ISCA Archive
http://www.isca-speech.org/archive

ISCA & IEEE Workshop on Spontaneous
Speech Processing and Recognition
Tokyo Institute of Technology, Tokyo, Japan
April 13-16, 2003

MODELLING ASPECTS OF REDUCTION AND ASSIMILATION OF CONSONANT SEQUENCES IN SPONTANEOUS FRENCH SPEECH

Danielle Duez

Laboratoire Parole et Langage, CNRS UMR 6057

ABSTRACT

The following paper presents spectrographic data of consonant sequences containing one or two consonants omitted and/or changed into another consonant when compared to an existing perception analysis. In most cases, perceptual and acoustic data are shown to strongly correspond, proving that consonants had indeed been changed, significantly reduced or deleted, mainly in a weak position, thereby preserving acoustic information crucial for lexical access, integration of prosodic structure and successful communication. Tentative rules summarise the tendencies observed in reduction and assimilation patterns.

1. INTRODUCTION

The present study analyses the effects of reduction and contextual assimilation on the acoustic structure of two-consonant sequences in spontaneous French speech, both consonants occurring initially or finally in a syllable or occurring in two consecutive syllables.

In his hypo/hyperarticulation model, [1] described natural speech production as an adaptive process by which speakers adjust their performance according to situation demands and communication goals. In casual speech styles such as conversation, speakers talk fast and hypoarticulate, decreasing the duration and amplitude of phonetic gestures and increasing their temporal overlap. At the acoustic level, hypoarticulation is reflected by greater undershoot and contextual assimilation of speech segments, with segments being reduced, omitted or combined with other segments compared to the same phonetic segments of the same words when read.

Hypoarticulation effects are not uniform. Fundamentally, hypoarticulation is governed by a speaker's need to produce an auditory signal possessing sufficient discriminatory power for successful word recognition and communication [1]. The investigation of reduction and contextual assimilation processes in conversational speech provides an explanation of their occurrence and nature and allows a better understanding of basic principles governing them. In this respect, the analysis of consonant sequences in conversational speech is of special interest since consonants in sequences are not discretely produced and are affected by a wide range of interacting linguistic influences.

Section 1 summarises spectrographic data of consonant sequences containing one or two consonants reported as omitted and/or changed into another consonant when compared to an existing perceptual investigation [2]. Consonant omission is the result of an extreme reduction process while consonant change reflects assimilation. Spectrographic data were compared with perceptual data and underlying phonological sequences. Section 2 reviews the interaction of reduction and assimilatory processes by type of word (grammatical or lexical), position in syllables (initial or final), and position in phrases (final or non-final). Section 3 proposes a certain number of rules identified when analysing reduction and assimilation, which are particularly adapted for speech technologies.

2. SPECTROGRAPHIC ANALYSIS

2.1. Perceptual criteria of selected consonant-sequences

In [2], 720 sequences with adjacent vowels, extracted from the conversational speech of two male French speakers (Spk1 and Spk2) were presented for identification to 16 listeners, who were asked to write down everything they heard using orthographic transcription. There were 360 $V_1C_1C_2V_2$ sequences per speaker, each containing the nine following groups OO, OF, OS, FF, FO, FS, SS, SO, SF, where O is an occlusive, F a fricative and S a sonorant. Each group had 40 C_1C_2 sequences. Each sequence was assigned a single identification value corresponding to the phonological underlying sequence reported by at least 75% of listeners. This yielded two main groups:(1) sequences identified as the underlying phonological sequences and (2) sequences identified as different. In the latter group, there were four cases:(2a) the phonological C1 and/or C2 was identified as another consonant (C), (2b) only one or two features were identified, (2c) neither a C nor a feature was identified, and (2d) the C was omitted.

Table 1. For each speaker, number of consonant sequences (CS) with C1 or C2, or both C1 and C2 reported as different.

	CS	C1	C2	C1+C2
Spk1	179	126	27	26
Spk2	166	109	28	29

As can be seen in table1, the number of sequences different from phonological sequences is high and mainly correlate to the high number of C1's different from phonological.

The following presentation is limited to C1's and/or C2's reported as changed into another C and/or omitted as seen in table 2, thereby permetting a focus on extreme cases of reduction and assimilation.

