

Reflections about the phonemic analysis of Yongning Na (Tibeto-Burman): perceptual transcription and acoustic data

Alexis Michaud* & Jacqueline Vaissière**

*Langues et Civilisations à Tradition Orale (LACITO), CNRS/Univ. Paris 3

**Laboratoire de Phonétique et Phonologie (LPP), CNRS/Univ. Paris 3

alexis . michaud (at) vjf . cnrs . fr

jacqueline . vaissiere (at) univ-paris3 . fr

Abstract

A preliminary phonemic inventory of the dialect of the Na language spoken in Yongning, China (a.k.a. Eastern Naxi 纳西语东部方言/ Mosuo 摩梭话) is presented, on the basis of first-hand fieldwork data. Some reflections are offered on three interrelated issues commonly encountered in fieldwork, namely (i) To what extent does perceived allophonic variation correspond to articulatory/acoustic reality, to what extent does it merely reflect the investigator's perceptual expectations? (ii) How can acoustic data help select International Phonetic Alphabet symbols for the phonemic units brought out by distributional analysis? (iii) How can acoustic data help characterise the vowels and consonants encountered in fieldwork, both in a structural (language-internal) perspective, and in a cross-language perspective?

Key words: phonemics; acoustics; allophony; perception; Na/Naxi/Mosuo.

Résumé

Ces quelques pages, fondées sur deux enquêtes de terrain, présentent un premier inventaire phonémique du dialecte na parlé à Yongning, en Chine (ce parler est également connu sous le nom de « naxi oriental », 纳西语东部方言, ou « moso », 摩梭话). Trois problèmes couramment rencontrés lors des enquêtes de terrain sont abordés. 1) Dans quelle mesure la variation allophonique perçue par l'enquêteur, qui n'est pas locuteur natif, renseigne-t-elle une réalité acoustique, et dans quelle mesure reflète-t-elle simplement les attentes perceptuelles de l'enquêteur ? 2) De quelle façon des données acoustiques peuvent-elles guider le choix de symboles phonétiques pour les unités phonémiques dégagées par l'analyse distributionnelle ? 3) Comment des données acoustiques peuvent-elles aider à caractériser voyelles et consonnes, au plan structural (celui du système de la langue concernée) et au plan de la comparaison entre langues ?

Mots-clefs: phonèmes ; acoustique ; allophonie ; perception ; na/naxi/mosuo.

1. Introduction

1.1. Usefulness of acoustic description. The project of an acoustic description of Yongning Na.

The selection of International Phonetic Alphabet symbols for the phonemes of a given language is essentially based on structural arguments and expert listening, rather than on acoustic analysis and modelling. This is true both for the most documented languages, because their description is usually still based on a tradition that predates modern acoustic

phonetics, and for newly described languages, because acoustic analysis is seldom a priority in fieldwork.

IPA notations are oriented towards the transcription of the language's phonemic oppositions. They offer some indications on the articulatory characteristics of the sounds at issue, but cannot be expected to provide a full specification of the pronunciation of the sounds under study: they do not encapsulate enough information to reproduce the sound at issue.

The acoustic theory of speech production [1] allows for a representation of acoustico-perceptual characteristics of speech sounds by reference to the resonant properties of the vocal tract: the F-pattern. Unlike IPA notation, a characterisation in terms of F-pattern is sufficiently detailed to serve as input for speech synthesis, via articulatory modelling [2]. (To be quite precise, the F-pattern needs to be supplemented by information on nasality, if any, and on phonation type.) In the belief that speech sounds can be characterised very accurately in terms of a target F-pattern and of the articulatory configuration used to attain this pattern, the second author of this article is currently developing a notation [3] to describe a given phoneme by reference to certain fixed acoustical properties. The aim of this research is to some extent comparable with that of Daniel Jones in defining the cardinal vowels: to provide a stable reference point for linguistic description. However, to our knowledge, the cardinal vowels have not been explicitly related to formant characteristics: they are defined in articulatory terms, and the audio renderings proposed to illustrate them fluctuate somewhat, e.g. some of the cardinal vowels recorded by Peter Ladefoged are somewhat different from those of Daniel Jones (see [3] for more details). It appears worth investigating in which ways an acoustic definition of the sounds of a given language can supplement the IPA notations for this language and facilitate cross-language comparisons.

The present article presents preliminary observations on how acoustic observations and reflections can help overcome some difficulties encountered in fieldwork. The object of study is the Na language (Tibeto-Burman family) as it is spoken in Yongning, China (Chinese coordinates: 云南省丽江市宁蒗彝族自治县永宁乡平静村). This language is also known as Mosuo (摩梭话) or Eastern Naxi (纳西语东部方言).

Subsection 1.2 of the Introduction offers explanations on the provisional notation of tones used in this article. Section 2 provides a (likewise provisional) phonemic inventory, with specific explanations about the palatalisation of velars (2.1), the three-way contrast between /k/, /k^h/ and /k^ʷ/ (2.2), consonantal nuclei (2.3) and rhotic rhymes (2.4). Acoustic analyses proper are set out in section 3. In conclusion, language-specific remarks (subsection 4.1) and remarks on acoustic analysis in fieldwork (4.2) open into the broader

project of a monographic study of the acoustics of dialects of the Na language (4.3).

1.2. Preliminary remarks on the prosodic system of Yongning Na

Like all the Na dialects documented to date – such as Lijiang Na, a.k.a. ‘Western dialect of Naxi’, and Namuyi [4]–, Yongning Na makes use of three tonal levels: High, Mid and Low, here transcribed as 1, 2, 3. However, unlike Lijiang Na, which essentially has a syllable-tone system (one tone per syllable), with only a few processes of tonal change [5, 6], Yongning Na appears to have word-tones: the lexical tonal schemes are partly neutralised in citation form. For instance, [gi+zu+1] ‘little brother’ and [gu+mi+1] ‘little sister’ have the same tonal pattern in citation form, yet they do not belong in the same tonal class, as demonstrated by their behaviour when followed by the copula [ŋi], e.g. in the carrier sentence [tʂ^hu. _ .ŋi], ‘This is...’ ([tʂ^hu] is a deictic): [tʂ^huŋgi+zu+1 ŋi] ‘this is little brother’ (tonal scheme: HMMH) vs. [tʂ^huŋgu+mi+1 ŋi] ‘this is little sister’ (tonal scheme: HMML).

