d/, they respond with /t/ more often if the stop is followed by *...ype* (so that *type* is a word and *dype* not), than when the stop is followed by *...ice*. In the latter case, interpreting the stop as /d/ gives rise to a word.

Norris, McQueen, and Cutler [13] showed that such effects can be interpreted as joint influences of perceptual and lexical information on the interpretation of the incoming signal. The assumption that the lexical information is fed back in a top-down fashion to a phonological level is not necessary. The best way to test for true top-down effects is by showing an *indirect* influence of the lexical knowledge.

One of the benchmarks is whether the lexically influenced identity of phone gives rise to the same context effect than the *real* thing—an unambiguous phoneme in a nonword context—itsself. Previous investigations focused on fricative-stop sequences. Mann and Repp [8] had shown that unambiguous /ʃ/ induced a /t/-bias for a following ambiguous stop between /t/ and /k/, while a preceding /s/ induces a /k/-bias. This reflects compensation for coarticulation, because an /t/ will surface somewhat /k/ like if preceded by /ʃ/. The crucial question now is what happens if the fricative is

ambiguous and the lexicon determines its perceived identity. An ambiguous fricative is likely to be perceived as /ʃ/ at the end of *fooli...* but as /s/ in *christma...*, because Christmas and foolish are words and Christmash and foolis are not. The question then is whether a stop following these words is more likely to be perceived as /k/ after *christmaX* than after *fooliX*, in which *X* denotes an ambiguous fricative. Elman and McClelland [1] found just this result and interpreted it as evidence for a true top-down effect. The lexically restored identity of the fricative as /s/ or /ʃ/ exerts just the same compensation-for-coarticulation effect as "real" /s/ or /ʃ/ do. However, subsequent investigations of this effect offered alternative interpretations and sometimes failed to replicate the effect [14, 15], although there are also successful replications [7] of lexically mediated compensation.

One possible reason for this unclear picture might be that compensation for coarticulation in fricative-stop sequences might not be the best place to look for lexical involvement. This context effect may arise on a pre-linguistic, auditory level of processing, which may preclude an influence of lexical processing [4-6]. Moreover, even visual influences, arguably more signal-based than lexical influences, do not affect compensation for coarticulation in this case. Vroomen [18] showed that the perception of an ambiguous fricatives can be biased towards /s/ or /ʃ/ by showing a face mouthing /s/ or /ʃ/, respectively. This direct influence of the visual speech gesture did not influence the perception of following fricative, as acoustic /s/ and /ʃ/ did.

Recently, I have shown that visual influences on the perception of a phoneme can nevertheless lead to compensation effects on surrounding phonemes [12]. These experiments first replicated the compensation effect in fricative-vowel syllables first established by Mann and Repp [9]: A fricative is more likely to be perceived as /s/ rather than /ʃ/ if followed by a rounded vowel (/y/) than if followed by an unrounded vowel (/i/). This context effect compensates for coarticulation in production: A fricative is produced with a lower fricative pole if adjacent to rounded vowel than if adjacent to an unrounded vowel. Hence, /s/ is somewhat /ʃ/-like if adjacent to rounded /y/, and listeners compensate for this. This context effect can also be triggered visually: In an experiment with audiovisual speech, an acoustically ambiguous vowel was not only perceived more often as rounded /y/ if accompanied by a face mouthing /y/, but also lead a preceding fricative to be perceived as /s/, just as an acoustically specified /y/ did [9]. Based on these results, I argued that context effects in speech perception may arise at different auditory and phonological processing levels. In the fricative-stop case, the data point towards an auditory level, while for the fricative-vowel case, the data point towards a phonological level.

A corollary of these findings is that lexical influences on compensation for coarticulation, just as visual influences, may arise only for certain compensation effects. Therefore,

three experiments tested whether lexical effects on compensation reliably occur in the fricative-vowel case. In order to maximize the lexical influence on the vowel, I changed the order to vowel-fricative sequences, which does not alter the nature of the context effect [10], but maximizes the lexical influence on vowel identification.

