Original Paper

Phonetica 2010;67:1–24
DOI: 10.1159/000319376

Received: September 4, 2009
Accepted: March 11, 2010

The Role of the Spectral and Temporal Cues in Consonantal Vocalization and Glide Insertion

Daniel Recasens Aina Espinosa

Universitat Autònoma de Barcelona and Institut d'Estudis Catalans, Barcelona, Spain

Abstract

This study investigates the perceptual role of several acoustic characteristics to glide generation processes affecting the consonants [ɫ], [β] and [ɲ], i.e., the vocalization of syllable-final [ɫ] and syllable-initial [β] into [w], and the insertion of [j] before syllable-final [ɲ]. Results from identification tests with synthetic speech stimuli performed on Catalan-speaking informants reveal that both the formant frequency characteristics (at the consonant steady-state period for [ɫ] and [β], and at the endpoint of the vowel transitions for [ɲ]), and the onset or onset/offset time of the vowel transitions may play an active role in vocalization and glide insertion. Mostly for the changes [ɫ] > [w] and [ɲ]> [jɲ], glide identification was triggered by formant frequency variations rather than by variations in the temporal implementation of the vowel transitions. The implications of the perception results for the interpretation of the sound changes of interest are evaluated.

1. Introduction

While a good deal of experimental research has been carried out on the causes of sound change, much knowledge is lacking about the perceptual strategies used by listeners for replacing one phonetic sound by another. The present paper investigates the perceptual mechanisms involved in the generation of a glide out of the dark alveolar lateral [ɫ], the bilabial approximant [β] and the alveolopalatal nasal [ɲ] in frequent sound changes affecting these three consonants: the vocalization of [ɫ] and [β] into [w], and the glide insertion process [ɲ] > [jɲ]. The study deals mostly with a Romance language, i.e., Catalan, where all five consonants [ɫ], [β], [ɲ], [w] and [j] occur. While both changes give rise to a glide, they differ in that the former involves the substitution of one sound by another and the latter the addition of a glide before the original consonant. The term 'vocalization' refers to a change in the consonant percept from [ɫ] to [w], not to the loss of tongue contact at the alveolar place of articulation. On the other hand, 'glide insertion' is the standard term used in segmental phonology for denoting a particular kind of phonetic development, i.e., the perceptual categorization of a transitional sound from the preceding vowel into the alveolopalatal consonant. Moreover,

KARGER

Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

0031–8388/10/0672–0001
$26.00/0
Accessible online at:
www.karger.com/pho

Daniel Recasens
Departament de Filologia Catalana
Universitat Autònoma de Barcelona, Bellaterra
ES–08193 Barcelona (Spain)
Tel. +34 3 4578183, Fax +34 3 5812782
E-Mail daniel.recasens@uab.es

vocalized realizations and realizations with a preposed glide may be found as alternating synchronic variants and as phonemicized sound changes depending on the dialectal scenario. An overview of the positional and contextual conditions involved in the implementation of these sound changes is presented next.

In the Romance languages, the vocalization of syllable-final [ɫ] is most prone to occur after low or back rounded vowels and before labial and velar consonants and, less so, before dentals and alveolars (Catalan dialects [ˈawβə] derived from Latin ALBARU ‘poplar’, [awkəˈɾiə] for [alkəˈɾiə] ‘farmhouse’, Sicilian [ˈsawtu] SALTU ‘jump’) [Recasens, 1996, pp. 314–315; Rohlfs, 1966, p. 343]. In Old English, too, [ɫ] vocalization occurred after back or open vowels and before labials and velars but not before alveolar stops [*walk*, *half*, *salt*; Gimson, 1970, p. 204]. [ɫ] vocalization may also take place word-finally and more rarely intervocalically (Occitan dialects [saw] SALE ‘salt’, [ˈawo] ALA ‘wing’) [Lafont, 1983, p. 56; Ronjat, 1932, vol. II, p. 144]. Evidence for the gradual application of this vocalization process is consistent with the presence of other, mostly back rounded, glide or vowel outcomes (Gascon [kuˈtɛɔ] CULTELLU ‘knife’, London speech [mɪo̯k] *milk*) [Fleischer, 1912, p. 56; Wells, 1982, vol. 1, p. 259]. Moreover, instances of in-between variants which are difficult to assign to either [ɫ] or a glide or a vowel have been reported in the literature [Stuart-Smith et al., 2006, regarding Glaswegian].

The vocalization of syllable-final [β] operates typically word-finally or before any consonant (Old Portuguese *ausoluto* for *absoluto* ‘sheer’, Catalan [ˈdɛwtə] DEBITU ‘debt’, [ˈbɛw] BIBET ‘he/she drinks’) [Williams, 1938, p. 41]. Judging from the syllable affiliation rules operating in the Romance languages, [β] vocalization appears to apply syllable-initially in tautosyllabic clusters such as [βl] and [βɾ] and more rarely in intervocalic position (dialectal Spanish [awˈlaɾ] for [aˈβlaɾ] ‘to speak’, Tuscan Italian [ˈfawa] FABA ‘faba bean’) [Lipski, 1994, pp. 74–75; Rohlfs, 1966, p. 293].

Regressive glide insertion before syllable-final [ɲ] accounts for the realization [jn], but also for the outcomes [jɲ, jŋ], in Romance languages and dialects [Recasens et al., 1995]. While these phonetic variants appear to occur often in free variation, [ɲ] > [jn] may also be a categorical sound change. Thus, in Majorcan Catalan, preconsonantal /ɲ/ takes a glide systematically after which the nasal assimilates to the place of articulation of the following consonant, e.g., /baɲ ˈturk/ ‘Turkish bath’ and /baɲ ˈbɔ/ ‘good bath’ are realized [bajn ˈturk] and [bajm ˈbɔ], respectively. Moreover, as proposed by several scholars, the same sound change occurred word-finally and preconsonantally in Old French. Thus, a word like *bain* ‘bath’ was generated through the derivation [baɲ] > [bajɲ] > [bajn] > [bɛ̃], where the alveolopalatal nasal shifted to [jn] and the glide merged with the preceding vowel afterwards [Lausberg, 1970, p. 273; Pope, 1934, p. 161]. As shown in section 1.3, [j] insertion may also take place next to other (alveolo)palatal, alveolar and velar consonants. [ɲ] may undergo vocalization intervocalically and word-finally (Dacoromanian [ˈvije] VINEA ‘vineyard’, Gers Gascon [luj] LONGE ‘far away’) [Lausberg, 1970, p. 392; Sampson, 1999, p. 153], and depalatalization without glide insertion syllable-finally (Alguerese Catalan /aɲ ˈnɔw/ ‘new year’ is realized [an ˈnɔw]).

The present paper investigates using synthetic speech stimuli the relative perceptual power of the steady-state consonant period and the vowel transitions in the implementation of the sound changes [ɫ], [β] > [w] and [ɲ] > [jn] since, as argued in sections 1.1 through 1.3, both acoustic characteristics appear to play a role in vocalization and glide insertion. Therefore, a central goal of the present study is to achieve a more thorough understanding of the relative power of static and dynamic acoustic cues in the

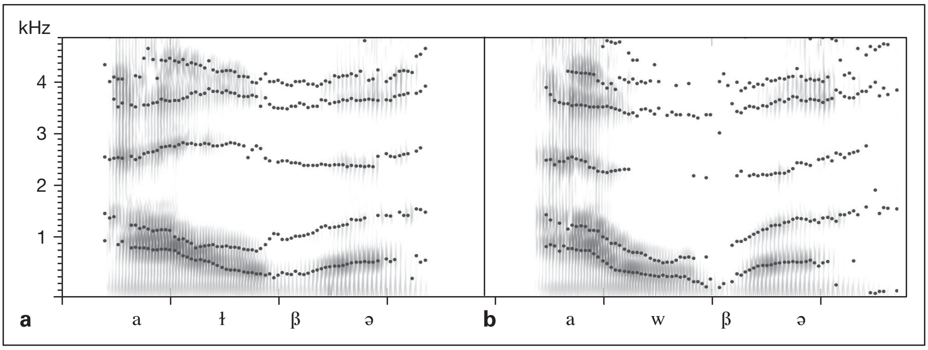

Fig. 1. Spectrographic displays for the Catalan words ['aɫβə] (**a**) and ['awβə] (**b**) produced by speaker DR. LPC trajectories have been overlaid on the formant displays.

generation of a glide from [ɫ], [β] and [ɲ]. [ɫ] vocalization and [ɲ] decomposition will be analyzed syllable-finally since this position favors both changes, while [β] vocalization will be investigated syllable-initially in tautosyllabic consonant clusters given that the bilabial approximant occurs typically in this syllable position in Catalan [see also Navarro Tomás, 1972, p. 173 for Spanish].

1.1. Dark Alveolar Lateral

1.1.1. Spectral Similarity

The replacement of [ɫ] by [w] should be associated primarily with spectral similarity, namely, with both sounds sharing a spectral configuration with a low F2 frequency about 1,000 Hz [von Essen, 1964; Ohala, 1974]. Data for American English [Lehiste, 1964] reveal indeed the existence of similar spectra for [ɫ] and [w] in word-initial position across vowel contexts (F1 = 305 Hz, F2 = 630 Hz, F3 = 2,180 Hz for [w]; F1 = 295 Hz, F2 = 950 Hz, F3 = 2,610 Hz for [ɫ]), and more so word-finally after low and mid high back rounded vowels (F1 = 545, 410 Hz, F2 = 850, 740 Hz, F3 = 2,325, 2,335 Hz for [w]; F1 = 415, 435 Hz, F2 = 870, 905 Hz, F3 = 2,225, 2,435 Hz for [ɫ]). A low F2 is associated with the formation of a large mid cavity between the alveolar closure for [ɫ] or the labial constriction for [w], and the postdorsal constriction at the rear of the vocal tract for the two consonants. There are, however, some relevant spectral differences which render the replacement of [ɫ] by [w] less straightforward. Thus, F2, F3 and possibly F1 are regularly lower for [w] than for [ɫ] in line with well-established articulatory characteristics for the two sounds: lip rounding and a narrow dorsovelar constriction for the glide, and a wider pharyngeal constriction, a small front cavity and considerable oral opening for the lateral [Bladon, 1979; Browman and Goldstein, 1995; Fant, 1960; Gick et al., 2002; Keating et al., 1994; Koneczna and Zawadowski, 1951, 1956; Lindblad and Lundqvist, 2003; Martins et al., 2008; Narayanan et al., 1997]. The spectrographic displays for the Catalan words ['aɫβə] ‘dawn’ and ['awβə] ‘poplar’ in figure 1, and to a large extent Lehiste’s [1964] data reported above, show indeed lower F2, F3 and F1 frequencies for [w] than for [ɫ]. Moreover, [ɫ] but not [w] exhibits spectral zeroes and a more or less audible release [Espy Wilson, 1992; Fant, 1960].