Table 2. Number of C1's and/or C2's omitted (O) and/or changed into another C (Ch).

	C1		C2		C1+C2		
	Ch	O	Ch	O	Ch+O	O+Ch	Ch+Ch
Spk1	18	31	2	13	1	1	1
Spk2	9	38	4	2	7	1	1

2.2. Method

2.2.1. C-segmentation

Spectrograms and oscillograms were generated for each sequence. Sequence onset and offset coincide with the preceding-V offset and following-V onset, respectively. The boundary between C1 and C2 was labeled as follows:

O+(O, F or S): the boundary occurs at the end of the burst of the first O (voiced or unvoiced). In case of absence of a burst, the boundary is the discontinuity (if visible) between the two occlusions (O+O), between occlusion end and noise beginning (O+F), and occlusion and formants (O+S)

F+(F, O or S): The different F's are distinguished by differences in intensity and noise zone. Therefore, the F/F-boundary location is based on changes in spectral zone, discontinuity in voicing, and change in intensity. The cessation of the F noise coincides with the beginning of occlusion (F+O) or F*2* (F+S).

S+(O,F or S). The cessation of the S-F*2* coincides with the appearance of voice bar or silence (S+O), of noise (S+F) or a change in harmonic density or formantic structure (S+S).

2.2.2. Acoustic-cue analysis

Acoustic patterns of C's were examined in relation to perceptual data and phonological C's. The analysis was restricted to cues proven to be relevant in C identification.

O's /*p,t, k, b, d, g*/. For each phonological and perceived O, the presence of an occlusion and/or a burst was controlled, the duration of visible occlusions and bursts measured. Burst frequency, known to be a place cue, was also measured with LPC. Voiced O's have low frequencies whose presence was checked and duration measured.

F's /*f, s, ʃ, v, z, ʒ*/. The presence of noise, typical of F's, was controlled. The limits and maxima of noise frequency, which are place cues, were measured. Voiced F's may have two or three formants, depending on the proportion of voicing and fricative elements. Voiced-F patterns were examined and fricative duration measured.

S's /*m, n, l, ʀ*/. The presence of formants was controlled for each S and measured. It is known that the spectrum of nasals is dominated by a strong low-frequency formant and has mid-frequency formants. Laterals have stronger mid-frequency formants than nasals. The French dorsovelar /ʀ/ is context-dependent : devoiced in an unvoiced context, vocalic in a voiced context.

2.2.3. Analysis of reduction and assimilation

For oral O's reported as nasals, the presence of mid-frequency formants was checked. Three patterns of nasalisation (N) were determined: no N (no formants), partial N (a separate stop exhibiting an occlusion and/or burst) and total N (no interruption of mid-frequency formants). Similarly, three patterns of voicing or devoicing were defined for O's and F's, depending on the presence of low frequencies. The same procedure was used to investigate frication of O's or S's. For place change of O's and F's, burst frequency, noise limits and maximum were measured, for nasals, change in V/C transitions. The absence of noise, closure and burst, and formants was controlled for omitted F's, O's and S's, respectively. Present segments (if any) were measured using the above procedure. To check for traces of underlying Cs, the duration of remaining C1's and C2's was measured and compared with the duration of the same intervocalic C's.

2.2.4. Syllabic structure, word type and position in syllables and phrases

Sequences were divided into homosyllabic [$\#C_1C_2$ or $C_1C_2\#$] and heterosyllabic ([C_1 $\#C_2$] or [C1#C1]). In [C_1 $\#C_2$]'s, C1 and C2 are coda and onset C's, respectively. [C1#C1]'s result from the deletion of an optional mute /ə/ within a word or in a grammatical word [C1(ə)#C1], each C is a C1. C's were also analysed as a function of location in grammatical words or lexical words. With no lexical stress in French, prominent syllables are mostly phrase-final syllables to which prominence was assigned. A prominent syllable is a syllable which stands out from others (i.e. longer, more intense and/or higher in pitch)

2.3 Results

2.3.1. Oral O's reported as nasals

__C1's.__ Six 6/d/'s and 2/t/'s perceived as /n/ in nasal vowel context have complete overlapping of occlusion and mid-frequency formants (about 1200 –Hz, 2000 Hz, 2700 hz and 4000 Hz). The /d/'s are 45 ms, 57 ms, 64 ms, 67 ms and 77 ms long, the /t/'s 94 ms and 55 ms. None has a burst. Surprisingly, 2 /b/'s and 1 /d/ reported as /m/ and /n/ are in an oral vowel context : they may be confusions or misproductions. One /m/ has mid-frequency formants (1100 Hz and 2600 Hz) and is 67 ms long ; the other has an occlusion (70 ms) and a burst (7ms, 1700 Hz), the /n/ has no mid-frequency formants.