The tonal system of Yongning Na is currently under analysis. It is reminiscent of the ‘tonal accent’/‘pitch accent’ of Japanese and of the ‘tonal schemes’ of Sub-Saharan languages such as Bambara [7], and has some analogues in other Tibeto-Burman languages (see [8] and references therein). The notation of tones in this article is provisional.

The tonal classes are exemplified in table 1. The carrier sentence used for nouns is ‘This is...’, [tʂ^hu. _ .ŋi], as has just been exemplified for the words ‘little brother’ and ‘little sister’. Monosyllabic nouns and disyllabic nouns are provisionally placed in different categories—four for monosyllables, seven for disyllables—; progress in the analysis of the tone system should allow for a unified account of the tonal classes of these two sets. Tonal classes for predicative words (verbs and adjectives) are established on the basis of their behaviour when followed by the particle /zeJ/, which conveys accomplished aspect, in the carrier phrase /le. _ ze/, e.g. [le+dzu+zeJ] ‘has eaten’. A first overview suggests that there are no disyllabic predicative words apart from reduplicated forms, and that their tonal scheme can be derived straightforwardly from that of the corresponding monosyllables, so they are not included in table 1.

Table 1. Provisional classification of Yongning Na lexical items into tonal classes, on the basis of their behaviour in carrier sentences (see main text). Categories m1 to m4: monosyllabic nouns; d1 to d7: disyllabic nouns. For predicates: four categories.

number assigned	tones in frame	examples		
		tones in citation+in frame	meaning	
m1	1 1	dæ1 → dæJ ŋi1	dust	
m2	1 2	kwɣ1 → kwɣ1 ŋi2	mountain	
m3	1 1	ŋv1 → ŋvJ ŋi1	silver	
m4	1 1	ʃi1 → ʃiJ ŋi1	bone	
d1	1 1 1	gu+mi+1 → gu+mi+1 ŋi1	little sister	
d2	1 1 1	gi+zu+1 → gi+zu+1 ŋi1	little brother	
d3	1 1 1	nyJmi1 → nyJmiJ ŋi1	heart	
d4	1 1 1	ku1dzu1 → ku1dzuJ ŋi1	Tibetan	
d5	1 1 1	ziJbæ1 → ziJbæ+1 ŋi1	lightning	
d6	1 1 1	aJku1 → aJkuJ ŋi1	home	
d7	1 1 1	zu+my1 → zu+myJ ŋi1	child	
predicates	1	1 1 1	li1 → le+1 li+1 ze+1	to watch
	2	1 1 1	se1 → le+1 se1 zeJ	to walk
	3	1 1 1	tʂ ^h u1 → le1 tʂ ^h uJ zeJ	to drink
	4	1 1 1	la1 → le+1 la+1 ze1	to strike

In what follows, the tonal category is indicated in superscript characters at the beginning of the syllable, e.g. /^{m2}dzi/ ‘water’ belongs in tonal category m2.

2. Phonemic inventory of Yongning Na

The present phonemic inventory of Yongning Na is based on fieldwork conducted in 2006 and 2007. As the word list comprises only some 2,000 entries, it is possible that more phonemic oppositions will be discovered in future.

The inventory presented here differs somewhat from that proposed by [9], due in part to dialectal differences ([9] is based on fieldwork in the village of Luòshuǐ 落水, whereas the present data come from 平静村, in the Yongning plain), in part to differences in phonemic analysis. A detailed comparison of the two analyses falls outside the scope of this paper.

Syllabic structure in Yongning Na is simply (C)(G)V, where C is a consonant, G a glide – with a severely restricted distribution – and V a vocalic element; the brackets indicate that C and G are optional. It therefore appeared simplest to present the inventory in table 2 in terms of initials, corresponding to consonants, and rhymes, corresponding to (G)V.

Table 2. A preliminary inventory of the phonemes of Yongning Na. Initials (left) and rhymes (right).

	bilabial	coronal	retroflex	velar	uvular	glottal
plosive	p ^h p b	t ^h t d	tʂ ^h t ɖ	k ^h k g		
affricate		tʂ ^h ts dz tɕ ^h tɕ dz	tʂ ^h tʂ dʂ			
nasal	m	n		ŋ		
fricative		s z ɕ ʒ	ʂ ʐ		ʁ	h
lateral approximant		l				
lateral fricative		ɬ				
approximant			ɹ			

i	u	u, ju
	y	
		ɤ, wɤ, jɤ
	ɹ	
	æ, wæ, jæ	
		a, wa, ja
+ /i/, /y/, /ü/, /wɤ/, /æ/, /ä/ after /h/, and /ɹ/ on its own as a syllable		

The following subsections provide some explanations about table 2.

2.1. The palatalisation of velars and its consequences

As compared to the Na dialects spoken in and around the city of Lijiang (a.k.a. Western dialects of Naxi [10, 11]), Yongning Na has undergone a thorough palatalisation of velars before /i/ and /j/. The empty slots in the phonemic system created by this evolution have been filled by syllables whose vowel underwent fronting, as shown in table 3.

Table 3. Some Yongning Na (“Eastern” Na) cognates of velar-initial syllables of Lijiang Na (“Western” Na).

	“Eastern” Na (Yongning)	“Western” Na (data from [3, 4])
water	/ ^{m2} dzi/	/gi-/
to fall (rain)	/(toneless)gi/	/gu-/
little brother	/ ^{d2} gi.zu/	/gu-+zu-/
to have	/ ^{m1} dzu/	/gy/

Clearly, the conservative form of the word for ‘water’ is /gi/, whereas /dzi/ is the result of palatalisation. It appears reasonable to hypothesise that the word for ‘to fall (rain)’ and the first syllable of the word for ‘little brother’ used to have a back vowel which later became fronted and came to occupy the slot left empty by the palatalisation of /*gi/, hence present-day /gi/, which now stands in a relation of phonemic contrast with /dzi/. The existence of this new contrast between /gi/ and /dzi/ probably contributes, in turn, to the further fronting of the palatalised syllables.

The case of ‘to have’ is slightly more complex. In addition to the palatalisation of the initial, from /g/ to [j] and further to [dz], the rhyme appears to have diphthongised to [ju]; the resulting syllable, in the present-day state of the language, can be analysed phonemically either as /dz/ + /ju/ or as /dz/ + /u/. The former analysis may carry an etymologising bias, reflecting diachronic hypotheses rather than synchronic, phonetic reality. We choose the analysis as /dz/ + /u/, which reflects the intuition that the frontness of the earlier rhyme has been entirely transferred onto the initial.

