

A close-up photograph of a marble sculpture of a woman's head. The woman has long, wavy hair that flows down and around her face. Her eyes are closed, and her mouth is slightly open. The lighting is soft, highlighting the texture of the marble and the intricate details of the hair.

Gina Maria Schneider, Maria Chiara Janner, Bénédicte Élie (eds.)

VOX & SILENTIUM

Études de linguistique et littérature romanes
Studi di linguistica e letteratura romanza
Estudios de lingüística y literatura románicas

Voix et silence se considèrent traditionnellement comme deux phénomènes opposés, s'excluant l'un l'autre. Les contributions contenues dans ce volume se proposent de dépasser une telle conception, en se centrant non seulement sur la valeur et les fonctions que les deux concepts peuvent recouvrir, mais aussi sur la relation complexe qui existe entre eux en linguistique et en littérature. Outre les deux pôles constitués par la voix et le silence, on peut repérer dans le domaine des langues romanes une grande variété de voix silencieuses ou de silences expressifs : la communication non verbale et son interaction avec le langage verbal, les différentes voix (plus ou moins silencieuses) donnant expression à ce qui ne peut pas être dit, ou bien la représentation graphique – et donc apparemment 'muette' – d'un phénomène potentiellement acoustique. Le cri silencieux de Daphné, rendu visible dans la sculpture de Gian Lorenzo Bernini illustrant la couverture, est la manifestation figurative de cette rencontre oxymorique entre la voix et le silence.

GINA MARIA SCHNEIDER a étudié la langue et la littérature espagnoles à Zurich et à Madrid. Elle travaille depuis 2011 comme assistante et chargée de cours à l'Université de Zurich. Sa recherche se centre sur la poésie espagnole du XX^e siècle.

MARIA CHIARA JANNER a étudié la langue et la littérature italiennes à Zurich et à Rome. Elle est doctorante en linguistique italienne à l'Université de Zurich et à la KU Leuven.

BÉNÉDICTE ÉLIE est assistante auprès de la chaire de littérature française moderne à l'Université de Zurich et termine une thèse sur l'Épopée au XIX^e siècle.

VOX & SILENTIUM

Gina Maria Schneider, Maria Chiara Janner, Bénédicte Élie (eds.)

VOX & SILENTIUM

Études de linguistique et littérature romanes

Studi di linguistica e letteratura romanza

Estudios de lingüística y literatura románicas

avec la collaboration de :

Cristina Albizu, Julie Dekens, Mario A. Della Costanza, Cyril Dubois,
Valeria Frei, Charlotte Meisner et Michael Schwarzenbach



PETER LANG

Bern · Berlin · Bruxelles · Frankfurt am Main · New York · Oxford · Wien

Information bibliographique publiée par «Die Deutsche Nationalbibliothek»
«Die Deutsche Nationalbibliothek» répertorie cette publication dans la
«Deutsche Nationalbibliografie»; les données bibliographiques détaillées sont
disponibles sur Internet sous <<http://dnb.d-nb.de>>.

Illustration de couverture : détail de Gian Lorenzo Bernini, *Apollon et Daphné*,
1622-1625, Galleria Borghese, Rome (photographie : Luciano Romano).

ISBN 978-3-0343-1611-8 br.

ISBN 978-3-0351-0823-1 eBook

Cette publication a fait l'objet d'une évaluation par les pairs.

© Peter Lang SA, Editions scientifiques internationales, Berne 2015
Hochfeldstrasse 32, CH-3012 Berne, Suisse
info@peterlang.com, www.peterlang.com

Tous droits réservés.

Cette publication est protégée dans sa totalité par copyright.

Toute utilisation en dehors des strictes limites de la loi sur le copyright est
interdite et punissable sans le consentement explicite de la maison d'édition.
Ceci s'applique en particulier pour les reproductions, traductions, microfilms,
ainsi que le stockage et le traitement sous forme électronique.