The spectral configuration of [ɫ] is influenced by the same positional and contextual factors which trigger vocalization. [ɫ] is often darker syllable-finally than syllable-initially, which is consistent with a trend for consonants to be articulated with more tongue body lowering and retraction in the former position than in the latter [Browman and Goldstein, 1995; Giles and Moll, 1975; Recasens, 2004]. Degree of darkness should also increase after a back vowel and before a labial or a velar consonant since the production of these contextual sounds involves a low predorsum and a back place of articulation (back vowels, velar consonants) or does not interfere with the lingual configuration for [ɫ] (labials). A relatively lowered tongue predorsum and perhaps some tongue body retraction may also be available for other contextual consonants allowing [ɫ] vocalization to occur, i.e., dentals [Dart, 1991], as well as [s], which exhibits lingual grooving for the passage of airflow. In agreement with this contextual scenario, [ɫ] in Majorcan Catalan was found to show less dorsopalatal contact and a lower F2 frequency before labials and dentals (850–900 Hz) than before alveolopalatals (1,100 Hz), and intermediate contact and F2 values before alveolars and velars [Recasens, 2009].

It has been argued that [ɫ] vocalization cannot only be achieved by spectral equivalence but also by alveolar contact loss [Grammont, 1971, p. 207; Straka, 1965]. This is a controversial research issue since the two hypotheses explain fairly successfully the outcome [w] in most contextual conditions, i.e., after low and back rounded vowels and before labials and velars. Alveolar contact loss for [ɫ] next to all these contextual sounds has been reported in the literature [Hardcastle and Barry, 1989; Recasens, 2009] and is consistent with their being articulated without a front lingual closure or constriction. The acoustic-based hypothesis seems to account better for [ɫ] vocalization before dentals and alveolars since these consonants cause the preceding alveolar lateral to maintain full apical contact at closure location.

1.1.2. Vowel Transitions

The generation of [w] out of syllable-final [Vɫ] sequences is based most feasibly not only on spectral similarity in the positional and contextual conditions mentioned above but also on the perceptual categorization of the vowel formant transitions as an independent phonetic segment by listeners [Gick and Wilson, 2001; Recasens, 1999]. The fact that glide insertion operates only at the regressive level in this case is in accordance with [ɫ] exerting prominent anticipatory coarticulation effects on the preceding vowel. Instances of [w] insertion before the alveolar lateral may be found in Sutselvan Romansh ([awlt] ALTU ‘high’) [Haiman and Benincà, 1992, p. 55] and Minho Portuguese (['siwrba] SILVA ‘forest’) [Leite de Vasconcellos, 1987, p. 96], and presumably in written forms taken from early texts (Old Venitian *aultro* ALTERU, *faulssamente* derived from FALSU, High Picard *saulz* SOLIDOS, *voult* VOLET, Middle French *paulmier* PALMARIU, *maulve* MALVA, Old and Middle English *al* > *aul* ‘all’, *salt* > *sault*) [Gossen, 1970, p. 73; Lass, 1980, p. 37; Tuttle, 1991; von Wartburg, 1922, vol. VI(1), p. 129, vol. VII, p. 515]. Glide insertion before [ɫ] and [ɫ] vocalization could occur either independently of each other ([VɫC] > [VwɫC], [VwC]) or else in succession ([VɫC] > [VwɫC] > [VwC]). In the latter event, listeners would first categorize the salient vowel transitions as an independent glide and then eliminate the alveolar lateral whenever it becomes hardly audible next to spectrally similar phonetic segments. Thus, it may be that ALTU has yielded *haut* in French through the intermediate form [awlt], which is still present in Sutselvan Romansh.

Several acoustic characteristics could cause the vowel transitions to be integrated as a labiovelar glide by listeners. Lowering the F2 frequency at the steady-state period of [ɫ] contributes to an increase in the frequency extent of the vowel transitions, thus rendering them more [w]-like. Another relevant acoustic cue could be absolute duration such that the longer the vowel transitions, the higher the chances that the lateral be confused with the glide. The spectrographic displays for ['aɫβə] and ['awβə] in figure 1 reveal indeed the existence of a somewhat longer F2 vowel transition for [w] than for [ɫ]. Data in the literature suggest, however, that it is the relative timing rather than the absolute duration of the vowel transitions which plays an important role in the identification of [ɫ] with [w]. Thus, according to lingual movement data for American English [ɫ], tongue dorsum retraction offset is nearly synchronous with tongue tip raising onset in *peal* but with tongue raising offset in *leap*; moreover, the longer the syllable rime in *peal*, the greater the degree of tongue dorsum lowering and retraction, the larger the F2–F1 difference and the greater the temporal lag between tongue dorsum lowering and tongue tip raising extrema [Browman and Goldstein, 1995]. An obvious acoustic consequence of the articulatory data just described is for the vowel transitions for [ɫ] to occur earlier in time, the more the tongue dorsum lowering is anticipated with respect to tongue tip raising; this should be so since F2 is strongly correlated with the tongue dorsum motion and the acoustic period of the alveolar lateral begins roughly at tongue tip displacement maximum. In agreement with this expectation, the onset time of the F2 vowel transition in the sentence-final sequence /il/ was found to vary with degree of darkness for [ɫ] in the progression American English > Catalan > Italian [Recasens and Farnetani, 1994].

Similar findings pointing to the relevance of the relative timing of articulatory events have been reported for the bilabial nasal consonant [Krakow, 1989, 1993, 1999]. Indeed, velum lowering offset coincides roughly with lip raising offset for initial [m] (sequences 'see more', 'pa made') but with lip raising onset for final [m] (sequences 'seem ore', 'palm aid'). In addition, while velum lowering lasts a similar amount of time for both positional allophones, final [m] involves a larger lowering movement, a lower velic position and longer low velic plateaus than initial [m]. Also for /w/, kinematic data for American English show that the lip constriction maximum often follows the tongue body backing maximum syllable-finally, while the former event always precedes the latter syllable-initially [Gick, 2003].

In the light of these data, our hypothesis is that preconsonantal [ɫ] in syllable-final position could be identified with [w] (or else with [wl]) by lowering F2 at the steady-state consonant period and/or by anticipating the VC transitions of a relatively constant duration.

1.2. Bilabial Consonant

The replacement of the approximant [β] by [w] parallels the sound change [ɫ] > [w] in some respects. On the one hand, [β] vocalization could be induced by spectral similarity since both consonant realizations [β] and [w] share a low F2 at about 1,000 Hz, which is mostly related to lip closing for the bilabial cognate and to lip closing and protrusion and to tongue dorsum backing for the labiovelar one [Ohala, 1978]. It should also be noticed that F2 is often lower for [w], i.e., always below 1,000 Hz, than for [β], i.e., between 600 Hz and 1,500 Hz depending on vowel context (see the

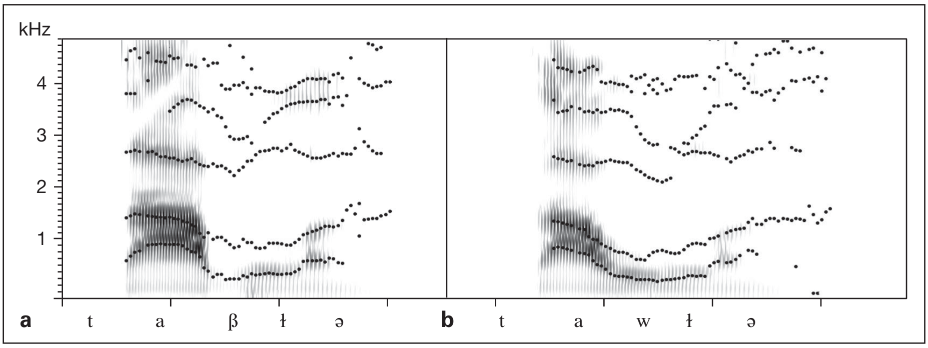

Fig. 2. Spectrographic displays for the Catalan words ['taβɫə] (**a**) and ['tawɫə] (**b**) produced by speaker DR. LPC trajectories have been overlaid on the formant displays.

spectrographic displays for ['taβlə] 'board' and ['tawlə] 'table' in figure 2, and data on Catalan [β] and [w] in Recasens [1986] and on English [b] and [w] in Mack and Blumstein [1983]). It appears then that some F2 lowering is needed for [β] to be categorized as [w], which may be possibly achieved when the labial closing gesture becomes especially prominent.

Instances of [w] insertion before labial consonants suggest, on the other hand, that [β] vocalization may also be triggered by the vowel transitions, e.g., Occitan, Francoprovençal ['mowblo] MOBILE, Wallon [ta^{w}p, ta^{w}f] TABULA (where [^{w}] represents an evanescent glide), Old Bourguignon *deaubles* DIABOLU, *estaubli* STABULU [Boutier et al., 1953–2006, vol. I, p. 259; Fouché, 1927, p. 83; von Wartburg, 1922, vol. VI(3), p. 1].

As argued for [ɫ] above, it may be that the F2 transition in VC sequences with a bilabial consonant is integrated as a separate glide by listeners and that the bilabial is eliminated from the resulting complex coda cluster at a later stage, i.e., [VβC] > [VwβC] > [VwC]. The prominence of the vowel transitions may increase with formant frequency and temporal variations. F2 lowering should increase the frequency extent of the VC transitions, thus rendering them more [w]-like. As for the temporal characteristics, research exploring the acoustic differences between [b] and [w] have found a greater F1 and F2 transition duration, and perhaps a smaller relative amplitude change in the vicinity of consonantal release, for [w] than for [b] [Diehl and Walsh, 1989; Liberman et al., 1956; Mack and Blumstein, 1983; Miller and Liberman, 1979; Nittrouer and Studdert-Kennedy, 1986; Shinn and Blumstein, 1984]. In agreement with this finding, spectrographic data in figure 2 show that the F2 vowel transition starts earlier for [aw] in the word ['tawlə] (about halfway through the vowel) than for [aβ] in ['taβlə] (just before the last vowel glottal pulses). Since the vowel transitions for the labiovelar glide are longer than those for the bilabial approximant, it may be hypothesized that the change [β] > [w] may be triggered by an increase in transition duration. Additional acoustic evidence for CV and VC sequences with a bilabial consonant indicates, however, that speech rate variations affect the duration of the vowel and the consonant steady-state periods rather than the duration of the vowel transitions, which is presumably in line with different degrees of anticipation of the consonant articulatory gesture. Indeed, a decrease in speech rate has been reported to lengthen the vowel transitions for [wa] from 50 to about 100–150 ms, but not those for [ba] [Gay, 1978; Miller and Baer, 1983]. Likewise, Catalan data for the word ['taβlə] produced by the present paper author reveal that slowing down the rate of speech causes some lengthening of the steady-state periods for [a], [β] and [ɫ] (from 160 to 250 ms, 65 to 140 ms and 120 to 155 ms, respectively), but not of the F2 vowel transitions for [aβ] (always about 40 ms long). Further support for variations in the timing of the vowel transitions may also be inferred from X-ray data for [apV] sequences showing some earlier vertical lip closing, but no differences in movement duration or displacement, at slower vs. faster speech rates [see fig. 2 in Gay et al., 1974].

The possible involvement of the tongue body in the production of bilabial consonants needs to be considered as well. Tongue body deactivation and F2 frequency lowering have been reported to occur during closure and even during the preceding vowel in the sequences [ipi] and [ibi] in American and British English, Swedish, German and Irish [Fuchs et al., 2004; Lindblom et al., 2002; McAllister and Engstrand, 1992; Svirsky et al., 1997; Vázquez and Hewlett, 2007]. This effect has been interpreted in terms of segment-by-segment activation [Lindblom et al., 2002], or else of active vocal tract expansion to maintain voicing for [b] and of passive pressure-driven forces to raise the intraoral pressure for [p] [Fuchs et al., 2004].