## 2. Experiments

In all experiments, participants heard a three-syllable sequence. The first two syllables were either [ʒen(i-y)] or [men(i-y)], with a last vowel from a semi-synthetic (i-y) continuum. In Dutch, [ʒeni] and [meny] are words (Engl. *genius* and *menu*) while [ʒeny] and [meni] are nonwords. The [men...] context should hence give rise to a /y/-bias and the [ʒen...] context to an /i/-bias. The third and last syllable of the stimuli started with a token from a /s/-/ʃ/ continuum. The remainder of the syllable in Experiments 1 and 2 was [...op], so that both interpretations of the fricative lead to a Dutch word (/sɔp/ *soap solution* and /ʃɔp/, *id.*). In Experiment 3, the remainder of the third syllable was [...aft], which lead to a nonword with both an /s/ and /ʃ/ interpretation of the initial fricative. The task of the listeners was to identify the sequence at the word boundary as either /i,ʃ/, /i,s/, /y,ʃ/, or /y,s/, by pressing on of four buttons labeled with the Dutch orthographic transcriptions "ie..s", "ie..sj", "u..s", and "u..sj". These responses were then again transformed in proportion /i/- and /y/-responses, by adding the proportions of "ie..s" and "ie..sj" responses on the one hand and the proportion of "u..s", and "u..sj" responses on the other. The response proportions for fricatives were derived accordingly.

### 2.1. Experiment 1

#### 2.1.1. Participants

Sixteen native listeners of Dutch participated for a small monetary compensation. All listeners reported to be a free of (a history) of hearing impairments.

#### 2.1.2. Stimuli

A male native speaker of Dutch was recorded saying all sequences of words and nonwords needed for the experiment. The last vowel of the first word and the initial fricative of the second were spliced out and replaced by a semi-synthetic vowel, based on the LPC-based estimate of the speaker's source, and a synthetic fricative. The vowels had the following formants (bandwidth in brackets) in Hz: F1 350 (100), F2 1900 (150), F3 2250-2750 (200), F4 3450 (300), F5 6000 (400). Seven vowel tokens were generated with equal bark intervals between the steps from the minimum of 2250 and the maximum of 2750 Hz, based on the measures F3 from this speaker for /y/ (2100-2300 Hz) and /i/ (2700-2800 Hz)

The fricatives were generated from a white-noise source with two formant filters, one fixed at 6 kHz (1kHz) and one ranging from 2890 Hz to 3500 Hz (0.5kHz), in seven stimuli separated by six equal steps in bark. We call this formant the Fricative Pole, in line with earlier work [16, 17].

#### 2.1.3. Results and Discussion

Figure 1 shows the results of Experiment 1 in the top row. Beside the simple signal-driven effects (higher F3 → more /i/

responses, higher fricative pole → more /s/ responses), there are three non-trivial effects. (All reported effects are significant at a  $p < 0.05$  level.) The vowel at the end of the first word is influenced by the lexical bias as predicted ([men...] → /y/-bias, [ʒen...] → /i/-bias). Secondly, we also see that in the vicinity of a lower F3, indicative of a rounded vowel, listeners are more likely to perceive the fricatives as /s/. This replicates the phonological, signal-based pattern of compensation of coarticulation in fricative-vowel sequences [9, 12, 16].

Most importantly, the lexical influence on vowel identification, visible in the upper left panel of Figure 1, also has repercussions for the perception of the following fricative, visible in the right upper panel of Figure 1. One can see that listeners give more /s/ responses in the /y/-biasing [men...]. This can be explained by the following causal chain. The lexicon influences the perception of the vowel on a pre-lexical level. On this level, the lexically top-down altered percept of the vowel has the same consequence on the perception of following fricative as a vowel that is perceived as /y/ due to its bottom-up signal properties: In the vicinity of /y/, a fricative is more likely to be perceived as /s/ rather than /ʃ/.

There is one caveat, the overall bias of listeners to perceive nearly all of the presented vowels as /y/ (78%), and listeners hardly respond to changes in F3 in the vowel. Although the influence is overall significant, there is only a slight increase in /i/ judgments as F3 rises (upper left panel of Figure 1). Hence, we altered the procedure in Experiment 2 to get a more balanced number of /i/ and /y/ responses, and clearer endpoint classifications of the vowels as /i/ and /y/.