Imprimé en Suisse

Sommaire

I Introduction.....	7
II Contributions	15
VOLKER DELLWO	
What does voice and silence tell us about speaker identity? An introduction to temporal speaker individualities and their use for forensic speaker comparison.....	17
FRANCESCA GALLI	
Silenzio, «voce di bocca» e «voce di cuore» nella predicazione in volgare di Giordano da Pisa e Bernardino da Siena.....	37
JULIEN MARSOT	
La Voix et le Silence : l'envers et l'endroit du Verbe fin-de-siècle	51
SIBILLA DESTEFANI	
Il rovesciamento del lessico religioso in Liana Millu	63
PATRICIA RIOSALIDO VILLAR	
La historia acallada y los «ruidos» en: <i>Autobiografía del general Franco</i> de Manuel Vázquez Montalbán.....	73
MARIANA MORAES MEDINA	
«Lo que el silencio nombra»: mística y poesía en la obra de Hugo Mujica	87
MARION COSTE	
Pascal Quignard. Entre le refus du langage et l'impossibilité de se taire, le silence musical	101

TANJA SCHWAN Entre ne rien dire et ne dire que des riens : les poétiques de Flaubert, Manzoni et Clarín	113
SANDRA CARRASCO Voz y silencio en la obra de Manuel Rivas: <i>Todo es silencio</i> (2010).....	129
CLAUDIA CATHOMAS, NICOLA FERRETTI & ANNE-DANIÈLE GAZIN Alternance codique dans les SMS écrits en Suisse italophone et romanchophone.....	147
ELISA TONANI Vedere il silenzio. Forme grafico-interpuntive di rappresentazione della voce e delle sue pause.....	165
LISA CADAMURO Punteggiatura del silenzio in <i>Fortezza</i> di Giovanni Giudici.....	181
SUSANA GONZÁLEZ AKTORIES Partituras silenciosas: pautas de lectura para textos pentagramados	197
GABRIELA SOARE Le rôle des éléments nuls à l'intérieur de quelques syntagmes nominaux et leur effet sur l'accord sujet-verbe en français et en roumain.....	211
VALENTINA BIANCHI Sintassi di gesto e parola: note sull'eloquio afasico.....	227
SABINA FONTANA & ERIKA RANIOLO Interazioni tra oralità e unità segniche: uno studio sulle labializzazioni nella Lingua dei Segni Italiana (LIS).....	241
III Résumés.....	259
IV Collaborateurs.....	273

I
Introduction

Le mythe ovidien d'Apollon et Daphné symbolise l'un des problèmes majeurs de la modernité, à savoir le dialogue entre voix (*vox*) et silence (*silentium*), annoncé dans le titre de ce volume. Approchée et effleurée par Apollon, la nymphe pousse un cri et se transforme en arbuste de laurier, privé de parole. La sculpture de Gian Lorenzo Bernini (1622-1625) interprète cette rencontre oxymorique entre la voix et le silence. Une femme sculptée dans le marbre ne devrait pas pouvoir crier ; or sa représentation, bouche ouverte, regard bouleversé, exprime un véritable cri pour qui la regarde. La matérialité inanimée de la sculpture empêche de percevoir le son et, dans le même temps, en exprime toute la force, tout en se transformant en écoute du silence.

La double valeur inhérente au laurier souligne ce paradoxe. D'un côté, il représente ici le silence de la nymphe Daphné après sa métamorphose. De l'autre, ce sont précisément les feuilles de cet arbre qui couronnent le dieu de la poésie. Pourrait-on conclure que la nymphe silencieuse poursuit sa vie dans la voix des poètes ?

Les notions de voix et de silence gagnent en complexité notamment lorsqu'on les met en relation, puisque le silence se révèle comme une entité qui ne peut pas se définir simplement comme l'absence de voix. En comparaison avec la voix, il prend une dimension particulière qu'on peut spécifier tant en linguistique qu'en littérature.

En linguistique, la signification du silence et de la voix est pertinente à tous les niveaux du système de la langue. La phonétique et la phonologie rendent compte de la description des sons langagiers et s'occupent d'identifier le locuteur à l'aide de sa voix. Au niveau morphosyntaxique, le silence peut être porteur de significations, par exemple par le concept de morphème nul ou d'ellipse. Il existe, en dernier lieu, des formes de communication non verbale ou de paralangage, complétant la langue parlée ou bien s'y suppléant totalement, comme c'est le cas aussi des différentes langues des signes.