Palatalisation is also having consequences on other phonemes of Yongning Na. The extreme fronting of the syllables /t^hi/, /t^hei/, /dzi/ brings them in the phonetic vicinity of the syllables /ts^hu/, /tsu/, /dzu/. The rhymes of both sets of syllables can be described as the voiced continuation of the initial: for instance, /tsu/ is realised as [tsz], and /t^hei/ can be approximately transcribed in IPA as [t^hɛ̣], rather than [t^hei]. (The diacritic indicating syllabic status is placed over the [ɛ̣], not underneath it, for the sake of clarity.)

It can be imagined that the continuous evolution towards fronting combined with an extreme degree of syllable-level coarticulation may threaten the distinction between these two sets (/t^hi/, /t^hei/, /dzi/ on the one hand, /ts^hu/, /tsu/, /dzu/ on the other). The high functional load of the difference between the two sets goes against their confusion, however. So does the ever growing influence of Mandarin Chinese, where both sets exist: alveolo-palatal as in 七 /t^hi/ ‘seven’ and 鸡 /t^hei/ ‘chicken’, with less palatalisation than in Yongning Na; dental as in 此 /ts^hu/ *DEICT.* and 子 /tsu/ ‘child’, less fronted than in Yongning Na. To venture a remark on possible future developments: If the growing proficiency of Na

speakers in Mandarin led them to adopt a Mandarin-like pronunciation for their palato-alveolars (i.e. less fronted than they are now in Na), this would restore a greater phonetic distance between them and the dental affricates.

As for the alveolo-palatal fricative /z/, it is exclusively found in combination with /i/. Examples include /^{m1}zi/ ‘monkey’, /^{m4}zi/ ‘family’, /⁴zi/ ‘to sleep’, and /³zi/ ‘to flow (river, water)’. It appears to be an example of an empty slot in the phonemic system being filled in the course of language evolution: Yongning Na has a voicing distinction with a high functional load, so that the introduction of a voiced counterpart to /ç/ fits in well within the system. How this evolution actually took place is an issue for future investigations.

2.2. The affricated velar stop /k^ɕ/ as a phoneme

The affricated velar stop [k^ɕ] is attested before /æ/, /a/, /wa/ and /y/. Before the first three, it can be treated as an allophone of /k^h/; indeed, it is often accompanied by some aspiration (hence the notation as [k^{hɕ}] in table 4 below). However, it contrasts with /k^h/ (as well as with /k/) before /y/. The combination /k^ɕy/ is found in the words for ‘hole’, ‘six’, ‘to fold’, ‘throat’; /k^hy/ in ‘dog’, ‘year’, ‘to cut (grass)’, ‘to steal’; /^{m4}ky/ means ‘garlic’. This quasi-complementary distribution is shown in table 4. As a conclusion, phonemic status has to be granted to the affricated velar stop /k^ɕ/.

Table 4. Distribution of [k^h], [k^ɕ]/[k^{hɕ}], and [k] in Yongning Na; “-” indicates unattested combinations.

	æ	a	wa	i	o	u	y
k	kæ	ka	kwa	ki	ko	ku	ky
k ^h	k ^{hɕ} æ	k ^{hɕ} a	k ^{hɕ} wa	k ^h i	k ^h o	k ^h u	k ^h y
k ^ɕ	-	-	-	-	-	-	k ^ɕ y

2.3. General remarks on consonantal nuclei

In Yongning Na, some consonantal sounds function as syllable nuclei – which amount to rhymes, in the absence of any final consonants. The IPA term for this is to call them ‘syllabic’, because they function as syllable nucleus (though they do not constitute a syllable by themselves). Fricative syllabics are an areal characteristic: in particular, they are also common in neighbouring Yi languages, belonging to the Nasoid subgroup of Loloish (a.k.a. Ngwi; see [12, page 70]). The local dialect of Mandarin also has a fricative allophone of /u/ in words such as ‘lake’, 湖, pronounced [fv] (Standard Mandarin: /hu/).

There are three sets of consonantal rhymes in Yongning Na:

(i) The voiced fricative /v/ can only appear as a rhyme, not as an initial. The diacritic / / is thus implied in what follows. After bilabial initials, it is slightly trilled (i.e. /bv/ is realised close to /β/, /pv/ close to /pβ/), though less markedly so than in Lijiang Na. On the whole, the friction for /v/ is much less salient than in Lijiang Na.

(ii) The vowel /u/ has fricative allophones after dental and retroflex fricatives and affricates, like the vowel /i/ in Mandarin Chinese. For instance, /ts^hu/ ‘hot’ is realised [t^hɛ̣], /^{m4}dzu/ ‘market, city’ is realised [m⁴dɛ̣].

(iii) A rhotic element is also present in some rhymes. From an acoustic point of view, studies of various languages

have shown that typical rhotic vowels always have a lowering of the third formant (as mentioned e.g. by [13, page 313]). Unlike Lijiang Na, whose rhotic vowel /ɜ/ clearly presents this characteristic, Yongning Na does not have a canonical rhotic rhyme; lexical correspondences confirm that rhoticity in Yongning Na and Lijiang Na are distinct phenomena, since out of 34 words containing the rhyme /ɜ/ in Lijiang Na and which have likely cognates in Yongning Na, 16 have the vowel /æ/ in Yongning, 10 have /v/. The situation in Yongning Na therefore warrants further analyses. Section 2.4 is a first step in this direction.

2.4. Rhotic rhymes

2.4.1. The phoneme /ɹ/

Yongning Na has several rhotic-like sounds. One of these is the retroflex approximant, /ɹ/, which can serve as an initial, as a rhyme, or constitute a syllable in itself.

It appears as an initial in the words /^{m1}ɹæ/ ‘seed’ and /^lɹwæ/ ‘to cry, to shout [of animals or people]’; also in the second syllable of /^{d1}su.ɹæ/ ‘table’, /^{d4}t^hæ.ɹæ/ ‘book, scripture’, /^{d2}k^hwæ.ɹæ/ ‘mat’. The phoneme /ɹ/ is not phonologically specified for voicing – indeed, there does not exist an IPA symbol for an unvoiced counterpart to /ɹ/, which means that such a contrast has not been reported to date in any language. In initial position, the /ɹ/ of Yongning Na appears to be partly devoiced.