The articulatory and acoustic data just reviewed indicate that, analogously to the scenario for [ɫ] described above, [β] could be rendered more [w]-like by lowering F2 as the lips become more constricted and the tongue body is slightly depressed, and/or by anticipating the vowel transitions as these articulatory maneuvers occur earlier in time.

The role of syllable position in consonant vocalization needs to be addressed. The possibility that vocalization may affect [β] and other approximants in syllable-initial tautosyllabic clusters (dialectal Spanish [aw'laɾ] for [a'βlaɾ] 'to speak', ['majɾe] for ['maðɾe] 'mother') is not at ease with the notion that C1 should undergo articulatory reduction and contact loss provided that it is located in syllable-final position [Recasens, 2004]. In order to account for this discrepancy, it has been argued that consonant clusters with [β, ð, ɣ] followed by [l, ɾ] must have undergone a resyllabification process rendering C1 syllable-final for this consonant to become confusable with [w] or [j] [Lipski, 1994; Malmberg, 1971, p. 409]. Independently of whether resyllabification has occurred or not in the Spanish dialects where [β] vocalization has taken place, consonant vocalization may possibly operate syllable-initially in line with the fact that consonants exhibit similar, though not identical, steady-state and vowel transition acoustic characteristics in the two syllable positions [Steriade, 1988]. This option will be explored by eliciting whether those acoustic characteristics which may contribute to the generation of a glide from syllable-final [ɫ] act as potential glide generation cues for [β] in syllable-initial position, which is where the intervocalic word-medial cluster [βl] occurs invariably in Catalan (see section 1).

1.3. Alveolopalatal Nasal

The replacement of [ɲ] by [jn] should result from the categorization of the F2 and F3 VC transitions for [ɲ] as [j] by listeners, mostly so next to low and back rounded vowels where these transitions rise towards 2,000 Hz (F2) and 3,000 Hz (F3) and exhibit a large frequency extent. In contrast to [w] insertion in [VɫC] sequences (see

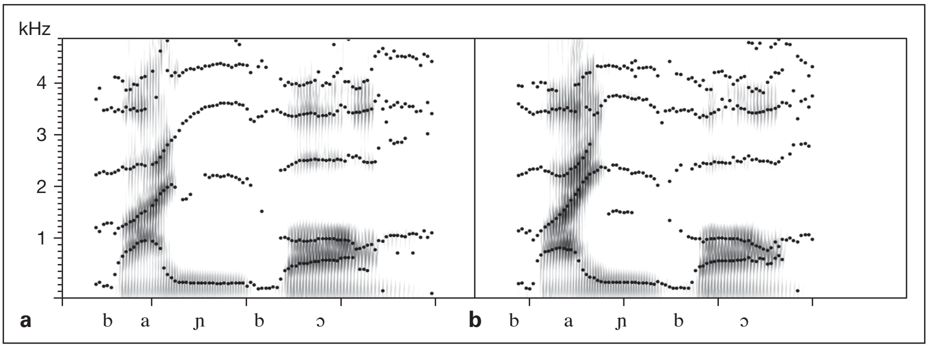

Fig. 3. Spectrographic displays for the Catalan two-word sequence [baɲ ˈbɔ] produced by speaker DR. The alveolopalatal nasal [ɲ] has been realized with a more prominent [j]-like component in the sequence displayed on the right than in the sequence displayed on the left. LPC trajectories have been overlaid on the formant displays.

section 1.1.2), [j] insertion may occur at the regressive and progressive levels, which is in accordance with alveolopalatal and palatal consonants exerting anticipatory and carryover effects on the adjacent vowels (Old French [sẽjˈɲeur] SENIORE 'gentleman', [ʃjɛvr] CAPRA 'goat') [Pope, 1934, pp. 161, 163]. Analogous [j] insertion cases may also take place before other consonants involving tongue body raising, i.e., alveolars and velars (dialectal Occitan [tɾejs] TRES, Sardinian [ˈkujlpa] CULPA, American English [læɪg] *leg*, [bæɪg] *beg*) [Contini, 1987, pp. 370–372; Lass, 1980, p. 38; Mowrey and Pagliuca, 1995, p. 102; Ronjat, 1930, vol. I, p. 366].

Variations in acoustic prominence of the vowel transitions in sequences with alveolopalatal consonants have been previously investigated in the literature [Recasens et al., 1995]. Data for Eastern Catalan show that, in comparison to the VC formant transitions for [aɲa], those for [aɲ] are longer (100–160 ms for [aɲ], 70–115 ms for [aɲa]) and exhibit larger frequency ranges (525–950 Hz for [aɲ], 450–900 Hz for [aɲa]). The presence of longer vowel transitions word-finally than intervocalically appears to be associated with an earlier onset of the dorsopalatal contact trajectory and with a greater temporal delay between the alveolar and dorsopalatal contact maxima, while analogous position-dependent differences in F2 transition endpoint frequency and frequency range follow from a larger tongue dorsum contact size in the former word position than in the latter. Differences in prominence of the F2 vowel transition are illustrated in figure 3 by two spectrographic displays for two consciously produced tokens of the Catalan sequence [baɲ ˈbɔ] 'good bath'. Longer and steeper F2 vowel transitions in the spectrogram displayed in figure 3b than in the one displayed in figure 3a result from the differences in the temporal course of events and in dorsopalatal contact size referred to above. Analogous acoustic variations may be induced by changes in speech rate: the Catalan word [aɲ] 'year' was found to lengthen the vowel steady-state period and the F2 vowel transition (from 60 to 90 ms and 120 ms, respectively), and to increase the F2 transition frequency range (from 1,400–1,850 to 1,400–2,030 Hz), at slower vs. faster speech rates.

In summary, segmental decomposition of the alveolopalatal consonant [ɲ] into [jn] in syllable-final position appears to be triggered by an increase in duration and/or frequency extent of the vowel transitions. This decomposition process should cause [ɲ] to be perceived as [Vjɲ] at first, and as [Vjn] later on presumably through some sort of perceptual differentiation between the spectrally similar phonetic segments [j] and [ɲ].

1.4. Summary

This paper investigates the relative perceptual role of spectral and temporal characteristics for the vocalization of syllable-final [ɫ] and of syllable-initial [β], and for [j] insertion before syllable-final [ɲ]. It is hypothesized that [ɫ] and [β] vocalization may result from a lowered F2 and/or by anticipating the vowel transitions, while [ɲ] decomposition may be triggered by an increased duration and/or an increase in the frequency extent of the vowel transitions.

This research should throw light on the intermediate processes occurring during the sound changes of interest. Thus, [ɫ] and [β] vocalization is prone to be a one-step process if cued by the spectral characteristics at the steady-state consonant period ([VC] > [Vw]), or as a two-stage period if cued by both the timing of the vowel transitions and the spectral properties of the steady-state consonant period ([VC] > [VwC] > [Vw]). Another research issue is whether the percentage of [w] and [jn] identification responses should increase gradually with incremental spectral and/or temporal changes or else categorically once a specific spectral and/or vowel transition configuration is achieved. Thus, for example, the identification of [ɫ] as [w] could increase continuously with successive F2 lowering steps, or else listeners could shift categorically from one phoneme to the other once a specific F2 frequency is reached. If the former is the case, there would be a good number of ambiguous stimuli not sounding quite as either [ɫ] or [w].

2. Method

The perceptual role of spectral and temporal characteristics for the vocalization and glide insertion processes of interest was tested by means of identification tests with synthetic speech stimuli. Essentially, the consonant spectral characteristics and the time of occurrence of the vowel transitions were manipulated in the stimulus continua. The aim of the study was to gain insight about the perceptual role of the two cues independently and in combination in glide generation. The synthetic stimuli were created based on data from the Catalan language, whose sound inventory includes all five consonants [ɫ], [β], [ɲ], [w] and [j].

2.1. Dark Alveolar Lateral

The acoustic data for the preparation of the perception test for [ɫ] were obtained from realizations of this consonant flanked by vowels and consonants causing it to exhibit a considerably low F2 frequency, i.e., after [ə, o] and before [b, d, s, k]. The four sequences [oɫb], [əɫd], [əɫs] and [əɫk] were embedded in the meaningful Majorcan Catalan sentences [lakə moɫ ˈbɔnə] *laca molt bona* 'very good varnish', [dɛw səɫ dodˈdzə] *deu ser el dotzè* 'it must be the twelfth', [trɛw əɫ suˈk ultim] *treu el suc últim* 'he/she pull out the last juice' and [mənʒət əɫ ˈkraŋk] *menja't el cranc* 'eat the crab'. The speech material was recorded 7 times by the 4 Majorcan Catalan speakers BM, MJ, ND and CA with an artificial palate in place. Formant frequency values and linguopalatal contact patterns for these recordings have been reported elsewhere [Recasens, 2009]. Analysis data for [ɫ] were gathered from Majorcan Catalan speakers because the alveolar lateral is strongly dark and, therefore, especially prone to be confused with [w] in this dialect.

The acoustic signal was sampled at 10 kHz. For each sequence and speaker, those tokens exhibiting the highest and lowest F2 frequency at the [ɫ] midpoint were selected for further acoustic analysis. F1, F2 and F3 frequency trajectories were tracked during the vowel steady-state period, the vowel transitions and the [ɫ] steady-state period using LPC with a 16 filter order and a 20-ms window. Unreliable formant frequency estimates were corrected by hand.

The formant frequency trajectories for the selected tokens of each sequence were averaged across speakers by lining them up at [ɫ] onset, and the onset and offset of the resulting average trajectories

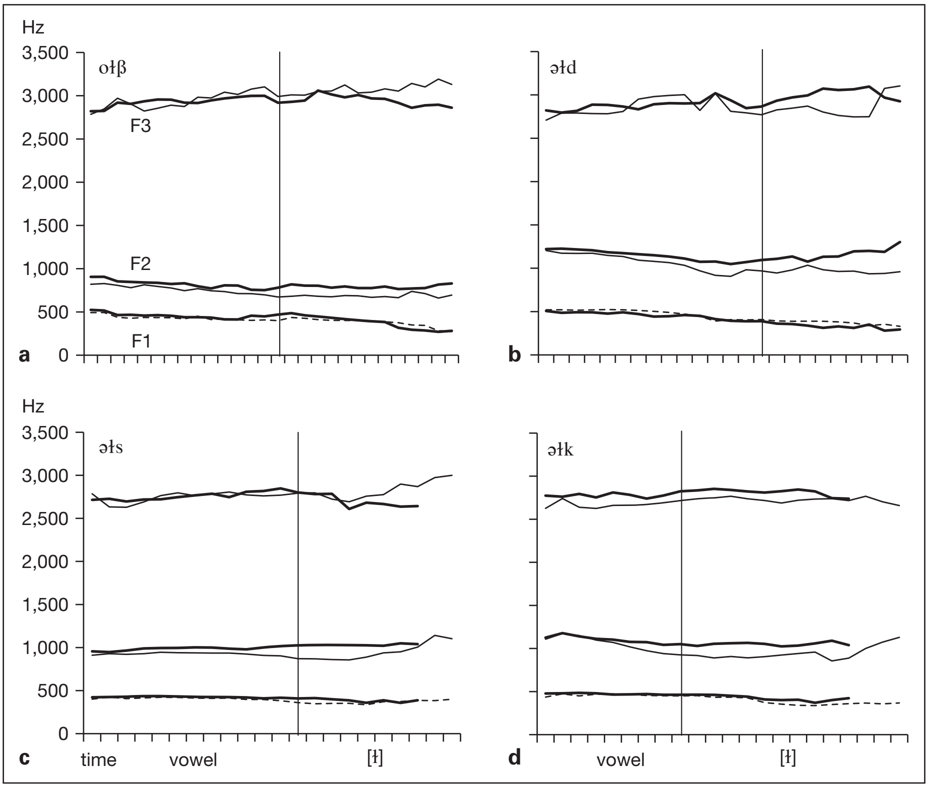

Fig. 4. [Vɫ] formant trajectories for the high (thick trace) and low (thin trace) formant frequency conditions. Data correspond to the sequences [oɫβ] (**a**), [əɫd] (**b**), [əɫs] (**c**) and [əɫk] (**d**). The division mark has been set at the temporal boundary between the acoustic periods for the vowel and for [ɫ]. Temporal frames are 5 ms long.

were assigned the same duration as those of at least two of the speakers' trajectories. According to figure 4, average formant trajectories for [oɫβ, əɫd, əɫs, əɫk] across speakers differ to a larger extent in F2 than in F1 and F3, thus indicating that differences in darkness degree for [ɫ] may be modeled by varying F2 alone. Moreover, F2 frequency differences extend back to V1 onset, which reflects anticipatory effects in tongue dorsum height and fronting for the consonant during the preceding vowel. In the light of this experimental evidence, the high and low F2 frequency values across speakers were selected as the endpoints of the continuum of perception stimuli.