Dans le domaine de la littérature, on peut étudier la pluralité des voix dans une œuvre littéraire (plurivocité et polyphonie), la dimension indicible du discours poétique ou bien la signification des pauses et des éléments graphiques dans la versification. La voix et le silence sont cependant aussi des éléments fondamentaux pour l'analyse littéraire, pouvant

rendre compte de divers aspects du texte, que ce soit le silence ou la voix des protagonistes ou du narrateur. Non seulement lorsqu'ils sont envisagés dans une perspective thématique, la voix et le silence peuvent acquérir, par conséquent, diverses significations.

Les contributions de ce volume, présentées lors du VII^e *Dies Romanicus Turicensis* à l'Université de Zurich le 21 et 22 juin 2013, s'intéressent à ces différentes formes de voix ou silence tant en linguistique qu'en littérature. Né en 2003, ce congrès biennal se veut un échange fructueux entre les jeunes chercheurs dans le champ des langues et littératures romanes, et a pour but d'approfondir le débat transdisciplinaire sur un thème spécifique. Dans son édition de 2013, une vingtaine de jeunes chercheurs venant des différentes régions de Suisse, d'Allemagne, de France, d'Italie, d'Espagne, du Canada et du Mexique y ont participé, ce qui prouve le caractère de plus en plus international du *Dies Romanicus Turicensis*.

Le volume commence par l'article de notre conférencier invité, Volker Dellwo, professeur assistant au Laboratoire de phonétique de l'Université de Zurich. Sa contribution, ayant pour titre « What does voice and silence tell us about speaker identity? », envisage les notions de voix et silence du point de vue de la phonétique judiciaire et confirme l'individualité des interlocuteurs à travers les caractéristiques temporelles du langage humain (tels les patrons prosodiques, les aspects rythmiques ou la durée de l'élocution).

Suivent une série de contributions qui ne s'intéressent pas tant à l'individu qu'aux différentes valeurs attribuées à la voix et au silence d'un point de vue littéraire ou culturel, et ce tout au long de l'histoire de la littérature du XIV^e au XX^e siècle. Francesca Galli (Università della Svizzera italiana) ouvre cette première section avec des considérations autour des discours de Giordano da Pisa et de Bernardino da Siena, deux des prédicateurs toscans les plus importants aux XIV^e et XV^e siècles. Partant de l'usage éthique de la parole, elle observe dans leurs sermons une valorisation particulière du silence, présente dès les débuts du christianisme, qui alterne avec un emploi raisonnable de la voix. Vient ensuite l'étude de Julien Marsot (Université de Montréal), centrée sur la poésie française de la fin du XIX^e siècle, qui relie la dévaluation axiologique de la voix par la tradition judéo-chrétienne avec la volonté contraire de l'esthétique décadente

de faire du poème une voix singulière entendue sur la scène du cabaret. À travers l'œuvre de Maurice Rollinat, Marsot analyse comment voix et silence constituent l'envers et l'endroit d'une même modernité fin-de-siècle. Les contributions de Destefani et de Riosalido, de leur côté, étudient la manière dont la littérature redonne une voix à l'histoire inexprimable du XX^e siècle. En partant du livre *Il fumo di Birkenau* de Liana Millu, survivante du camp de concentration de Auschwitz-Birkenau, Sibilla Destefani (Universität Zürich) fait voir comment l'écrivaine arrive à surmonter, à travers le renversement du lexique biblique, le silence occasionné autour de la Shoah. Une thématique comparable est proposée par Patricia Riosalido Villar (Universidad Nacional de Educación a Distancia, UNED) dans son étude centrée sur le livre *Autobiografía del general Franco* (1992) de Manuel Vázquez Montalbán. Elle nous montre dans quelle mesure, à travers un dialogue fictif entre le personnage de Franco et l'antifasciste Marcial Pombo, ce roman donne voix aux perdants de la guerre civile, exprimant ce qui autrefois n'a pas pu être exprimé. Enfin, la problématique de l'indicible au XX^e siècle est reprise dans le travail de Mariana Moraes Medina (Universidad de Navarra), qui termine cette première partie avec une réflexion sur la valeur du silence dans le mysticisme du prêtre-écrivain argentin Hugo Mujica (1942). Le silence étant à l'origine de sa vocation de poète, le désir d'exprimer l'inexprimable – un désir relié à celui de la transcendance – se traduit dans son œuvre littéraire dans une conception de la parole poétique comme entité entourée du vide, du silence.