The phoneme /ɹ/ is observed as a rhyme in three contexts: after /t/; after /t^h/; and on its own, the morpheme /ɹ/ being a verbal suffix, e.g. /^lhu/ ‘wait’, [dɹ+hu+ɹ] ‘Wait!’ (/dɹ/ is the numeral ‘one’).

The syllable /tɹ/ contrasts with several syllables that are articulatorily close to it, having either back vowels or consonantal rhymes: /tv/, /tɹ/ (e.g. in /^{d7}k^hwæ.tɹ/ ‘turban’), /tɹ/ (e.g. /^ltɹ/ ‘to pull’), and /tɹɹ/ (e.g. /^ltɹɹ/ ‘to call out [of cock], to cock-a-doodle-doo’; tone not verified).

In the process of comparing spectrograms of these various phoneme combinations, we realised that an understanding of their acoustics required a comparison of all five members of the /t/-initial set with one another, with /t/-initial syllables, and with affricated-initial syllables. After obtaining these data, we still required additional information on the vowels, since their realisation after /t/ can hardly be studied without a knowledge of their intrinsic acoustic characteristics, which in turn requires observations on these vowels in a variety of consonantal contexts. The conclusion, which will be taken up in section 4.3 below, is that such analyses require a systematic approach – a monograph on the acoustics of the language.

2.4.2. The phoneme /ɹ̃/

The phoneme /ɹ̃/ is the nasal counterpart of the phoneme /ɹ/ studied in the previous section. This phoneme, /ɹ̃/, is used to transcribe a syllable that occurs only once in the vocabulary list: /ɹ̃/ ‘bone’ (also found in many compounds, such as /su.ɹ̃.mi/ ‘vertebral spine’ and /ɹ̃.ku/ ‘knee’). It is a nasal, retroflex, approximant sound. Other notations would be possible, for instance as a nasal rhotic vowel: /ɜ̃/. The symbol /ɹ̃/ was chosen because the units transcribed as /ɹ/ and /ɹ̃/ have similarities, and it appears plausible that they actually stand in a relation of structural opposition, the one nasal, the other oral.

The other nasal rhymes of Yongning Na are found only after /h/, where they contrast with oral rhymes, e.g. /hæ̃/ ‘wind’ vs. /hæ/ ‘lime’, /hã/ ‘evening’ vs. /ha/ ‘food’, /hĩ/ ‘human being; relativizing particle’ vs. /hi/ ‘rain’. Before oral rhymes, /h/ is realised with a friction source at a point in the vocal tract determined by the following vowel, e.g. palatal before /i/, velar before /a/ and /æ/, and labiodental before /y/ (hence [çi], [xɑ], [xæ], and [fỹ]), whereas the friction is very weak, due to an already lowered velum, in [hĩ], [hã], [hæ̃], [hỹ]. The analyses proposed in [5] concerning the diachronic origin of nasal/oral vowel oppositions in the Na dialect of Fèng Kē apply to the Yongning Na nasal vowels.

3. Acoustic analysis of the close, front vowel of Yongning Na (“/i/”)

3.1. The issue: coarticulation, or categorical allophony?

One of the issues encountered in fieldwork was the notation of high vowels. The first, tentative notations of Yongning Na included [e] as well as [i], [o] as well as [u]. Looking back at these transcriptions, it appears that close vowels [i], [u] are in complementary distribution with close-mid vowels [e] and [o]. Table 5 shows their distribution¹.

Table 5. Distribution of close and mid-close vowels in the first field notations of Yongning Na vocabulary.

rhyme	[e], [o]	[i], [u]
initial	retroflex (fricatives,	zero initial; labials;
consonant	affricates and stops);	dental stops; laterals;
	dental fricatives and	alveolo-palatals;
	affricates	velars; glottal; nasals

The standard way of reporting on such fieldwork observations is to describe the phoneme at issue as having two allophones, and to make a reasoned choice of one of these allophones to represent the phoneme: for instance, one could choose to transcribe /i/ for [i, e], and /u/ for [u, o]. This amounts to saying that the difference between the phonetic realisation of these phonemes in the two sets of consonantal contexts listed in table 5 goes beyond what can be predicted from coarticulatory properties of the preceding consonant: that the two contextual variants are phonetically different sounds.

This raises an important issue: does our classification of the realisations of one phoneme into two phonetic categories (e.g. transcribing [i] in [m¹i] ‘spot, pimple’ vs. [e] in [m⁴tse] ‘earth’) actually reflect an acoustic difference between these two realisations in Yongning Na, or does it merely reflect our perception as non-native speakers? Our perceptual expectations may not be adequate for the language under description, detracting from the precision and usefulness of the descriptions we produce.

Acoustic analysis can arguably provide decisive evidence on this issue, supplementing auditory impressions. Untrained listeners go by the categories of their native languages (for

¹ Note that [9] analyses the ‘fricative vowels’ mentioned in section 2.3 above as allophones of /i/ (i.e. the way things stand in Mandarin Chinese): [dʒ], e.g. in the verb ‘to eat’, is analysed as /dʒi/, whereas we analyse it as /dzui/. Thus, [9] does have a contrast between /i/ and /e/, because she transcribes with an /e/ the syllable which we transcribe as /dʒi/, e.g. /^{d1}dʒi.lu/ ‘wheat’.

experimental evidence of an effect of linguistic experience on the perception of front vowels, see, e.g., [14]). It is not very likely that trained phoneticians can fully cast off this bias. Phonetic transcriptions done in the field are commonly considered to reflect a (narrow) *phonetic* level, on top of which *phonemic/phonological* analyses can be built; however, a different perspective can in fact be adopted: field notations can be considered as perceptual materials for linguistic analysis and reflection. An important step consists in reflecting on the reasons why the investigators, given their linguistic and scholarly background, chose one notation rather than another.

In the case of the present investigation, conducted by native speakers of French, the issue can be phrased as: What are the acoustic properties that led to the perception of the sounds [i] and [e]? To preview the result, it will appear

- that the phoneme at issue differs acoustically both from French /i/ and from French /e/;
- that our perception of [i] in some contexts and [e] in others can be explained by the coarticulatory effects of the preceding consonant.