The synthetic speech stimuli were prepared with the Analysis Synthesis Laboratory (ASL) program of Kay Pentax. ASL makes it possible to edit and resynthesize the formant frequency, intensity and duration values obtained through LPC analysis of real speech data. The editing procedure was carried out on the LPC formant frequency trajectories of a token of the meaningful word ['aɫβə] 'dawn' produced by speaker DR (the first paper author) of Eastern Catalan, where [ɫ] is moderately dark and the vowel [a] is neither clearly front nor back as a general rule. Three events were identified for that purpose, i.e., the vowel steady-state period, the vowel transitions and the [ɫ] steady-state period. The acoustic boundary between the two former events was determined as accurately as possible at the temporal point where F2 starts moving downwards along the frequency scale, and that between the two latter events was placed at F2 transition offset.

Figure 5 reports the F1, F2 and F3 frequency values and the transition onset/offset times for all stimuli conditions. As shown by figure 5a, F2 at the [ɫ] steady-state period was varied from 650 up to

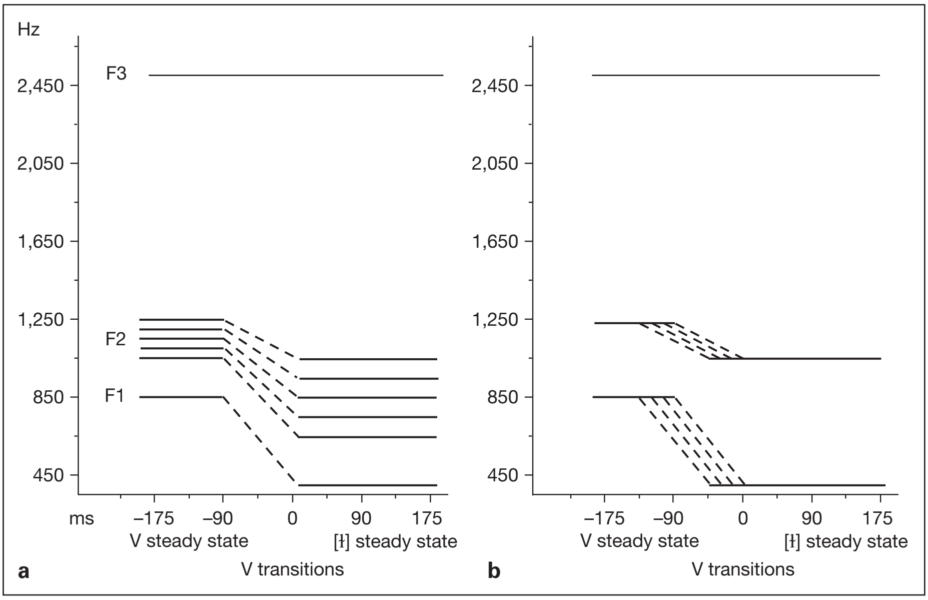

Fig. 5. Parameter values for the [ɫ] perception stimuli. **a** Formant frequencies at the consonant steady-state period (illustrated for the vowel transitions 0 ms offset time condition). **b** Vowel transition onset/offset times (illustrated for the 1,050-Hz F2 frequency condition).

1,050 Hz in 100-Hz steps, thus yielding 5 stimuli. In order to match the consonant-to-vowel effects observed on the spectrographic displays, this formant frequency was varied simultaneously at the vowel steady-state period from 1,050 to 1,250 Hz in 50-Hz steps. F1 was kept fixed at 850 Hz during the vowel and at 400 Hz during the consonant, and the F3 frequency was 2,500 Hz throughout the entire VC sequence. As shown in figure 5b, the F2 and F1 vowel transitions were taken to end at four different temporal points, i.e., at the onset of the [ɫ] steady-state period (0 ms condition) and at 15, 30 and 45 ms before that temporal point (–15, –30 and –45 ms conditions). Since the vowel transitions lasted invariably 90 ms, their onset turned out to vary in 15-ms steps as well. The duration of the vowel and the consonant steady-state periods ranged, respectively, between 55 and 100 ms and between 180 and 225 ms depending on the vowel transition onset/offset time.

In order to test the relative contribution of the two parameters to the identification of [ɫ] as [w], each transition condition was combined with each [ɫ] steady-state period condition, thus yielding 20 combinations. The perception test included 90 tokens, i.e., 4 tokens of the basic 20 combinations, plus 5 tokens of a synthesized version of an original production of the words ['aɫβə] 'dawn' and ['awβə] 'poplar' produced by speaker DR, which could serve as reference of an acceptable [ɫ] and an acceptable [w]. Another reason for including several tokens of [w] in the test was to make sure that listeners would not be strongly biased in favor of the [ɫ] identification responses.

The synthetic stimuli were randomized and presented for identification using Power Point on a laptop computer with a highly acceptable 16-bit sound card. Seventeen informants with Catalan as the first language took the test without headphones in groups of 3 in a sound-treated room at the phonetics laboratory of the Universitat Autònoma of Barcelona. They were asked to identify the perception stimuli as either /l/ or /w/ by writing 'L' or 'U' on an answer sheet. The graphic symbol 'W' was not chosen since in written Catalan the grapheme 'w' is only found in words of foreign origin. The option /wl/ (i.e., 'UL') was not given since the segmental combination [VwɫC] is not possible in Catalan and also because having three options would have rendered the identification task harder than it actually was. It was soon found that several subjects could not hear [ɫ] as [w] perhaps since not all acoustic cues which could contribute to the misidentification between the two sounds had been subject to manipulation

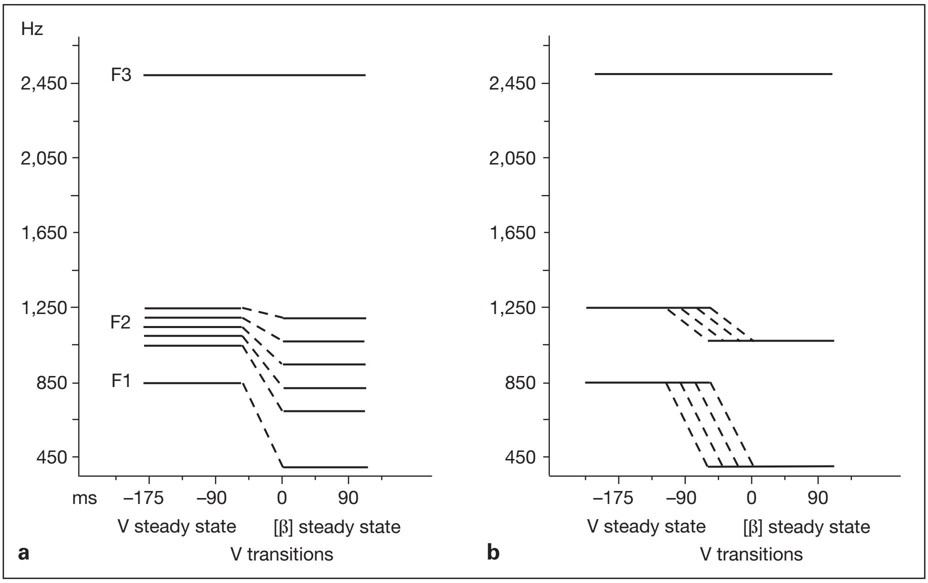

Fig. 6. Parameter values for the [β] perception stimuli. **a** Formant frequencies at the consonant steady-state period (illustrated for the vowel transitions 0 ms offset time condition). **b** Vowel transition onset/offset times (illustrated for the 1,075-Hz F2 frequency condition).

during the preparation of the synthetic stimuli (see 'Discussion'). For this reason, our informants were instructed to mark as 'U' not only those stimuli which were heard as [w] but also those which sounded [w]-like and, therefore, could be confused with [w] in an informal speech situation.

2.2. Bilabial

Data for creating the perception stimuli for testing [β] vocalization were taken from LPC formant frequency trajectories for the [aβ] portion of the Catalan word ['taβɫə] 'board' recorded by speaker DR at different speech rates. F2 was found not to vary much across rate conditions whether during the steady-state portion of the vowel (1,200–1,300 Hz), [β] (950–1,100 Hz) or [ɫ] (900–950 Hz), or at vowel offset (1,050–1,100 Hz).

The editing task was performed with the ASL program on a token of the word ['taβɫə] produced by the same speaker. Three events were identified for manipulation applying the same criteria described in section 2.1: the vowel steady-state period, the vowel transitions and the [β] steady-state period. Formant frequency changes were performed based on the ranges for speaker DR's productions and on spectral data in the literature. As shown in figure 6, the stimulus continuum was prepared by varying F2 at the [β] steady-state period from 700 to 1,200 Hz in 125-Hz steps (5 stimuli; fig. 6a) and the offset time of the F2 and F1 vowel transitions from 0 ms at vowel offset to 20, 40 and 60 ms before vowel offset (4 stimuli; fig. 6b). Since the transition duration was always 50 ms, the onset of the vowel transitions varied in 20-ms steps as well. Simultaneous F2 frequency variations during the vowel were also performed from 1,050 to 1,250 Hz in 50-Hz steps. The F1 frequency was 850 Hz during the vowel steady-state period and 400 Hz during the consonant, and the F3 frequency was 2,500 Hz throughout the entire [aβ] sequence. The duration of the vowel steady-state period ranged from 125 to 170 ms and that of [β] from 60 to 115 ms depending on the transitions onset/offset time, while [ɫ] was always 120 ms long. All possible combinations of the 5 consonant steady-state and 4 transition conditions yielded an overall number of 20 stimuli.