C2's. A /p/ reported as /m/ is followed by a nasal vowel (/ɔ̃/) : it is 74 ms long and has a 1000-Hz formant. A /t/ reported as /n/ in an oral vowel context has no formants and burst.

2.3.2. Voiced or devoiced O's and F's

__C1's.__ No low frequencies for 4/ʒ/'s, 1/v/ and 1/z/ and 4/d/'s reported as unvoiced; on the contrary, there are low frequencies for 4/s/'s, 1/p/ and 1/t/ reported as voiced. For all of them, there is an anticipatory effect of a voiced or C2. Although reported as unvoiced 1/ʒ/ is partially voiced (25 ms out of 50ms) and 1/ʒ/, 1/v/ and 1/d/ are fully voiced.

C2's. No low frequencies, but a short frication noise (51ms) for a /v/ reported as /f/ in /sv/.

C1+C2's. A /ts/ sequence is totally voiced: both C's have low frequencies. On the opposite, a voiced sequence (/dʒ/) reported as unvoiced has no voicing. Six single C's result from the coalescence of C1-manner-and-place and C_2 voicing: /sd/=>/z/, /sv/=>/z/, /fz/=>v/, /ʒs/=/ʃ/, /tv/=>/d/. The /z/'s and the /v/ have low frequencies, their respective duration is 150 ms, 120 ms, and 94 ms. The /d/ has both a voice bar and burst (3750 Hz), its duration is 120 ms. The two /ʃ/'s have no low frequencies, their duration is 100ms and 175 ms, respectively. Each reported C is longer than the mean duration of intervocalic /z/'s (70ms, SD: 20), /d/'s (74ms, SD: 34), /ʃ/ (112ms) and /v/'s (65 ms, SD: 25).

2.3.3. *Other cases*

C1's. Two /n/'s reported as /m/ are preceded by a labial vowel. One 20-ms /m/ has a 2500-Hz formant and flat V/C transitions. The other is 65 ms long, has three visible formants (1117 Hz, 2194 Hz and 4209 Hz) and a clear downward movement of the second and third formants at the preceding-V end. A /ʀ/ reported as /b/ is followed by /l/ : with no occlusion and/or burst, this may be a confusion. A /f/ results from /t/-manner-and-place change (/tʀ/ =>/f/), the other from a /s/-place change (/st/=>/f/). Both exhibit a weak noise whose limit is around 1000 Hz (duration is 60 ms and 115 ms, respectively).

C2's. A /p/ in /sp/ is changed into a /f/ with a weak frication noise (limit: 500 Hz, duration: 65 ms). A /l/ in /sl/ is changed into a typical /t/ (silence :50ms; burst : 14 ms, frequency: 4200 Hz). A /b/ reported as /g/ has a voiced bar (62ms) and a burst (18ms, frequency: 1669Hz). A 105-ms /d/ changed into /v/ has a 1200-Hz formant. The /t, g and v/ may be misproductions.

C1 and C2. A /z/ resulting from the merger of /d/ place and /v/ manner has a visible noise (limit: 3500 Hz, maximum: 4500 Hz).