### 2.2. Experiment 2

In this experiment, we pre-tested participants on the perception of the vowel continuum. Pre-tested started on the second F2 level of the vowel continuum (see Stimuli). If a participant was biased toward /y/, we raised the level F2 to get overall more /i/ responses, but kept the range of F3 constant. If a participant had an /i/-bias, the F2 level was lowered. Pre-testing was terminated if participants perceived the endpoints of the continua with 90% consistency as /i/ and /y/.

#### 2.2.1. Participants

Sixteen native listeners of Dutch participated for a small monetary compensation. All listeners reported to be a free of (a history) of hearing impairments.

#### 2.2.2. Stimuli

The same stimuli were used as for Experiment 1, only with new vowels added with three levels of F2: 1900 Hz (=Exp. 1), 2000 Hz, and 2100 Hz.

#### 2.2.3. Results and Discussion

The leftmost panel of the middle row of Figure shows the vowel responses in Experiment 2. It shows that participants now respond more strongly to variations in F3 in the vowel, but still show a strong lexical bias. A vowel is more likely to be perceived as /i/ if it occur in a context with a lexical bias toward /i/ (*gen...*). The stronger effect of F3 on vowel perception also has repercussion for the phonological, signal-based, context effect of the fricative on the vowel, which is clearly stronger than in Experiment 1. Replicating the

pattern, vowels with a low F3, likely to be perceived as /y/, induce /s/-percepts for a following fricative. There is, however, no consistent lexical effect consistent anymore over the fricative continuum. That is, the lexical effect on compensation for coarticulation found in Experiment 1 is not replicated here.

A possible reason for the absence of the effect may lie in semantic biases at a phrase level. The two-word phrases in these experiments are not very meaningful, but the most meaningful phrase clearly is "menu shop" (Engl. id.). Hence, the context biasing toward /y/ vowel percepts, and hence /s/ fricative percepts might also induce a /ʃ/-bias at the phrase level, an issue addresses in Experiment 3.

### 2.3. Experiment 3

In this experiment, the rime of the last syllable was [...aft]. Independent of the interpretation of the onset as /s/ or /ʃ/, this yields a nonword in Dutch. This eliminates any semantic bias on a phrase level, which may have influenced the results of the previous experiments.

#### 2.3.1. Participants

Sixteen native listeners of Dutch participated for a small monetary compensation. All listeners reported to be free of (a history) of hearing impairments.

#### 2.3.2. Stimuli

The same stimuli were used as for Experiment 2, only with a different rime of the syllable being [...aft] rather than [...op].

#### 2.3.3. Results and Discussion

The lower panel of Figure 1 shows that the first two effects are replicated. Listeners tend to identify in such a way to make the first stimulus a word. Moreover, a low F3 in the vowel exerts a context effects on the following fricative, which is thereby more likely to be perceived as /s/ rather than /ʃ/. There is also a consistent effect of the lexical context on compensation, which is, however, reverse than in Experiment 1. The lexical context that gives rise to more /y/ responses, *men...*, and as an indirect effect should give rise to more /s/ responses gives rise to more /ʃ/ responses.