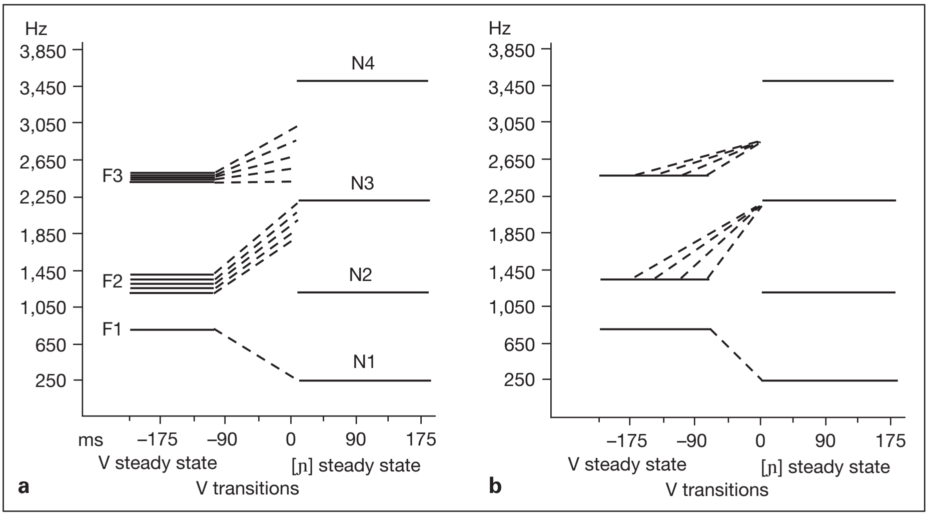

Fig. 7. Parameter values for the [ɲ] perception stimuli. **a** Formant transition endpoint frequencies (illustrated for the vowel transitions –110 ms onset time condition). **b** Vowel transition onset times (illustrated for the 2,200-Hz F2 transition endpoint condition).

The perception test was composed of 90 stimuli, i.e., 4 tokens of the basic 20 stimuli plus 5 tokens of synthesized versions of an original production of ['taβɫə] 'board' and ['tawɫə] 'table' recorded by speaker DR. Several tokens of [w] were included in order to prevent the informants from being strongly biased in favor of the [β] identification responses. Stimuli were randomized and presented for identification under the same conditions used in the perception experiment for [ɫ] vocalization. Eighteen native speakers of Catalan took the test in groups of 3 in a sound-treated room without headphones. They were asked to identify the stimuli as /b/ or /w/ by writing 'B' or 'U' on an answer sheet. The option /wb/ (i.e., 'UB') was not given since the combination [Vwβɫ] does not occur in Catalan.

2.3. Alveolopalatal Nasal

A set of stimuli which could yield different degrees of a [j] percept in the sequence [aɲ] were created using the ASL program. The editing procedure was performed on LPC formant frequency trajectories for an acoustic recording of the Catalan two-word sequence [baɲ 'bɔ] 'good bath' uttered by speaker DR. Two events were identified for manipulation, i.e., the vowel steady-state period and the vowel transitions. The nasal murmur was left unmodified due to its complex acoustic structure and also because it does not seem to be a highly relevant place of articulation cue for nasal consonants [Malécot, 1956; Recasens, 1983]. The vowel [a] was chosen since the [ɲ] transitions are particularly long and exhibit a large frequency extent and a high intensity level in this particular vowel context [Recasens and Martí, 1990].

The perception stimuli were prepared by varying the duration and endpoint frequency values of the F2 and F3 vowel transitions. As shown in figure 7a, the endpoint frequency of the vowel transitions was varied from 1,800 to 2,200 Hz in 100- Hz steps (F2) and from 2,400 to 3,000 Hz in 150-Hz steps (F3). F2 and F3 frequency changes were also performed at the vowel steady-state portion from 1,200 to 1,400 Hz in 50-Hz steps (F2) and from 2,400 to 2,500 Hz in 25-Hz steps (F3). As can be seen in figure 7b, the onset time of the vowel transitions occurred at 70, 110, 140 and 170 ms before vowel offset, and ended at the acoustic boundary between the vowel and the nasal murmur. Since the vowel was 215 ms long, the duration of the vowel steady-state period ranged from 45 to 145 ms depending on the onset time of the vowel transitions. F1 was located at 800 Hz during the vowel and had a lowering transition ending at 250 Hz at vowel offset. The nasal murmur was 370 ms long (though it shows a shorter duration in fig. 7) and its formant frequencies were kept fixed at 250 Hz (N1), 1,200 Hz (N2), 2,200 Hz (N3) and 3,500 Hz (N4) [Recasens, 1983].

The perception test was composed of 90 stimuli. There were 4 tokens of the basic 20 stimuli resulting from all possible combinations of the 5 transition endpoint frequencies and the 4 transition onset times ranging from more [ɲ]-like to more [jɲ]-like. The test also included 5 tokens of synthesized versions of an original production of the sequences [baɲ ˈbɔ] 'good bath' and [ban ˈbɔ] 'good proclamation' recorded by speaker DR. The rationale for having a few tokens of [ban ˈbɔ] in the test was to make sure that listeners would not be strongly biased in favor of the /jɲ/ vs. /ɲ/ identification responses. Stimuli were randomized and presented for identification using Power Point. Seventeen native Catalan-speaking informants in groups of 3 without headphones were asked to indicate whether they perceived a glide or not by identifying the stimuli as 'INY' or as plain 'NY', where the digraph 'NY' corresponds to the orthographic representation of the consonant [ɲ] in Catalan. They were also told to categorize the [n]-like stimuli as 'NY' or as 'N'. Only the 'INY' responses were coded as identifying /j/; both responses 'NY' or as 'N' were computed as instances of no glide insertion.

This sound identification task differs from the identification of [ɫ] and [β] as /w/ in that, while [ˈaɫβə] and [ˈawβə] as well as [ˈtaβɫə] and [ˈtawɫə] are Catalan words, [bajɲ ˈbɔ] is just a possible realization of /bajɲ ˈbɔ/ in Catalan. It is the case, however, that differences in degree of palatality in the alveolopalatal nasal consonant may be distinctively perceived in a similar fashion to differences in degree of darkness in /l/. It is indeed likely that the informants who took the perception test are exposed to variants of /ɲ/ exhibiting different degrees of the palatal on-glide in their everyday life given that there are speakers of specific Catalan dialectal areas who produce /ɲ/ with a high degree of palatality and also that, as indicated in section 1, Majorcan has a phonological rule transforming preconsonantal /ɲ/ into [jn]. It is believed that listeners who did well in the test marked the option INY whenever they heard a clear chirp corresponding to long and steep F2 and F3 vowel transitions.

2.4. Statistics

Repeated measures ANOVAs were performed on the /w/ identification scores for the [ɫ] and [β] stimuli, and on the /jɲ/ identification scores for the [ɲ] stimuli. Factors were 'formant frequency' for all tests (5 levels), and 'transition offset time' for the tests for [ɫ] and [β] and 'transition onset time' for the test for [ɲ] (4 levels). Even though the onset and offset of the F1 and F2 transitions were varied simultaneously for [ɫ] and [β], we just entered the offset time values in the statistical analysis; likewise, the F2 but not F3 data were analyzed statistically in the case of the synthetic continuum for the alveolopalatal nasal. The experimental units of the ANOVAs were the identification scores elicited by the 17 or 18 informants who took the perception tests. Scores ranged from 0 to 4 since there were four tokens for each stimulus, such that 0, 1, 2, 3 and 4 corresponded to 0, 25, 50, 75 and 100% of /w/ or /jɲ/ identification responses. Huynh-Feldt corrected degrees of freedom were applied to the main effects and interactions whenever the sphericity requirement was not fulfilled. Bonferroni post-hoc tests were run on those factor values yielding a significant main effect at the $p < 0.05$ level of significance. Significant interactions were explored by performing one-way RM ANOVAs on values for all levels of a given factor while keeping each level of the other factor constant.

3. Results

3.1. Dark Alveolar Lateral

Two-way RM ANOVAs for the [ɫ] identification test yielded a highly significant 'formant frequency' effect [$F(1.8, 28.5) = 19.62$, $p = 0.000$], a moderately significant

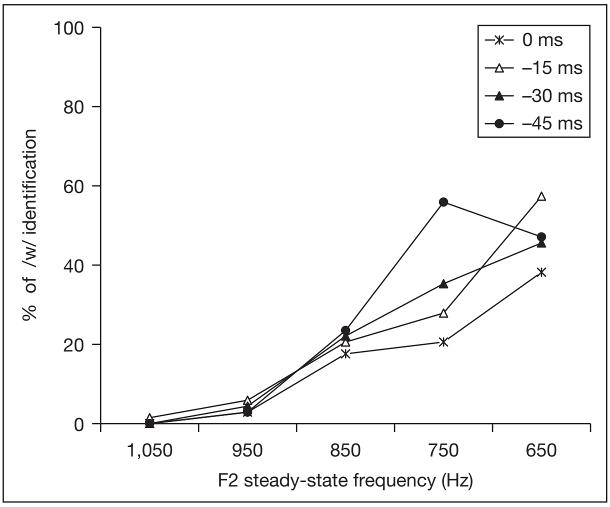

Fig. 8. Percentages of /w/ identification for the 20 stimuli of the [ɫ] perception test plotted as a function of F2 steady-state frequency and transition offset time (in milliseconds).

'transition offset time' effect [F(3.8, 44.4) = 3.55, p = 0.024], and a significant two-factor interaction [F(8.6, 137.4) = 3.85, p = 0.000]. The F and p values just given suggest that, while both acoustic characteristics play a role in the identification of [ɫ] as /w/, the F2 frequency at the consonant steady-state period is a more effective cue than the offset time of the vowel transitions. Informants accurately identified the unmodified original productions of [ɫ] and [w], i.e., the percentages of /w/ identification were 5.9% for the former and 100% for the latter.

Identification percentages for the 20 stimuli of the perception continuum are displayed in figure 8. According to this figure, the two highest 1,050-Hz and 950-Hz F2 frequency conditions yielded practically no /w/ identification responses irrespective of differences in the onset and offset time of the vowel transitions. Moreover, percentages of /w/ identification responses increase to about 20, 35 and 50% as F2 shifts down to 850 Hz, 750 and 650 Hz, respectively. As indicated in table 1a, /w/ identification responses turned out to be significant for all F2 frequency pairs except for those involving adjacent values at the highest end of the continuum (1,050/950 Hz) and at the lowest end (750/650 Hz). Moreover, the /w/ identification curves plotted in figure 8 suggest that the contribution of F2 to the identification of [ɫ] as /w/ is partly categorical (/w/ identification responses occur provided that F2 lies below 950 Hz) and partly gradual (/w/ identification responses increase gradually as F2 lowers from 850 Hz down to 650 Hz).

Post-hoc tests for the 'transition offset time' factor yielded no significant effects (table 1b). Indeed, as revealed by the /w/ identification curve displayed in figure 8, the percentage of /w/ identification responses does not always improve as the vowel transitions are anticipated in time and thus, their offset occurs at 0, –15, –30 and –45 ms.

