

Voicing-feature opacity of whispered French fricatives: the interplay between duration and position in voicing perception

Yohann Meynadier, Noël Nguyen, Sophie Dufour
Aix Marseille Univ, CNRS, LPL, UMR7309, Aix-en-Provence, France
Aix Marseille Univ, ILCB, Marseille, France

Whisper creates opacity of the voicing-feature between the phonetic surface and the phonological representation of French words. Opacity comes from the fact that the voicing-feature is no longer conveyed by the vocal fold vibration at the phonetic level but by other secondary acoustical properties. Indeed, the whispered [+voiced] obstruents are recognized at better than chance level [1, 2, 3, 4, 5] and whispered speech production studies have shown that the voicing contrast is maintained by phonetic acoustic traces [4, 6, 7, 8, 9]. As in modal speech, the [+voiced] obstruents are shorter than the [-voiced] obstruents. Lastly, in modal speech the manipulation of obstruents' duration influences the identification of their voicing feature [10]. However, no study has addressed this point in whispered speech. In that respect, consonant duration could also play a main role in the phonological recovery of the obstruent voicing feature when phonetic voicing is missing. To clarify the voicing-feature opacity in whisper, three questions were addressed here: (1) Does duration impact the identification of consonant voicing-feature? (2) At which level of processing (pre-lexical vs. lexical) is phonetic duration used to recover the intended voicing-feature? (3) Is the duration effect word-position-dependent, knowing that a coda consonant in word-final position undergoes prosodic final lengthening? To do this, we manipulated the duration and position (initial vs. final) of /ʃ/ vs /ʒ/ fricatives within words and non-words in an identification task. Using Praat's Vocal Toolkit, we resynthesized /ʃ/ and /ʒ/ embedded in CVC stimuli on a duration continuum of 11 steps of 20 ms, from 50 to 250 ms. According to the position, the remaining _VC and CV_ parts of the stimuli were phonetically neutralized by being re-synthesized at mid-way in duration and spectral shape between the [+voiced] and [-voiced] exemplars. 32 participants were tested on words and had to decide whether they heard *char* (tank) or *jarre* (jar) for the word-initial position and *cache* (cover) or *cage* (cage) for the word-final position. 32 other participants were tested on non-words and had to decide whether they heard *cheur* or *jeur* for the initial position and *queuche* or *queuge* for the final position. To answer our first question, the percentages of correct identification as a function of the duration continuum (Figure 1) were analyzed using an LME model, with a binomial linking function ('0' = incorrect, '1' = correct) that included *Duration* (11 levels), *Voicing* ([+voiced], [-voiced]), *Position* (initial, final), *Lexicality* (word, non-word) and their interactions as fixed effects. The model revealed a significant *Duration*Voicing* interaction [$\chi^2=5341.61$; $p<.0001$]. The longer the /ʃ/ is, the better it is perceived [-voiced], and conversely the shorter the /ʒ/ is, the better it is identified as [+voiced]. Thus, as expected, the voicing-feature of whispered fricatives is better recovered as expected from their duration in production. The model also revealed a significant *Duration*Voicing*Lexicality*Position* interaction [$\chi^2=174.86$; $p<.0001$]. We performed a second analysis on the durational values of the categorical perceptual boundary of voicing to clarify the role played by *Position* and *Lexicality*. For each subject in each condition, we calculated the boundary by detecting of the duration value matching the 50% intercept of the cubic spline function of the voicing-identification curves. 11 subjects over 64 were excluded because no categorical switch could be extracted by the modelisation in one condition at least. We conducted a 3-way ANOVA with *Voicing* and *Position* as within factors and *Lexicality* as between factor. A significant *Voicing*Position*Lexicality* interaction was found [$F(1,54)=4.32$; $p<.05$]. Subsequent pairwise comparisons (with Bonferroni correction) examining the effect of voicing for the words and the non-words and within each position revealed significant boundary differences between /ʒ/ and /ʃ/ when the