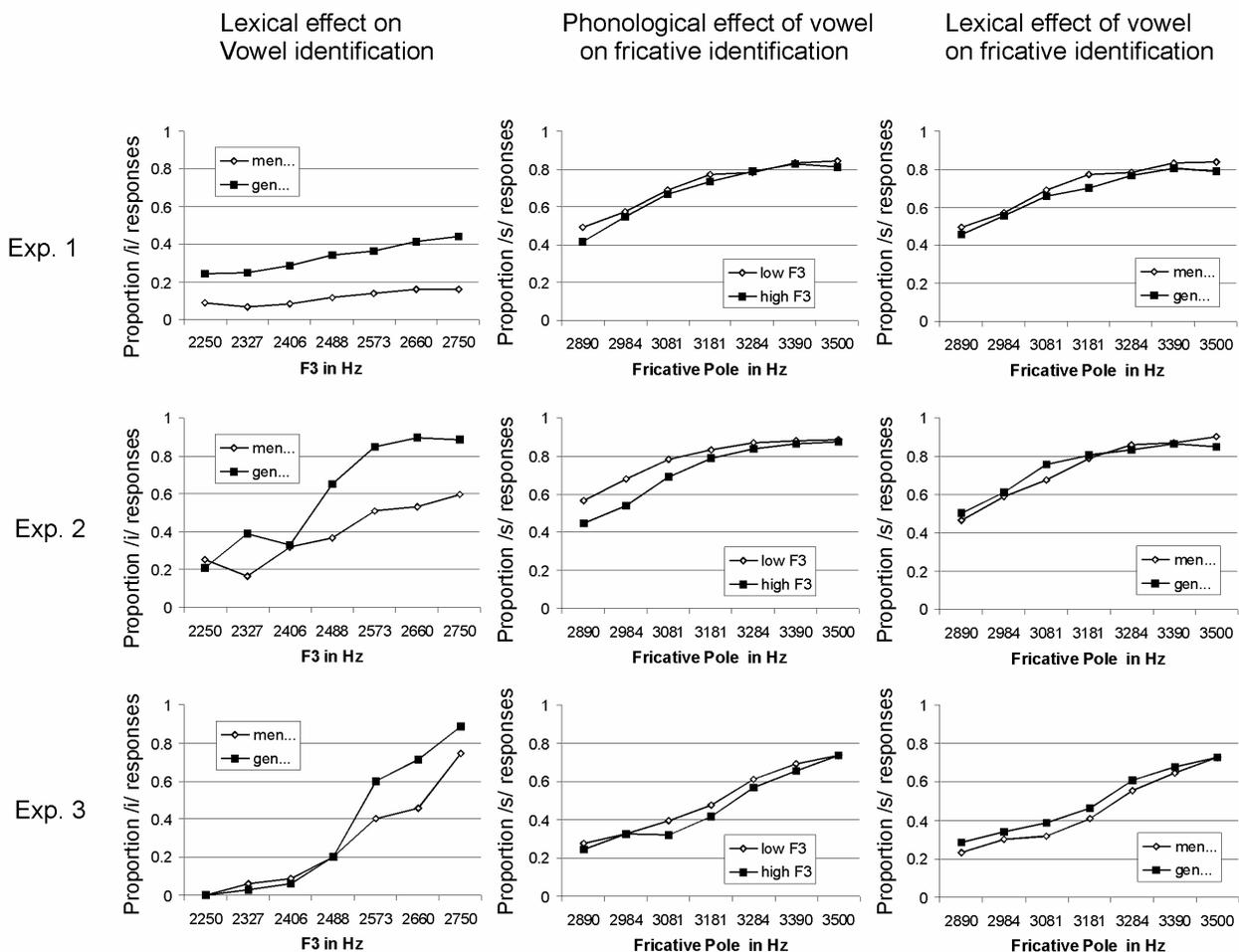


Figure 1: Results from the 4AFC task. The left panels show that lexical knowledge consistently influences the perception of a vowel in all three experiment, The middle panels show a replicable phonological context effect. The right panels show the influence of the lexicon on compensation for coarticulation with a significant positive effect in Experiment 1, a null-effect in Experiment 2, and a reverse effect in Experiment 3.

It is unclear, how this reversed effect could be explained. It is possible that the syllables [ʒen] and [men] give rise to a long-distance acoustic context effect [3].

### 3. General Discussion

In a series of three experiments, we investigated three types of context effects in the perception of vowel-fricative sequences. In all three experiments, we found a significant lexical context effect on vowel perception. That is, vowels were perceived in such a way to give rise to the perception of an existing word. Because in Dutch, [ʒeni] and [meni] are words (Engl. *genius* and *menu*) while [ʒeny] and [meni] are nonwords, listeners tended to perceive the vowel V as /i/ in /ʒenV/, but as /y/ in /ʒenV/ in all three experiments.

A phonological context effect was similarly consistent. Replicating earlier studies [9, 12, 16], we found that fricatives adjacent to the rounded vowel /y/ were more likely to be labeled /s/ than fricatives in the vicinity of the unrounded vowel /i/. This effect was stable over all experiments.

The crucial indirect lexical effect, which would be indicative of true top-down processing, however, was only found in one of the three experiments. Two things are noteworthy in this respect. First of all, it is especially damaging for the assumption of on-line top-down feedback that in two experiments, there was a direct lexical influence on vowel identification *and* a phonological context effect of vowel identity on fricative perception, but no indirect lexical effect. The absence of a lexical effect on compensation for coarticulation has sometimes been explained by "streaming": The two words of an utterance are perceived as "not-belonging-together", which precludes a context effect [15]. This would explain why there is a lexical effect on vowel identification but not on compensation for coarticulation: The vowel and the fricative are not perceived to be *co*-articulated. The presence of a phonological context effect, however, falsifies a streaming account. If the vowel would be perceived to belong to another utterance, it should not exert a phonological context effect on the following fricative.