3.2. Acoustic observations on a female speaker

The native language of the authors, ‘Standard’ (‘Parisian’) French, has phonemic /i/ and /e/, as part of a vocalic system which – in its more conservative varieties – has the full set of cardinal vowels, /i e ε a, u o ɔ α/. In the notation currently being developed by the second author of this article, prototypical French /i/ corresponds to the acoustic configuration {palatal (↑F3F4)^{3200Hz}}, i.e. F3 is maximal, it corresponds to a resonance of the front cavity, and since F3 comes in the vicinity of F4, the amplitude of one of the two formants, or of both, is reinforced. For a male voice this spectral prominence is located at about 3,200 Hz; however, the exact value in itself is less important than the concentration of energy – a phenomenon which the notion of F’2 (read “F2-prime”) captures well [15, 16]. This vowel (French /i/) is “the most acute voiced, noise-free sound that a vocal tract can generate” [3]. This characterisation has recently been verified by the statistical examination of formant values extracted from large databases of broadcast speech in eight languages: the average distance between F3 and F4 for /i/ is much smaller in French than in the seven other languages investigated, namely German, English, Arabic, Spanish, Italian, Mandarin Chinese and Portuguese [17]. On the other hand, the acoustic configuration for French /e/ is one in which F1 is higher than the F1 of /i/, and where, for an adult male speaker, F2, F3 and F4 are roughly equidistant, i.e. not clustering together, unlike in /i/ (clustering of F3 and F4: ↑F3F4) or /y/ (clustering of F2 and F3: ↓F2F3).

An examination of spectrograms will offer some insights into the Yongning Na sound which had at first been transcribed as /i/. Spectrograms 1a, 2a and 3a (placed at the end of the article) show data from a female speaker aged about 55 at the time of the recordings, 1b, 2b and 3b from a male speaker aged about 35. The items are pronounced in isolation. Spectrogram 1 shows a realisation of Yongning Na ^mi/ ‘spot, pimple’.

The syllable begins with a semi-consonant: in Yongning Na (as in Lijiang Na), syllabic vowels are phonetically preceded by a semi-consonant acting as empty-onset filler. In

Lijiang Na, the notation that has been proposed by [10] for syllabic /i/ is [ji]. Given the amount of friction noise observed on spectrograms 1a-b, transcription of the onset as the fricative [j] appears more adequate for Yongning Na.

The vocalic part of the syllable in spectrogram 1a is clearly diphthongised. The measurements in table 6 are taken at 1/4th of the vowel, a time point chosen after [18, p. 94], and at 5/6^{ths} of the vowel, where F1 is at its highest. Formant frequencies were estimated by using the PRAAT software package (www.praat.org).

Table 6. Formant frequency measurements on spectrogram 1a, showing evidence of vowel diphthongisation. The time point is indicated as a function of the total length of the vowel.

time point	F1	F2	F3
1/4 th	420	2740	3190
5/6 ^{ths}	500	2510	3310

The first formant rises in the course of the vowel, while the distance between F2 and F3 increases. During the first half of the vowel, much intensity is concentrated around 3000 Hz; this cluster appears to be made up of F2 and F3. A superficial examination would conclude that these characteristics are very much unlike those of French [i]: to a French ear, the perception of [i] requires, in terms of a male voice, an F3 frequency that is closer to that of F4 than to that of F2, or, said differently, a strong difference between F3 and F2 frequencies: F3-F2 ≈ 1000 Hz. On the other hand, from a perceptual point of view, it appears to be the frequency of the resulting spectral prominence that plays a key role in perception, whether such an acoustic prominence results from the clustering of F3 and F4 in a male voice or of F2 and F3 in a female voice. “It appears that in female and children’s voices the relative role of individual formants may differ while an overall pattern aspect, yet to be defined, could be similar” [18, p. 98]; see also the discussion of speaker normalisation in speech perception, and of the notion of ‘auditory *Gestalt* recognition’, by [19].

This is in keeping with our auditory impressions. When listening to the first half and the second half of the vowel separately, our perceptual impression is [i] for the first half, [e] for the second half. It is known that a higher F1 can lead French subjects to perceive /e/ as opposed to /i/.

Spectrogram 2a shows a realisation of Yongning Na ^m2tɕi/ ‘cloud’ said in isolation by the same female speaker. Again, F1 rises (from 220 Hz to about 625 Hz) in the course of the rhyme, and F2 decreases (from 2420 to about 2170). F2 values are slightly lower than on spectrogram 1a; the amplitude of F2 movement is similar.

Spectrogram 3a shows the word for ‘earth’, initially transcribed as ^m4tɕe/. Interestingly, its vowel actually resembles those represented on spectrograms 1a and 2a. F2 is even higher than on spectrogram 2a. These facts are taken up in the following section.

3.3. Comments and perceptual hypotheses

The allophonic variation between [i] and [e] transcribed on the basis of auditory impressions does not match with the observations made on the female voice: the vowel itself is realised by a fairly constant formant pattern. It may be described as an opening diphthong, somewhat like [i] at its

beginning and like [e] at its end. The next step in analysis consists in finding out the reason for our original perception of two distinct allophones.

Our perception of [e] in the word for 'earth', as in the other words with initial retroflex fricatives and affricates, may in fact be due to characteristics of the consonant /tʂ/, rather than of the vowel itself. The friction in /tʂ/ has an overall low frequency, with a peak at about 4,100 Hz as against 6,700 for the friction in /tʃ/ (spectrogram 2a). It is plausible that the perception of the initial portion of the vowel is influenced by the spectral characteristics of the initial consonant (for a review of evidence for perceptual context dependence in various languages, see [20, p. 61]). In spectrograms 2a and 3a, the fricative noise continues until after voicing begins, i.e. into the beginning of the vowel; by itself, this is suggestive of a close vowel – corresponding to a configuration where the area of greatest constriction has a small area, hence the production of turbulence noise by egressive airflow. However, since the retroflex fricative (spectrogram 3a) has an overall lower spectrum than other fricatives – alveolo-palatal, as in spectrogram 2a, or dental –, the superposition of its friction noise onto the beginning of the following vowel may lower slightly the perceived main spectral prominence of the following vowel. Since the vowel at issue is acoustically in-between our reference points for [i] and [e] in the first place, this small perceptual influence from the preceding consonant could account for the perception of [e] rather than [i]. This hypothesis is tested by means of a (sketchy) experiment reported in the next section.