fricative occurs in final [F(1,52)=38.01; $p < .0001$] but not in initial position of words, as for non-words, [F(1,52)=144.46; $p < .0001$] in final position. Hence, significant displacements toward higher durational boundary values for the [+voiced] fricative occurred only in final position for both words and non-words (Figure 2). Interestingly, pairwise comparisons also indicated that the final position induced a delayed boundary in comparison to initial position only for the [+voiced] fricative in both words [F(1,52)=48.19; $p < .0001$] and non-words [F(1,52)=66.26; $p < .0001$]. Altogether, the durational correlates of voicing seem to be recruited pre-lexically since they are perceived and processed similarly in words and non-words. It also appears that the identification of the [+voiced] feature better resists to lengthening in the (non)word-final position. Therefore, the question is also why [-voiced] is not concerned by the position effect? A proposal in two points can be made. First, [-voiced] fricatives are phonetically long by default; so, no special processing seems necessary to associate a perceptual long fricative to the [-voiced] feature. Rather, the longer a whispered [+voiced] is, the more it runs the risk to be confounded with the [-voiced] one. A different processing could have to be used by the listener (or the speaker) to compensate lengthening of whispered [+voiced] fricatives. It is in line with the process evidenced in whispered speech by Dufour & Meynadier [5] who showed that the identification of the [+voiced] feature requires additional time compared with the immediate recognition of the [-voiced] feature. Secondly, the listener (and/or the speaker) should use his prosodic knowledge (e.g. on final lengthening) to restore (and/or to compensate) segmental opaque features in adverse speech.

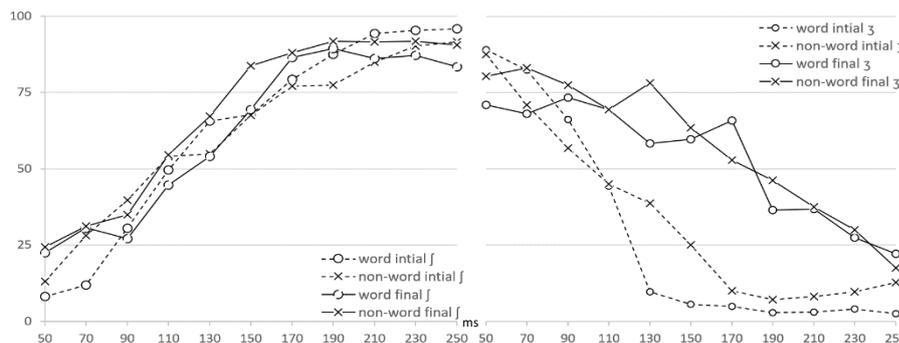


Figure 1: Identification percentages (y axis) of the voicing-feature in whispered words and non-word for /ʃ/ and for /ʒ/ in initial and final position as a function of consonant duration (x axis)

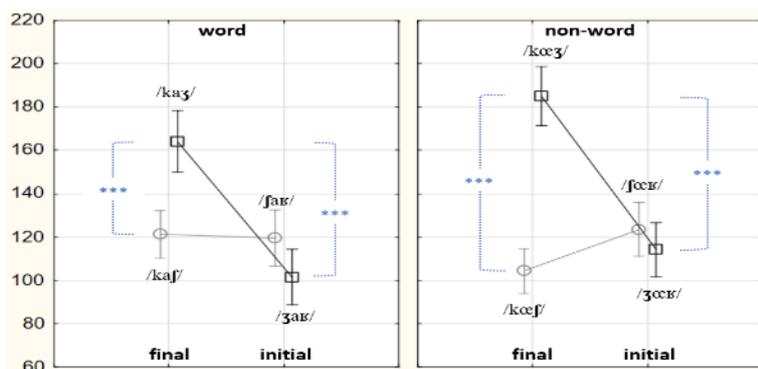


Figure 2: Mean duration value of the categorical perceptual boundaries of the voicing of whispered fricatives as a function of their voicing-feature and their position for words and non-words. Bars for 95% confident interval, *** for significant differences with $p < .0001$. No * for non-significant differences ($p > .05$)

[1] Dannenbring (1980). Perceptual discrimination of whispered phoneme pairs. *Percept Motor Skill* 51(3): 979-985. [2] Gilichinskaya, Strange (2011). Perception of final consonant “voicing” in whispered speech. *JASA* 129: 2420. [3] Munro (1990). Perception of “voicing” in whispered stops. *Phonetica* 47: 173-181. [4] Tartter (1989). What’s in a whisper? *JASA* 86: 1678-1683. [5] Dufour, Meynadier (2019). Temporary ambiguity in whispered word recognition: a semantic priming study. *J Cogn Psychol* 31(2): 157-174. [6] Jovičić, Šarić (2008). Acoustic analysis of consonants in whispered speech. *J Voice* 22: 263-274. [7] Kohlberger, Strycharczuk (2015). Voicing assimilation in whispered speech. *18th ICPHs*. Glasgow. [8] van de Velde, van Heuven (2011). Compensatory strategies for voicing of initial and medial plosives and fricatives in whispered speech in Dutch. *17th ICPHs*. Hong-Kong. [9] Meynadier, Gaydina (2013). Aerodynamic and durational cues of phonological voicing in whisper. *14th Interspeech*. Lyon. [10] Denes (1955). Effect of duration on the perception of voicing. *JASA* 27(4): 761-764.