Secondly, proponents of top-down effects have argued that a meta-analysis of studies on lexical involvement in compensation for coarticulation gives a clear positive picture [11]. This obviously ignores the publication bias for positive results. The current results show clearly that a lexical context effect on a given phoneme and a phonological context effect of this phoneme on an adjacent phoneme can co-occur without an influence of the lexical context effect on the phonological context effect. This suggests the independence of phonological, perceptual, processing from lexical processing.

### 4. References

- [1] Elman, J. L. and McClelland, J. L. "Cognitive penetration of the mechanisms of perception: Compensation for coarticulation of lexically restored phonemes", *Journal of Memory and Language*, 143-165: 1988.
- [2] Ganong, W. F. "Phonetic categorization in auditory word perception", *Journal of Experimental Psychology: Human Perception and Performance*, 6: 110-125, 1980.
- [3] Holt, L. L. "Temporally non-adjacent non-linguistic sounds affect speech categorization", *Psychological Science*, 16: 305-312, 2005.
- [4] Holt, L. L. and Lotto, A. J. "Behavioral examinations of the level of auditory processing of speech context effects", *Hearing Research*, 167: 156-169, 2002.
- [5] Lotto, A. J. and Kluender, K. R. "General contrast effects in speech perception: Effect of preceding liquid on stop consonant identification", *Perception & Psychophysics*, 60: 602-619, 1998.
- [6] Lotto, A. J., Kluender, K. R., and Holt, L. L. "Perceptual compensation for coarticulation by Japanese quail (*Coturnix coturnix japonica*)", *Journal of the Acoustical Society of America*, 102: 1134-1140, 1997.
- [7] Magnuson, J., McMurray, B., Tanenhaus, M. K., and Aslin, R. N. "Lexical effects on compensation for coarticulation: A tale of two systems?" *Cognitive Science*, 27: 801-805, 2003.
- [8] Mann, V. A. and Repp, B. H. "Influence of preceding fricative on stop-consonant perception", *Journal of the Acoustical Society of America*, 69: 548-558, 1981.
- [9] Mann, V. A. and Repp, B. H. "Influence of vocalic context on perception of the sh-s distinction", *Perception & Psychophysics*, 28: 213-228, 1980.
- [10] Mann, V. A. and Soli, S. D. "Perceptual order and the effect of vocalic context of fricative perception", *Perception & Psychophysics*, 49: 399-411, 1991.
- [11] McClelland, J. L., Mirman, D., and Holt, L. L. "Are there interactive processes in speech perception?" *Trends in Cognitive Sciences*, 10: 363-369, 2006.
- [12] Mitterer, H. "On the causes of compensation for coarticulation: Evidence for phonological mediation", *Perception & Psychophysics*, 68: 1227-1240, 2006.
- [13] Norris, D., McQueen, J. M., and Cutler, A. "Merging information in speech recognition: Feedback is never necessary", *Behavioral and Brain Sciences*, 23: 299-324, 2000.
- [14] Pitt, M. A. and McQueen, J. M. "Is compensation for coarticulation mediated by the lexicon?" *Journal of Memory and Language*, 39: 347-370, 1998.
- [15] Samuel, A. G. and Pitt, M. A. "Lexical activation (and other factors) can mediate compensation for coarticulation", *Journal of Memory and Language*, 48: 416-434, 2003.
- [16] Smits, R. "Evidence for hierarchical categorization of coarticulated phonemes", *Journal of Experimental Psychology: Human Perception and Performance*, 27: 1145-1162, 2001.
- [17] Smits, R. "Hierarchical categorization of coarticulated phonemes: A theoretical analysis", *Perception & Psychophysics*, 63: 1109-1139, 2001.
- [18] Vroomen, J. and de Gelder, B. "Lipreading and the compensation for coarticulation mechanism", *Language and Cognitive Processes*, 16: 661-672, 2001.
- [19] Warren, R. M. "Perceptual restoration of missing speech sounds", *Science*, 167: 393-393, 1970.