3.4. A pilot cross-splicing test

Cross-splicing, made easy by digital sound processing, can shed light on issues such as the one set out in the preceding section. The initial /tʃ/ of spectrogram 2a was placed before the vowel in spectrogram 3a. Our perceptual categorisation of the resulting stimulus confirms the expectations: after /tʃ/, we perceive this vowel as [i]. (This stimulus is available, together with the original sounds, at the internet address indicated in section 6.) Conversely, when the initial /tʂ/ of spectrogram 3a is placed before the vowel in spectrogram 2a, this vowel is not perceived as [i] anymore, but as a more open front vowel, [e] or [ɛ].

This result can be extended to the [u]-vs.-[o] allophonic pairs of the first fieldwork notations.

3.5. Observations on a male speaker

The data from the male speaker are shown on spectrograms 1b, 2b and 3b. The vowels in spectrograms 1b and 2b have a similar F-pattern, whereas 3b shows a rising F2. These data suggest that in the case of the male speaker, who is one generation younger than the female speaker, there is actually a certain amount of allophonic variation of the close front vowel phoneme under investigation: the acoustic realisation of this phoneme after a retroflex initial differs from its realisation in isolation and after a palato-alveolar fricative. Needless to say, this observation will need to be confirmed by the examination of more data, and of more speakers of the same age group.

4. Conclusions and perspectives

4.1. On the hypothesised allophonic variation of close vowels

The acoustic observations on the sounds initially transcribed as distinct allophones, [e] and [i], differ for the female speaker and for the male speaker.

In the data from the female speaker, a member of the older generation who speaks Yongning Na on a day-to-day basis, the contextual variants of the phoneme at issue are much closer than the first fieldwork notations suggested. This phoneme is an opening diphthong; to a foreign ear expecting a stable vowel quality, it can be perceived as [i] due to the high frequency of the spectral prominence found at vowel onset, or as [e] when the spectral properties of the preceding consonant have the effect of lowering the perceived spectral prominence.

On the other hand, in the data from the male speaker, who is one generation younger and has much greater proficiency in Chinese (local dialect, 西南官话, and Standard Mandarin, 普通话), preliminary observations actually suggest the presence of some allophonic variation of this vowel.

4.2. Implications for fieldwork: usefulness of acoustic observations

It is well known that field workers need to be careful to describe the phonetic categories of the language they investigate in their own right, rather than basing their judgments on analogies with sounds found in other languages that they know, such as their mother tongue or other varieties of the language under analysis. In the case of Yongning Na, the first author's earlier experience of fieldwork in 2002 and 2004 on a closely related language variety, Lijiang Na, certainly exercised an unwanted influence on phonetic notations: in Lijiang Na, there is a contrast between /i/ and /e/ (the latter actually closer to [ɛ]), e.g. /mbe/ 'village', /mbi/ 'urine') and between /o/ and /u/ (e.g. /k^ho/ 'noise, sound', /k^hu/ 'door'), which may have unwittingly led to an expectation that Yongning Na, too, had close vowels and mid-close vowels.

The results in section 3 illustrate the fact that a simple acoustic analysis can offer some evidence to overcome the uncertainties of auditory impressions. On the other hand, these results also suggest that the acoustic study needs to be conducted in a systematic way, collecting data from several speakers and building a corpus containing each phoneme in a variety of contexts, in order to obtain reliable results. This conclusion opens into a project which is outlined in the next section.

4.3. Issues of transcription: necessity of a systematic acoustic overview. Project of monograph on the acoustics of Yongning Na.

Concerning the notation of the high front diphthong studied in section 3, a decision can hardly be made in isolation, without considering the broader picture of the acoustic characteristics of the phonemes of Yongning Na. The choice between a notation as /ɪ/, /i/, or again a notation that would represent both the beginning and endpoint of the diphthong (such as /i^e/ or /i^ɛ/) will depend in part on how this sound compares with other vowel units in the language, and in part on how the formant structure observed in Yongning Na compares with the reference values available for an increasing number of

languages: American English [21], Swedish [18, p. 94], French, German [22]...

A consequence of the structural/functional notion of phonemic *system* is that the full picture of the language's contrasts should be taken into account before an adequate acoustic characterisation of individual sounds can be proposed. We therefore plan to produce a monograph on the acoustics of Yongning Na, on the basis of recordings by two women and two men.

The ultimate aim would be to propose a characterisation of each vocalic and consonantal phoneme by means of the fine-grained acoustic notation put forward by [3]. To this end, the more phonetic instruments can be taken into the field, the better: acoustic modelling requires acoustic-articulatory modelling, and the articulatory component requires more information than spectrograms can offer. Comparable acoustic results can be obtained by different articulatory

configurations, so that spectrographic analysis often leads to hypotheses that can only be verified by means of physiological data. We therefore plan to supplement audio data by recording multisensor data, such as oral and nasal airflow for the study of nasality, and electroglottography for the study of the voice source.

5. Acknowledgments

The first author expresses his heartfelt thanks to Latami Dashi (拉他咪·达石) for his help and advice during fieldwork in Yongning, and to the language consultants in Yongning for their hospitality and their patience. Our thanks to Liberty Lidz for kindly providing additional explanations about the phenomena that she observed in Luòshuǐ and elsewhere.