2.3.4. *Omitted C1's or C2's*

C1's. There are 1 /l/ and two /ʀ/'s omitted despite visible mid-frequency formants (1500Hz; 1472Hz and 2281 Hz; 2300 Hz). Their short duration (30ms, 18 ms, 30 ms) may have not given listeners sufficient time to identify them. There is no trace of formants for the other S's, which are deleted. The duration of the remaining C2's is often equal or inferior to reference C's. Two identical C's usually coalesced into a single C longer than reference C's. For example, /ʀ/ duration is 96ms, 76 ms for Spk1 and 77 ms for Spk2: mean reference duration: 58 ms and 52 ms for the two speakers. C1's and C2's having the same place and manner tend to combine into a single C also longer than reference C's: Spk1 (/s/:156 ms, reference /s/: 104 ms); Spk2 (//f/: 145 ms, reference //f/: 88 ms; /t/: 94 ms and 122 ms, reference /t/: 91 ms). For C1's and C2's with the same manner and voicing, the trend is similar although less acute. The remaining /z/, /p/ and /d/ in /gz/, /tp/ and /kd/ is 100 ms, 114ms and 105 ms, respectively: the mean reference duration of /z/'s, /p/'s and /d/'s is 70 ms, 90 ms and 61ms. In addition, there are often traces of the underlying C in preceding V transitions: e.g. in case of omitted velar, there is a merger of the second and third formants at the end of the preceding V.

C2's. A few omiited S's (1/m and 2 /l/'s) have visible formants. Again their short duration (from 15ms to 20 ms) may explain listeners' omission. There are no visible formants for the other omitted S's (Spk1: 2 /ʀ/'s and 3/ /l/'s; Spk2: 1 /ʀ/ and 2/ /l/'s) and no trace of C2 in the duration of the remaining C. Four omitted O's have the same articulation place as their preceding C1's : 1 /d/ in /zd/, 1/d/ in /td/ and 2 /t's belonging to /st/'s. The remaining /z/ has a duration (75 ms) close to reference /z/'s (78 ms), contrary to the two remaining /s/'s which are much longer (145 ms and 175 ms) than reference /s/'s. The remaining /t/ of /td/ has no voice bar and a duration equal to reference /t/'s.

2.3. Discussion

Listener reports differ occasionnaly from spectral cues, due to confusions caused by artificial truncation of vowels used in perception tasks. Misproductions are frequent in spontaneous speech, there may be some in the above acoustic data. However, generally, perceptual and acoustic data strongly correspond, proving that C's had been changed, significantly reduced or deleted. Two main tendencies emerge for assimilated C's:1) the nasalisation of O's, which can be viewed as the result of overlapping of the velum-lowering gesture of a preceding (or more rarely following) nasal vowel and the closure gesture of an occlusive and 2) the devoicing or voicing of voiced or unvoiced C1's and/or C2's (O's or F's) due to the anticipatory and/or the carryover effect of an unvoiced or voiced C2 or C1. Less frequent are the cases of O's frication, mainly due to the fricative gesture overlapping the closure gesture or the merger of C1 place and manner with C2 voicing. Strikingly, most changed C's keep their place of articulation, confirming results on intervocalic voiced stops [3]. Omitted S's are very short or deleted, reflecting a strong decrease in the magnitude of the movements associated with the gesture, or the complete deletion of the gesture [4]. Other omissions reflect assimilatory processes leading to the deletion of O's or F's. In some cases, there is trace of residual gestures in V/C transitions and/or in the long duration of single remaining C's.

3. SYLLABLE STRUCTURE, WORD TYPE, LOCATION IN SYLLABLES AND PHRASES

The four above factors significantly affect the realisation of C's. Out of the 31 deleted C1's, 30 are in heterosyllabic sequences, 28 in coda position, 29 in non-final syllables and 19 in function words. Out of the 20 assimilated C1's, 16 belong to heterosyllabic sequences, 17 to non-final syllables. The effect of position in syllables and word type is weaker than for deleted C1's: 8 are in function words, 9 are coda C's and 7 C1's in C1#C1's. For Spk2, there is also a significant effect of syllable structure and location in syllables. Out of the 39 deleted C1's, 34 are in heterosyllabic sequences (4 in C1#C1's) and 30 in coda position; out of the 16 assimilated C1's, 14 are in heterosyllabic sequences (4 in C1#C1's), 10 in coda position. Deleted and assimilated C1's are mainly located in non-final syllables (38 and 15, respectively).

For C2's, there is a significant effect of location in phrases, but not in syllables for Spk1. For Spk2, it is the opposite. Omitted and assimilated C2's are rare, however they are interesting cases such as the deletion of function words with a dropped /ə/ (l', d'), the deletion of the first C of a verb after the clitic /j'/ and the deletion or assimilation of /t/ in homosyllabic

/st/'s.