Let us now turn to the significant 'formant frequency' × 'transition offset time' interaction. One-way RM ANOVAs yielded a significant transition offset time effect when F2 exhibits a low frequency at the [ɫ] steady-state period, i.e., 750 Hz [F(2.6, 41.1) = 10.63, p = 0.000] and also 650 Hz through significance was not reached in this case [F(2.8, 45) = 2.84, p = 0.051]. As shown in table 2a, significant differences occur

Table 1. p significance level for pairwise comparisons performed on statistical factors yielding a significant main effect (p values above 0.05 are left blank)

a [ɫ] test — Consonant steady-state F2 frequency (Hz)

	1,050	950	850	750	650
1,050			0.015	0.002	0.001
950			0.041	0.004	0.003
850				0.030	0.012
750					
650					

b [ɫ] test — Vowel transitions offset time (ms)

	0	–15	–30	–45
0				
–15				
–30				
–45				

c [β] test — Consonant steady-state F2 frequency (Hz)

	1200	1075	950	825	700
1,200			0.000	0.000	0.000
1,075			0.000	0.000	0.000
950				0.000	0.000
825					0.000
700					

d [β] test — Vowel transitions offset time (ms)

	0	–20	–40	–60
0		0.000	0.000	0.000
–20			0.000	0.000
–40				0.000
–60				

e [ɲ] test — F2 transition endpoint frequency (Hz)

	1800	1900	2000	2100	2200
1,800			0.006	0.011	0.003
1,900				0.005	0.002
2,000					0.032
2,100					0.050
2,200					

f [ɲ] test — Vowel transitions onset time (ms)

	–70	–110	–140	–170
–70			0.040	0.034
–110				0.039
–140				
–170				

between the –45 ms and all other transition offset times in the case of the 750-Hz F2 frequency condition. In agreement with these statistical results, identification percentages plotted in figure 8 show a trend for the percentage of /w/ identification responses to increase gradually as the vowel transitions occur earlier in time provided that F2 exhibits a low frequency value mostly at 750 Hz (from 20.6 to 55.9%) and, less so, at 850 Hz (from 17.6 to 23.5%).

One-way RM ANOVAs yielded a significant 'formant frequency' effect for all four transition offset times at the p = 0.000 level, and an increase in F value with the degree of temporal anticipation of the vowel transitions, i.e., 9.62, 12.94, 14.26 and 19.05 for transition offsets at 0, –15, –30 and –45 ms, respectively. As revealed by table 2d, the number of significant effects and, therefore, the degree of perceptual sensitivity to F2 frequency variations, is greater for the early as opposed to late vowel transition offset times. Differences between the lowest and the highest F2 frequency values are generally significant for all four transition offset time conditions, while smaller F2 differences achieve significance only whenever one of the F2 frequency values subjected to comparison is relatively low.

In summary, the perception of [ɫ] as /w/ is favored by a low F2 frequency at the consonant steady-state period and, to a lesser extent, by an early onset and offset time of the preceding vowel transitions. While both acoustic properties appear to reinforce each other, the former characteristic contributes more than the latter to the identification of [ɫ] as /w/.

Table 2. p significance level for pairwise comparisons performed on significant interactions between formant frequency (in Hz) and transition onset or offset time (in ms) (Top: transition effects for each formant frequency condition; bottom: formant frequency effects for each transition condition, p values above 0.05 are left blank)

[ɫ] test

a

Hz	ms	0	–15	–30	–45
1,050	0				
	–15				
	–30				
	–45				
950	0				
	–15				
	–30				
	–45				
850	0				
	–15				
	–30				
	–45				
750	0				0.004
	–15				0.013
	–30				0.009
	–45				
650	0				
	–15				
	–30				
	–45				

[β] test

b

Hz	ms	0	–20	–40	–60
1,200	0				
	–20				
	–40				
	–60				
1,075	0				
	–20				
	–40				
	–60				
950	0			0.001	0.000
	–20			0.004	0.004
	–40				
	–60				
825	0		0.047	0.000	0.000
	–20			0.000	0.000
	–40				
	–60				
700	0		0.000	0.000	0.000
	–20				
	–40				
	–60				

[ɲ] test

c

Hz	ms	–70	–110	–140	–170
1,800	–70				
	–110				
	–140				
	–170				
1,900	–70				
	–110			0.010	
	–140				
	–170				
2,000	–70				
	–110				
	–140				
	–170				
2,100	–70				0.001
	–110				0.035
	–140				0.003
	–170				
2,200	–70			0.052	
	–110				
	–140				
	–170				

d

Hz	ms	1,050	950	850	750	650
0	1,050				0.008	0.004
	950					
	850					
	750					
	650					
–15	1,050			0.040	0.005	0.002
	950					0.039
	850					
	750					
	650					
–30	1,050				0.008	0.003
	950				0.021	0.006
	850					0.048
	750					
	650					
–45	1,050				0.001	0.001
	950			0.031	0.001	0.001
	850					
	750					
	650					

e

Hz	ms	1,200	1,075	950	825	700
0	1,200					0.006
	1,075					0.006
	950					0.006
	825					0.006
	700					
–20	1,200				0.035	0.000
	1,075					0.000
	950					0.000
	825					0.000
	700					
–40	1200			0.001	0.000	0.000
	1,075			0.003	0.000	0.000
	950				0.000	0.000
	825					
	700					
–60	1,200			0.001	0.000	0.000
	1,075			0.000	0.000	0.000
	950				0.000	0.000
	825					
	700					

f

Hz	ms	1,800	1,900	2,000	2,100	2,200
–70	1,800					
	1,900					
	2,000					
	2,100					
	2,200					
–110	1,800					0.013
	1,900				0.010	0.000
	2,000					0.020
	2,100					
	2,200					
–140	1,800					0.003
	1,900					0.007
	2,000					0.041
	2,100					0.001
	2,200					
–170	1,800			0.003	0.002	0.003
	1,900				0.005	0.009
	2,000					
	2,100					
	2,200					

3.2. Bilabial

Two-way RM ANOVAs for the identification test of [β] as /w/ yielded a highly significant effect of 'formant frequency' [F(2.8, 48.2) = 202.48, p = 0.000] and 'transition offset time' [F(2.6, 44.9) = 119.39, p = 0.000], and a significant two-factor interaction [F(4.9, 83.52) = 25.25, p = 0.000]. F and p values, and identification percentages, indicate that informants were more prone to identify [β] than [ɫ] as /w/. In parallel to the results for the perception test for [ɫ], the consonant steady-state spectral characteristics were a more effective vocalization cue than the degree of temporal anticipation of the vowel transitions. The unmodified original stimuli were correctly identified by our informants, i.e., the percentage of /w/ identification responses were 0% for the synthesized original version of ['taβɫə] and 99% for the synthesized original version of ['tawɫə].

As shown by figure 9, the two highest F2 frequency conditions, 1,200 Hz and 1,075 Hz, yielded virtually no /w/ identification responses irrespective of differences in the offset time of the vowel transitions. Moreover, the percentage of /w/ identification responses averaged across transition offset times was found to increase to about 25, 50 and 80% as F2 lowers down to 950, 825 and 700 Hz, respectively. Differences in F2

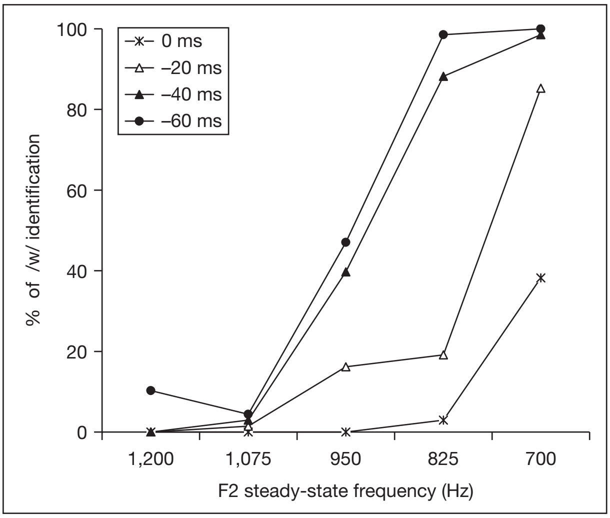

Fig. 9. Percentages of /w/ identification for the 20 stimuli of the [β] perception test plotted as a function of F2 steady-state frequency and transition offset time (in milliseconds).

frequency turned out to be significantly different in all cases except for the 1,200/1,075-Hz pair (table 1c). Analogously to results for the [ɫ] perception test, the contribution of F2 to the identification of [β] as /w/ appears to be both categorical (confusions occur provided that F2 does not exceed 1,000 Hz) and gradual (the percentage of /w/ identification responses increases gradually as F2 lowers from 950 Hz down to 700 Hz).

According to figure 9, percentages of /w/ identification increased with the degree of anticipation of the vowel transitions most of the time. When averaged across F2 frequency conditions they amounted to 10% when the transition offset time occurs at 0 ms, 25% (–20 ms), 45% (–40 ms) and 50% (–60 ms). Such an increase is more prominent than that obtained for [ɫ], thus meaning that the contribution of the vowel transitions to [w] identification was greater for [β] than for the alveolar lateral. All transition offset values turned out to be significantly different from each other thus meaning that the temporal occurrence of the vowel transitions is a robust [β] vocalization cue (table 1d).

As for the two-factor interaction (table 2b), differences in transition offset time achieved significance when the F2 frequency for [β] is particularly low, i.e., 950, 825 and 700 Hz ($p = 0.000$) and, less so, 1,200 Hz ($p = 0.037$), but not 1,075 Hz. Significant post-hoc effects occurred between the most extreme transition offset values, i.e., basically between 0 ms and –20, –40 and –60 ms and, less so, between –20 ms and –40 and –60 ms, but not for the 1,200-Hz condition. The /w/ identification curves in figure 9 indicate a clear increase in the percentage of /w/ identification responses for the three lowest F2 frequencies as the vowel transitions retract from 0/–20 ms to –40/–60 ms.

F2 differences at the [β] steady-state period were highly significant for all four transition offset times ($p = 0.000$), and the F value increased with the degree of temporal anticipation of the vowel transitions (F equals 17.44, 76.03, 134.35 and 113.15 for the 0, –20, –40 and –60 ms transition offset times, respectively). According to post-hoc tests, significant effects occurred for most F2 frequency pairs in the case of the two early transition offset times (–40, –60 ms), and basically between the lowest (700 Hz) and all other F2 frequency values for the two late transition offset times (0 ms, –20 ms) (table 2e, fig. 9).

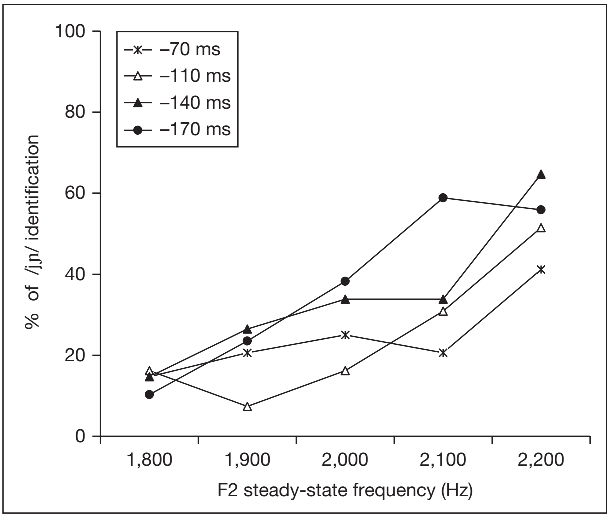

Fig. 10. Percentages of /jɲ/ identification for the 20 stimuli of the [ɲ] perception test plotted as a function of F2 transition endpoint frequency and transition onset time (in milliseconds).

In summary, the perception of [β] as /w/ appears to be favored by the presence of low F2 frequency values and an early occurrence of the vowel transitions. Both cues play a relevant role in /w/ identification.