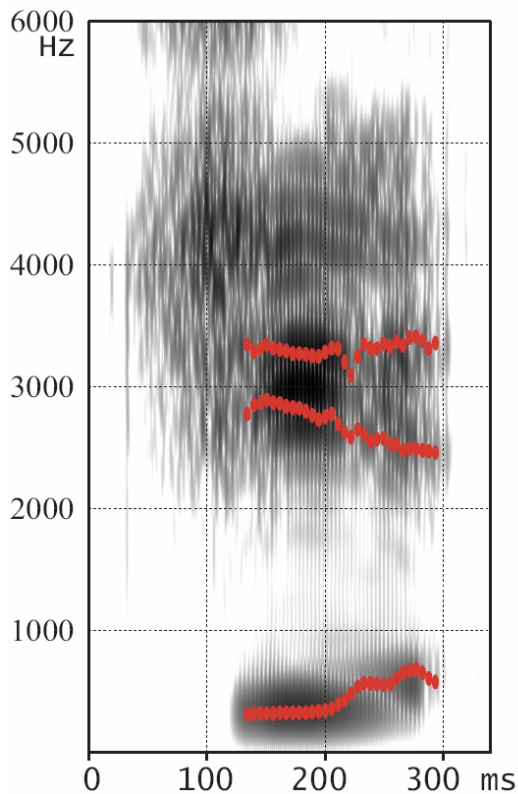
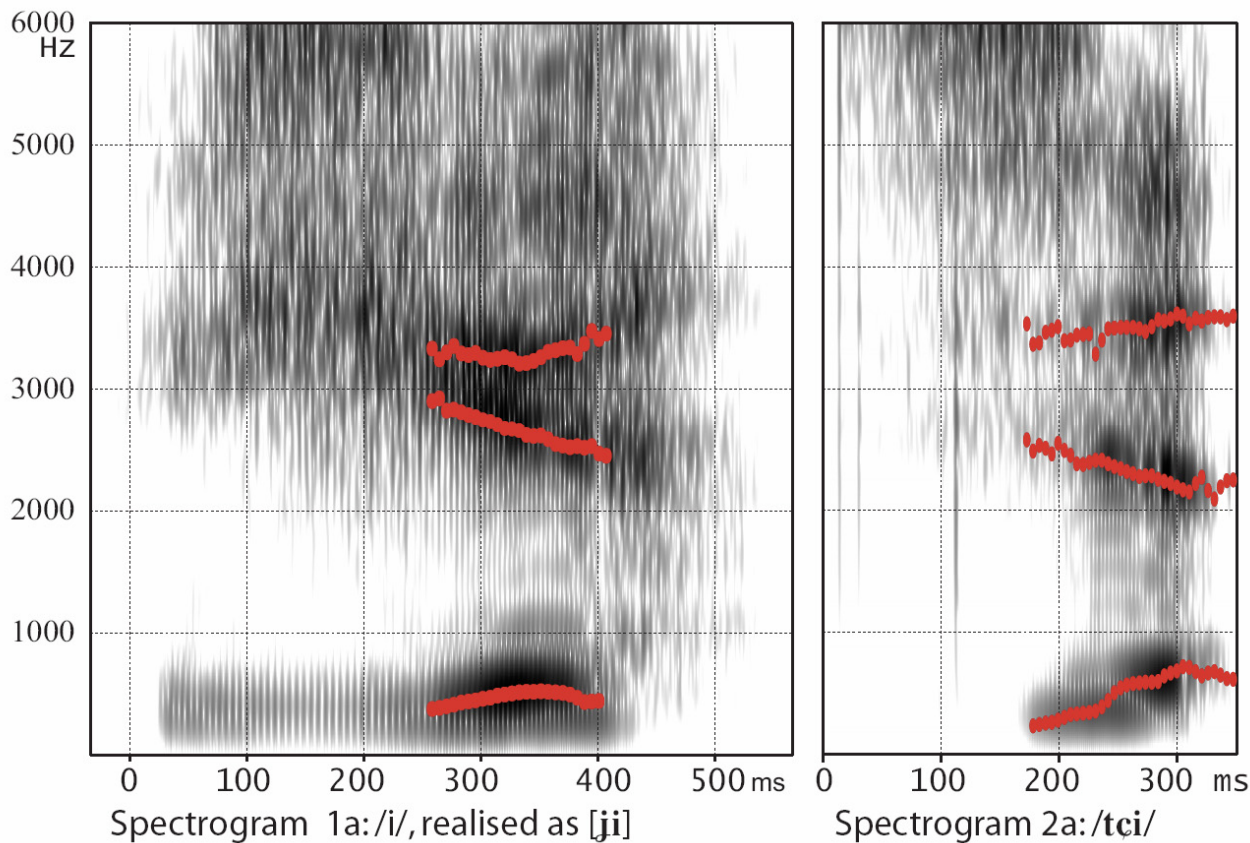
6. Link to sound files

The sounds corresponding to the spectrograms are available online from:

http://ed268.univ-paris3.fr/lpp/pages/EQUIPE/michaud/NAXI/Yongning_Na_HighFrontVowel.htm