Reduction and assimilatory processes particularly affect coda C's in non-final syllables, thereby preserving sequences which carry all information crucial for lexical access, integration of prosodic structure and successful communication.

4. TENTATIVE RULES

Spontaneous speech is well known for its extreme variability in speech segments. Even so, the acoustic data obtained reveal certain regularities in patterns of reduction and assimilation. The following tentative rules summarise some of the aspects of reduction and assimilation.

Each rule is structured X=>Y/W__Z where X is rewritten after left-hand context W and before right-hand context Z. If Y is absent, there is a deletion rule. Each symbol is composed of a phonetic segment (V and C), shown as a function of binary features. Full stop marks syllable boundary. As assimilated and omitted C's are mainly in non-prominent syllables, the feature /-PROM/ is not represented. Variables α, β… mark compatibility between the signs of features.

Nasalisation of voiced and unvoiced O's after a nasal vowel

$$\begin{bmatrix} C \\ -nas \\ +occ \end{bmatrix} \Rightarrow \begin{bmatrix} +nas \end{bmatrix} / \begin{bmatrix} V \\ +nas \end{bmatrix} \_\_\_ .C$$

Voicing of voiceless coda obstruents before a voiced obstruent

$$\begin{bmatrix} C \\ -voice \\ +\mathrm{obst} \end{bmatrix} \Rightarrow \begin{bmatrix} +voice \end{bmatrix} / \_\_\_ . \begin{bmatrix} C \\ +voice \\ +obst \end{bmatrix}$$

Devoicing of voiced obstruents before unvoiced obstruents

$$\begin{bmatrix} C \\ +voice \\ +obst \end{bmatrix} \Rightarrow \begin{bmatrix} -voice \end{bmatrix} / \_\_\_ . \begin{bmatrix} C \\ -voice \\ +obst \end{bmatrix}$$

Coalescence of coda voiced or voiceless obstruents with their voiceless or voiced counterparts

$$\begin{bmatrix} C \\ +obst \\ \alpha lab \\ \beta dent \\ \varepsilon vel \\ \alpha voice \end{bmatrix} \Rightarrow \begin{bmatrix} -\alpha voice \end{bmatrix} / \_\_\_ . \begin{bmatrix} C \\ +obst \\ \alpha lab \\ \beta dent \\ \varepsilon vel \\ -\alpha voice \end{bmatrix}$$

Reduction of double C's

$$\begin{bmatrix} C \\ \alpha cont \\ \beta obst \\ \varepsilon lab \\ \phi dent \\ \gamma vel \\ \gamma voice \\ \xi nas \end{bmatrix} \Rightarrow \varnothing \quad / \_\_\_ . \begin{bmatrix} C \\ \alpha cont \\ \beta obst \\ \varepsilon lab \\ \phi dent \\ \gamma vel \\ \gamma voice \\ \xi nas \end{bmatrix}$$

Vocalisation of /ʀ/ in coda position. Vi represents all features of the preceding vowel

$$[R] \Rightarrow [Vi] \; / \; [Vi] \_\_ . C$$

Deletion of /l/ in "il"-clitics

$$[l] \Rightarrow \varnothing \; / \; \_\_ . C$$

Deletion of voiced C's after "j'" [ʒ]- clitic

$$\begin{bmatrix} C \\ +voice \end{bmatrix} \Rightarrow \varnothing \; / \; [j'] \_\_\_\_\_\_$$

REFERENCES

[1] B. Lindblom, "Explaining phonetic variation: a sketch of the H and H theory". In *Speech production and speech modelling* (W. Hardcastle and A. Marchal, editors), Vol 55, pp. 403-439. NATO ASI Series. Dordrecht, Boston and London: Kluwer Academic Publishers, 1990.

[2]D. Duez,"Consonant sequences in Spontaneous French speech", *SPoSS,* ESCA, La Baume les Aix, .24-26, 1998.

[3]D. Duez, "On spontaneous French speech: aspects of the reduction and contextual assimilation of voiced plosives*",* *Journal of Phonetics,* 25. 407-427, 1995.

[4]C.P. Browman, and L. Goldstein, "Articulatory phonology: An overview", *Phonetica,* 49, 155-180, 1992.