3.3. Alveolopalatal

Two-way RM ANOVAs performed on the /jɲ/ identification responses yielded a highly significant effect of 'formant frequency' [$F(2.07, 33.14) = 15.55$, $p = 0.000$], a somewhat weaker effect of 'transition onset time' [$F(3, 48) = 6.87$, $p = 0.001$], and a significant two-factor interaction [$F(9.2, 147.25) = 3.54$, $p = 0.001$]. It thus appears that, while both acoustic cues may promote glide insertion perceptually, glide insertion depends more on the frequency endpoint value than on the onset time of the F2/F3 vowel transitions. The /jɲ/ identification scores for the unmodified original versions were successful for [ban ˈbɔ] (0%), while those for [baɲ ˈbɔ] were split between /ɲ/ and /jɲ/ responses (64.7 and 35.3%, respectively), which may be taken to indicate that listeners had a harder time in telling whether /ɲ/ had a glide or not than in identifying original productions of [ɫ], [β] and [w].

As shown in figure 10 and in contrast with results for the perception tests for [ɫ] and [β], the lowest transition endpoints at 1,800 and 1,900 Hz yielded a moderate number of /jɲ/ identification responses. Moreover, percentages of /jɲ/ identification turned out to increase gradually with an increase in the endpoint frequency of the vowel transitions, i.e., they amount to 15, 20, 28, 35 and 55% for F2 transition endpoint frequencies at 1,800, 1,900, 2,000, 2,100 and 2,200 Hz, respectively. Percentages achieved significance if involving pairs of F2 transition endpoints which were sufficiently distant along the frequency scale, i.e., those between 2,200 Hz and 1,800, 1,900, 2,000 and 2,100 Hz, between 2,100 Hz and 1,800 and 1,900 Hz, and between 2,000 Hz and 1,800 Hz (table 1e).

Post-hoc tests for the 'transition onset time' factor also yielded significant effects between vowel transitions exhibiting distant onset times, i.e., –70 ms and –110 ms vs. –170 ms, and –70 ms vs. –140 ms (table 1f). According to figure 10, /jɲ/ identification responses may increase slightly as the F2/F3 vowel transitions occur earlier in time; /jɲ/ identification percentages were 25% for transitions beginning at –70 and –110 ms and 35–40% for those beginning at –140 and –170 ms.

Results for the two-factor interaction are presented next. According to one-way RM ANOVAs, there is a significant effect of transition onset time for most F2 transition endpoint frequencies, i.e., 1,900 Hz (p = 0.008), 2,000 Hz (p = 0.046), 2,100 Hz (p = 0.000) and 2,200 Hz (p = 0.015). Significance occurred between early and late onset times for F2/F3 transition endpoint frequency values at 2,100 Hz and, less so, for those at 1,900 Hz and 2,200 Hz, and for no pair of transition onset times in the case of the 2,000-Hz condition (table 2c). The general trend is then for differences among transition onset times to become more perceptually salient as the endpoint frequency value becomes higher (fig. 10).

One-way RM ANOVAs yielded a significant 'formant frequency' effect for all transition onset time conditions, i.e., at the p = 0.000 level for the –110 ms, –140 ms and –170 ms conditions and at the p = 0.033 level for the –70 ms condition. According to results from post-hoc tests, the degree of perceptual sensitivity to endpoint frequency variations increases as the F2/F3 transitions begin earlier. Indeed, as shown in table 2f, significant effects occur mostly between the highest F2 transition endpoint at 2,200 Hz and lower endpoint frequencies whenever the onset of the vowel transitions is located at –110, –140 and –170 ms. Significant differences between lower F2 endpoints, i.e., between 1,800 and 2,000 Hz and 2,100 Hz and between 1,900 and 2,100 Hz, were found to hold mostly for the earliest –170 ms transition onset time. Post-hoc tests yielded no significant effects in the case of the –70 ms condition.

In summary, the perception of [ɲ] stimuli as /jɲ/ is favored by high F2/F3 transition endpoint frequencies and, less so, by long vowel transitions. The percentages of /jɲ/ identification increase as higher endpoint formant frequencies co-occur with earlier transition onset times, and the former acoustic property appears to be a better glide insertion cue than the latter. This relative power of cues resembles the one obtained for the [ɫ] and, less so, the [β] perception tests.

4. Discussion

A relevant finding of the present investigation is that the formant frequency at the consonant steady-state period or at the endpoint of the vowel transitions, and the onset or onset/offset time of the transitions, contribute to [ɫ] and [β] vocalization and to glide insertion before [ɲ]. Indeed, for all sound changes under analysis, ANOVAs yielded a main effect of both factors. Moreover, the perceptual effectiveness of the formant frequency characteristics was found to exceed that of the temporal characteristics of the vowel transitions in all cases, i.e., higher significance levels and greater numbers of significant pairwise comparisons were obtained for the former factor than for the latter. This finding is in support of Ohala's hypothesis that the equivalence between the spectral steady-state configurations for [ɫ] and [β], on the one hand, and for [w], on the other hand, may cause the two former consonant realizations to be identified as the labiovelar glide by listeners [Ohala, 1974]. A novel finding is that the perceptual relevance of the onset/offset time of the vowel transitions in consonant vocalization varies as a function of the original consonant, i.e., it was found to be higher for [β] than for [ɫ] though, as referred to later on in this section, this may also be associated with different degrees of success in synthesizing the two consonants.

Taken together, both findings suggest that [ɫ] and [β] vocalization may be achieved through two mechanisms: direct identification of the spectral configuration of the original consonant with the glide, and a double-stage evolution involving glide insertion followed by simplification of the outcoming complex coda cluster. The formulation of these mechanisms assumes that the perceptual role of the vowel transitions is negligible in the former scenario and quite prominent in the latter. Based on our perception results, it appears that, while the two mechanisms may be involved, the former mechanism is more prone to apply than the latter.

Identification responses indicate that a very low F2 frequency is needed for listeners to be able to take [ɫ] and [β] for [w]. In other words, the two consonant realizations need to be quite dark for them to be heard as the labiovelar glide. Below a specific formant frequency, the percentage of /w/ identifications increases gradually with darkness degree in the consonant. On the other hand, the chances that a glide is inserted before [ɲ] increase with an increase in the F2/F3 transitions endpoint frequency and such an increase proceeds continuously independently of whether the endpoint frequency is lower (about 1,800–1,900 Hz) or higher (about 2,000–2,200 Hz). Moreover, it seems then that the frequency ranges triggering vocalization are more precisely defined than those causing glide insertion to occur.

Judging from the identification percentages obtained in the present study, our informants did better in identifying [β] with [w] than in categorizing [ɫ] as [w] and [ɲ] as [jɲ]. Moreover, this outcome turned out to be associated with both the spectral and the transition cues, while suggesting that listeners may be sensitive to acoustic properties which were not taken into consideration in the preparation of the synthetic speech stimuli. As for the contrast between [ɫ] and [w], these other acoustic cues include differences in the F3 transition (which may rise slightly for the lateral and lower for the glide; fig. 1) as well as the oral release and the presence of spectral zeroes for [ɫ]. It may be that glide identification for [aɲ] could have improved if a nasal murmur had been added to the original stimuli since it is known that the F2 and F3 transitions from the vowel to the alveolopalatal nasal may continue to rise during the initial portion of the murmur [Recasens and Martí, 1990].

Perception data for [β] reported in the present study suggest that vocalization may occur not only syllable-finally but also syllable-initially in tautosyllabic consonant clusters. As pointed out in the 'Introduction', there may be no need for C1 onsets in consonant sequences such as [βl] and [βɾ] to become codas for them to undergo vocalization (just as intervocalic [β] may vocalize in syllable-initial position in Romance; see section 1). The vocalization of syllable-initial consonant realizations appears to be often contextually determined, i.e., consonants may vocalize next to vowels exhibiting a similar quality (e.g., [ʎ] > [j] next to front vowels, [ɫ] > [w] next to [a] and back rounded vowels), and most prone to affect consonant realizations of relatively little salience such as approximants in tautosyllabic consonant clusters as in the present paper.

None of the sound changes [ɫ]> [w] and [ɲ] > [jn] needs to be attributed to segmental complexity, which assumes that glide generation occurs because the target consonant is specified for an underlying dorsal gesture. Thus, one could be tempted to view the segregation of [w] out of [ɫ] and of [j] out of [ɲ] as associated with the temporal lag between two independent lingual gestures, i.e., apical and back dorsal for the lateral and laminal and dorsal for the alveolopalatal. It has been shown, however, that alveolopalatal consonants are not complex segments [Recasens et al., 1995], and there is no obvious reason why the presence of a [w]-like tongue body configuration during the production of [ɫ] should result from the formation of two controlled apical and postdorsal constriction locations. Moreover, darkness appears to be an attribute of bilabial consonants for which a single articulator is clearly involved. Therefore, the perceptual salience of the transitional glide may very well result from differences in anticipation for a single articulatory event, i.e., tongue predorsum lowering and postdorsum backing for [ɫ], lip closing with some possible concomitant tongue depression for [β], and tongue dorsum raising and fronting for [ɲ] (see section 1.1 through 1.3 regarding evidence for temporal anticipation in articulatory activity in the case of all three consonants). The prominence of the glide in question is prone to increase with syllable-final lengthening and at slow speech rates.

5. Conclusion

The present study has shown that specific combinations of acoustic cues may cause [ɫ] and [β] to be confused with [w] and a glide to be heard at the onset of [ɲ]. It also yields some support for sound changes being achieved through more than one evolutionary path, e.g., the vocalization of [ɫ] may be associated exclusively with the acoustic characteristics of the steady-state formant frequency period ([VɫC] > [VwC]), or else with those spectral properties and with the timing of the vowel transitions ([VɫC] > [VwɫC] > [VwC]). Further research could apply the same or a similar experimental paradigm to the investigation of vocalization and glide insertion processes in other language families and for different variants of the same consonants used in the present study or for other consonants as well (e.g., the vocalization of clear /l/ into [j], the insertion of [j] before an alveolar or a velar consonant). It should also test whether better identification results may be achieved by improving the quality of the synthetic speech stimuli. Moreover, the use of articulatory synthesis appears to be needed in order to explore the articulatory mechanisms which give rise to the acoustic cues triggering vocalization and glide insertion.

Acknowledgments

This research was funded by projects HUM2006-03743 and FFI2009-09339 of the Ministry of Education/Innovation and Science of Spain and FEDER, and 2005SGR864 and 2009SGR3 of the Catalan Government. I would like to thank Francis Nolan, Fernando Sánchez Miret, Maria Josep Solé, Marzena Zygis and two anonymous reviewers for comments on previous manuscript versions.

References

Bladon, R.A.W.: The production of laterals: some acoustic properties and their physiological implications; in Hollien, Hollien, Current issues in the phonetic sciences, pp. 501–508 (Benjamins, Amsterdam 1979).

Boutier, G.; Counet, M.-Th.; Lechanteur, J.; Legros, E.; Remacle, L.: Atlas linguistique de la Wallonie, 15 vols (Université de Liège, Liège 1953–2006).

Browman, C.P.; Goldstein, L.: Gestural syllable position effects in American English, in Bell-Berti, Raphael, Producing Speech: Contemporary Issues. A Festschrift for Katherine Safford Harris, pp. 19–34 (American Institute of Physics, Woodbury 1995).

Contini, M.: Étude de géographie phonétique et de phonétique instrumentale du sarde (Edizioni dell'Orso, Torino 1987).