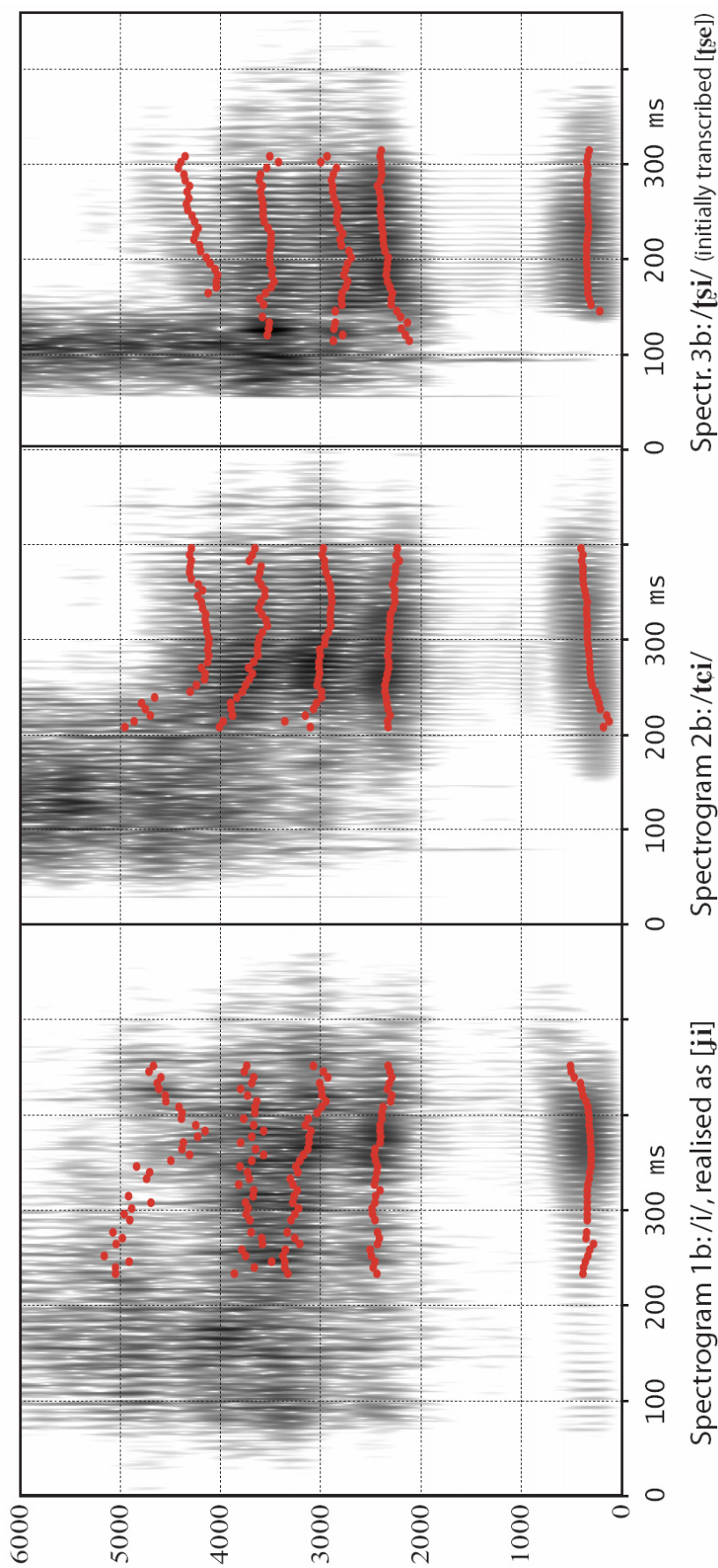
7. References

- [1] Fant, G., 1960. *Acoustic theory of speech production, with calculations based on X-ray studies of Russian articulations*. The Hague/ Paris: Mouton.
- [2] Maeda, S., 1996. Phonemes as concatenable units: VCV synthesis using a vocal-tract synthesizer. In: *Sound patterns of connected speech: Description, models and explanation*. *Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung* 31, A. Simpson and M. Patzod (eds.), Kiel, 127-232.
- [3] Vaissière, J., 2007. Area functions and articulatory modeling as a tool for investigating the articulatory, acoustic and perceptual properties of the contrast between the sounds in a language. In *Experimental Approaches to Phonology, in honor of John Ohala*, P.S. Beddor, M.J. Solé, and M. Ohala (eds.). Oxford: Oxford University Press, 54-72.
- [4] 黄布凡, 1991. 纳木兹语. In 藏缅语十五种, 戴庆厦 *et al.* (eds.). 北京: 北京燕山出版社.
- [5] Michaud, A. and He Xueguang, 2007. Reassociated tones and coalescent syllables in Naxi (Tibeto-Burman). *Journal of the International Phonetic Association* 37(3), 237-255.
- [6] Michaud, A., 2006. Tonal reassociation and rising tonal contours in Naxi. *Linguistics of the Tibeto-Burman Area* 29(1), 61-94.
- [7] Rialland, A.; Sangaré, M.B., 1989. Réanalyse des tons du bambara: des tons du nom à l'organisation générale du système. *Studies in African Linguistics* 20(1), 1-27.
- [8] 齐卡佳 (Ekaterina Chirkova), 2007. 在重音与声调的连续体上看史兴语. *东方语言学 (EastLing)* (2).
- [9] Lidz, L., 2006. A synopsis of Yongning Na (Mosuo). Handout circulated at the 39th International Conference on Sino-Tibetan Languages and Linguistics, University of Washington, Seattle.
- [10] Michailovsky, B.; Michaud, A., 2006. Syllabic inventory of a Western Naxi dialect, and correspondence with Joseph F. Rock's transcriptions. *Cahiers de linguistique - Asie Orientale* 35(1), 3-21.
- [11] Michaud, A., 2006. Three extreme cases of neutralisation: nasality, retroflexion and lip-rounding in Naxi. *Cahiers de linguistique - Asie Orientale* 35(1), 23-55.
- [12] Bradley, D., 1979. *Proto-Loloish*. Scandinavian Institute of Asian Studies, Monograph series n°39. London & Malmö: Curzon Press.
- [13] Ladefoged, P.; Maddieson, I., 1996. *The Sounds of the World's Languages*. Oxford, U.K. & Cambridge, Massachusetts: Blackwell.
- [14] Willerman, R.; Kuhl, P.K., 1996. Cross-language speech perception: Swedish, English, and Spanish speakers' perception of front rounded vowels. In *Proceedings of the 4th International Conference on Spoken Language Processing*, Philadelphia, 442-445.
- [15] Carlson, R.; Granström, B.; Fant, G., 1970. Some studies concerning perception of isolated vowels. *Speech Technology Laboratory Quarterly Progress and Status Report (STL-QPRS)* 2-3, 19-35.
- [16] Bladon, A.; Fant, G., 1978. A two-formant model and the cardinal vowels. *Speech Transmission Laboratory Quarterly Progress Status Report (STL-QPRS)* 1, 1-8.
- [17] Gendrot, C.; Adda-Decker, M.; Vaissière, J., submitted. Les voyelles /i/ et /y/ du français : aspects quantiques et variations formantiques. *Proceedings of Journées d'Etude de la Parole 2008*, Avignon, France.
- [18] Fant, G., 1973. *Speech sounds and features*. Cambridge, Massachusetts: MIT Press.
- [19] Johnson, K., 2004. Speaker normalization in speech perception. In *Handbook of Speech Perception*, D. B. Pisoni, and R. E. Remez (eds). Oxford: Blackwell, 363-389.
- [20] Kluender, K.R., 2003. Sensitivity to change in perception of speech. *Speech Communication* 41, 59-69.
- [21] Peterson, G.E.; Barney, H.L., 1952. Control methods used in a study of the vowels. *Journal of the Acoustical Society of America* 24, 175-184.
- [22] Gendrot, C.; Adda-Decker, M., 2005. Impact of duration on F1/F2 formant values of oral vowels: an automatic analysis of large broadcast news corpora in French and German. *Proceedings of Eurospeech/Interspeech 2005*, Lisboa, 2453-2456.



Spectrogrammes 1a, 2a and 3a are based on data from an adult female subject.

The superimposed dots indicate the frequency of formants F1, F2 and F3 as evaluated by the software PRAAT.



Spectrogram 1b: /i/, realised as [ji]
Spectrogram 2b: /tci/
Spectrogram 3b: /tʂi/ (initially transcribed [tʂe])

The superimposed dots indicate the frequency of formants F1 to F5 as evaluated by the software PRAAT.