Dart, S.: Articulatory and acoustic properties of apical and laminal articulations. UCLA Working Papers Phonet. *79* (1991).
Diehl, R.L.; Walsh, M.A.: An auditory basis for the stimulus-length effect in the perception of stops and glides. J. acoust. Soc. Am. *85:* 2154–2164 (1989).
Espy Wilson, C.Y.: Acoustic measures for linguistic features distinguishing the semivowels /wjrl/ in American English. J. acoust. Soc. Am. *92:* 736–757 (1992).
Essen, O. von: An acoustic explanation of the sound shift [ɫ] > [u] and [l] > [i]; in Abercrombie, Fry, MacCarthy, Scott, Trim, In honour of Daniel Jones, pp. 53–58 (Longman, London 1964).
Fant, G.: Acoustic theory of speech production (Mouton, The Hague 1960).
Fleischer, F.: Studien zur Sprachgeographie der Gascogne (Karras, Halle 1912).
Fouché, P.: Études de phonétique générale (Les Belles Lettres, Paris 1927).
Fuchs, S.; Hoole, P.; Brunner, J.; Inoue, M.: The trough effect: an aerodynamic phenomenon?. Proc. From Sound to Sense, MIT 2004, C25–C30.
Gay, T.: Effect of speaking rate on vowel formant movements. J. acoust. Soc. Am. *63*:223–230 (1978).
Gay, T.; Ushijima, T.; Hirose, H.; Cooper, F.S.: Effect of speaking rate on labial consonant-vowel articulation. J. Phonet. *2:* 47–63 (1974).
Gick, B.: Articulatory correlates of ambisyllabicity in English glides and liquids; in Local, Ogden, Temple, Papers in Laboratory Phonology. VI: Phonetic Interpretation, pp. 222–236 (Cambridge University Press, Cambridge 2003).
Gick, B.; Min Kang, A.; Whalen, D.H.: MRI evidence for commonality in the post-oral articulations of English vowels and liquids. J. Phonet. *30:* 357–371 (2002).
Gick, B.; Wilson, I.: Pre-liquid excrescent schwa: what happens when vocalic targets conflict. Proc. 7th Eurospeech, Aalborg Center for Person Kommunikation 2001, pp. 273–276.
Giles, S.B.; Moll, K.L.: Cinefluorographic study of selected allophones of /l/. Phonetica *31:* 206–227 (1975).
Gimson, A.C.: An introduction to the pronunciation of English; 2nd ed. (Arnold, Bristol 1970).
Gossen, Ch.T.: Grammaire de l'ancien picard (Klincksieck, Paris 1970).
Grammont, M.: Traité de phonétique; 9th ed. (Delagrave, Paris 1971).
Haiman, J.; Benincà, P.: The Rhaeto-Romance languages (Routledge, London 1992).
Hardcastle, W.; Barry, W.: Articulatory and perceptual factors in /l/ vocalisation in English. J. Int. Phonet. Ass. *15:* 3–17 (1989).
Keating, P.; Lindblom, B.; Lubker, J.; Kreiman, J.: Variability in jaw height for segments in English and Swedish VCVs. J. Phonet. *22:* 407–422 (1994).
Koneczna, H.; Zawadowski, W.: Przekroje Rentgenograficzne glosek Polskich (Panstwowe Wydawnictwo Naukowe, Warsaw 1951).
Koneczna, H.; Zawadowski, W.: Obrazy Rentgenograficzne glosek Rosyjskich (Panstwowe Wydawnictwo Naukowe, Warsaw 1956).
Krakow, R.: The articulatory organization of syllables: a kinematic analysis of labial and velar gestures; PhD thesis Yale University (1989).
Krakow, R.A.: Nonsegmental influences on velum movement patterns: syllables, sentences, stress, and speaking rate; in Huffman, Krakow, Phonetics and phonology: nasals, nasalization, and the velum, vol. 5, pp. 3–59 (Academic Press, San Diego 1993).
Krakow, R.: Physiological organization of syllables: a review. J. Phonet. *27:* 23–54 (1999).
Lafont, R.: Eléments de phonétique occitane (Vent Terral, Valderiès 1983).
Lass, N.: On explaining language change (Cambridge University Press, Cambridge 1980).
Lausberg, H.: Lingüística románica (Gredos, Madrid 1970).
Lehiste, I.: Some acoustic characteristics of selected English consonants (Indiana University Research Center in Anthropology, Folklore and Linguistics, 1964).
Leite de Vasconcellos, J.: Esquisse d'une dialectologie portugaise; 3rd ed. (Instituto Nacional de Investigaçao Científica, Lisboa 1987).
Liberman, A.M.; Delattre, P.C.; Gerstman, L.J.; Cooper, F.S.: Tempo of frequency change as a cue for distinguishing classes of speech sounds. J. exp. Psychol. *52:* 127–137 (1956).
Lindblad, P.; Lundqvist, S.: [l] tends to be velarised, apical as opposed to laminal, and produced with a low jaw, and these features are connected. Proc. 15th ICPhS, Barcelona 2003, pp. 1899–1902.
Lindblom, B.; Sussman, H.; Modarresi, G.; Burlingame, E.: The trough effect: implication for speech motor programming. Phonetica *59:* 245–262 (2002).
Lipski, J.M.: Spanish stops, spirants and glides: from consonantal to [vocalic]; in Mazzola, Issues and theory in Romance linguistics, pp. 67–86 (Georgetown University Press, Washington 1994).
Mack, M.; Blumstein, S.E.: Further evidence of acoustic invariance in speech production: the stop-glide contrast. J. acoust. Soc. Am. *73:* 1739–1750 (1983).
Malécot, A.: Acoustic cues for nasal consonants: an experimental study involving a tape-splicing technique. Language *32:* 274–284 (1956).
Malmberg, B.: Phonétique générale et romane (Mouton, The Hague 1971).
Martins, P.; Carbone, I.; Pinto, A.; Silva, A.; Texeira, A.: European Portuguese MRI based speech production studies. Speech Commun. *50:* 925–952 (2008).
McAllister, R.; Engstrand, O.: Interpretations of tongue movement patterns in VCV sequences. Proc. Fonetik 92. Papers from the 6th Swed. Phonet. Conf., Gothenburg 1992, pp. 115–119.

Miller, J.L.; Baer, T.: Some effects of speaking rate on the production of /b/ and /w/. J. acoust. Soc. Am. *73:* 1751–1755 (1983).
Miller, J.L.; Liberman, A.M.: Some effects of later-occurring information on the perception of stop consonant and semivowel. Perception Psychophysics *25:* 457–465 (1979).
Mowrey, R.; Pagliuca, W.: The reductive character of articulatory evolution. Riv. Ling. *7:* 37–124 (1995).
Narayanan, S.S.; Alwan, A.A.; Huker, K.: Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. Part I. The laterals. J. acoust. Soc. Am. *101:* 1064–1077 (1997).
Navarro Tomás, T.: Manual de pronunciación española; 17th ed. (Consejo de Investigaciones Científicas, Madrid 1972).
Nittrouer, S.; Studdert-Kennedy, M.: The stop-glide distinction: acoustic analysis and perceptual effect of variation in syllable amplitude envelope for initial /b/ and /w/. J. acoust. Soc. Am. *80:* 1026–1029 (1986).
Ohala, J.J.: Phonetic explanation in phonology; in Bruck, Fox, La Galy, Papers from the Parasession on Natural Phonology, pp. 251–274 (Chicago Linguistic Society, Chicago 1974).
Ohala, J.J.: The story of [w]: an exercise in the phonetic explanation for sound patterns. Rep. Phonol. Lab., Berkeley *2:* 133–155 (1978).
Pope, M.K.: From Latin to Modern French with special consideration of Anglo-Norman (Manchester University Press, Manchester 1934).
Recasens, D.: Place cues for nasal consonants with special reference to Catalan. J. acoust. Soc. Am. *73:* 1346–1353 (1983).
Recasens, D.: Estudis de fonètica experimental del català oriental central (Publicacions de l'Abadia de Montserrat, Barcelona 1986).
Recasens, D.: An articulatory-perceptual account of velarization and elision of dark /l/ in the Romance languages. Lang. Speech *39:* 63–89 (1996).
Recasens, D.: Predicting directionality patterns in assimilatory and epenthetic processes from patterns of coarticulatory directionality. Proc. XIVth ICPhS, San Francisco 1999, vol. 3, pp. 1847–1850.
Recasens, D.: The effect of syllable position on consonant reduction (evidence from Catalan consonant clusters). J. Phonet. *32:* 435–453 (2004).
Recasens, D.: Articulatory and acoustic factors involved in the vocalization of dark /l/ and /l/ elision in Romance; in Sánchez Miret, Romanística sin complejos. Homenaje a Carmen Pensado, pp. 455–482 (Lang, Bern 2009).
Recasens, D.; Farnetani, E.: Spatiotemporal properties of different allophones of /l/: phonological implications; in Dressler, Prinzhorn Rennison, Phonologica 1992, pp. 195–204 (Rosenberg & Sellier, Torino 1994).
Recasens, D.; Fontdevila, J.; Pallarès, M.D.: A production and perceptual account of palatalization; in Connell, Arvaniti, Papers in Laboratory Phonology, IV, pp. 265–281 (Cambridge University Press, Cambridge 1995).
Recasens, D.; Martí, J.: Perception of unreleased nasal consonants in Catalan. J. d'Acoustique *3:* 287–299 (1990).
Rohlfs, G.: Grammatica storica della lingua italiana e dei suoi dialetti. Fonetica (Einaudi, Torino 1966).
Ronjat, J.: Grammaire (h)istorique des parlers provençaux modernes. Vol. I: Voyelles et diphthongues (1930); vol. II: Consonnes et phénomènes généraux (1932) (Société des Langues Romanes, Montpellier 1930–1941).
Sampson, R.: Nasal vowel evolution in Romance (Oxford University Press, Oxford 1999).
Shinn, P.; Blumstein, S.E.: The role of the amplitude envelope for the perception of [b] and [w]. J. acoust. Soc. Am. *75:* 1243–1252 (1984).
Steriade, D.: Gemination and the Proto-Romance syllable shift; in Birdsong, Montreuil, Advances in Romance linguistics, pp. 371–409 (Foris, Dordrecht 1988).
Straka, G.: Naissance et disparition des consonnes palatales dans l'évolution du latin au français. Travaux de Linguistique et de Littérature, Univ. Strasbourg *3:* 117–167 (1965).
Stuart-Smith, J.; Timmins, C.; Tweedie, F.: Conservation and innovation in a traditional dialect: L-vocalization in Glaswegian. English World-Wide *27:* 71–87 (2006).
Svirsky, M.; Stevens, K.; Matthies, M.; Manzella, J.; Perkell, J.; Wilhems-Tricarico, R.: Tongue surface displacement during bilabial stops. J. acoust. Soc. Am. *102:* 562–571 (1997).
Tuttle, E.: Considerazione pluristratica sociale degli esiti di AU e AL + alveodentale nell'Italia settentrionale. Actes XVIIIe Congr. Int. de Ling. et de Philologie Romanes, Tübingen 1991, vol. 3, pp. 571–582.
Vázquez, Y.; Hewlett, N.: The trough effect: an ultrasound study. Phonetica *64:* 105–121 (2007).
Wartburg, W. von: Französisches etymologisches Wörterbuch (Schroeder, Bonn/Mohr, Tübingen/Zbinden, Basel 1922–...).
Wells, J.C.: Accents of English (Cambridge University Press, Cambridge 1982).
Williams, E.B.: From Latin to Portuguese (University of Pennsylvania Press, Philadelphia 1938).