Dans une deuxième série de contributions, le lecteur observe la présence de la voix et du silence sur un niveau thématique ou narratif, en relation avec les formes d'expression (parfois silencieuses) des différents types de locuteurs. L'essai de Marion Coste (Université Paris III Sorbonne Nouvelle) montre le refus paradoxal du langage observé chez les personnages mutiques ou quasi-mutiques de Pascal Quignard, qui devient un taire sans taire, un « silence expressif ». Particulièrement importante à ce propos, l'expression musicale se transforme en un lieu pour habiter le silence. Si l'essai de Coste se centre sur le silence des personnages qui refusent le langage, Tanja Schwan (Universität Leipzig) se propose de montrer deux autres formes de silence, notamment le fait de ne rien dire (« das Nichtsagen » en tant qu'indicible), produit par le narrateur même, et celui de ne dire que des riens (« das Nichtssagende » en tant qu'échange d'idées reçues dans le brouhaha des personnages et des narrateurs), propre à quelques personnages. Selon une approche comparatiste, elle analyse trois romans

européens : *Madame Bovary* de Flaubert, *La Regenta* de Clarín et *I promessi sposi* d’Alessandro Manzoni. La contribution de Sandra Carrasco (Université St. Gallen) porte sur le roman *Todo es silencio* (2010) de l’écrivain contemporain Manuel Rivas et analyse la relation entre la voix et le silence qui s’y présente non seulement comme un motif mais aussi comme un procédé littéraire. Partant du concrétisme – concept qu’elle définit à l’aide des théories psychanalytiques et sociologiques –, elle esquisse la mise en scène d’une voix narrative incapable de différencier le niveau figuré du langage de celui non figuré. Dans le domaine de la linguistique, enfin, Claudia Cathomas, Nicola Ferretti et Anne-Danièle Gazin (Universität Bern) consacrent leur étude à la présence simultanée de différentes langues dans les SMS écrits en Suisse italophone et romanchophone. Comme le démontrent les auteurs à partir du corpus analysé, bien que les phénomènes de polyphonie puissent se manifester sous diverses formes, elles ont en commun la nécessité d’une compétence bilingue assez limitée et se cantonnent le plus souvent en début et fin de message dans des formules de salutations.

Plusieurs contributions montrent ensuite que la thématique de la voix et du silence ne se manifeste pas seulement au niveau du contenu ou de la narration, mais a aussi des résonances au niveau de l’expression. D’une manière générale, l’étude d’Elisa Tonani (Università degli Studi di Genova), intitulée « *Vedere il silenzio* » (‘Voir le silence’), met l’accent sur la ponctuation dans les textes littéraires, en tant que représentation graphique de la voix et de ses silences. Elle nous fait voir ainsi comment la ponctuation établit un pont entre l’oral et l’écrit, en particulier lorsqu’elle fournit, par le biais d’indications graphico-visuelles, des indications pour la reproduction acoustique. L’aspect de la ponctuation est approfondi par Lisa Cadamuro (Università di Pavia), qui analyse dans la poésie de Giovanni Giudici l’emploi original des deux-points et du trait d’union à la fin du vers. Comme le souligne la chercheuse, dans la mesure où ces deux signes élargissent l’espace vide qui suit, ils conduisent à une véritable « ponctuation du silence » qui relie le niveau de la forme aux thèmes du texte et en renforce la signification. Avec l’étude de Susana González Aktories (Universidad Autónoma de México) le lecteur observe que dans la poésie expérimentale la représentation graphique finit par remplacer le texte dans son intégrité. En partant d’exemples tirés de l’œuvre de Francesco Cangiullo, de Dick Higgins et d’Augusto de Campos, sa contribution montre que la soi-disant ‘poésie pentagrammée’, en tant que « *partitura silenziosa* », est capable de reproduire dans la tête du lecteur-contemplateur une sphère à la fois poétique et musicale, dotée de voix.

Les trois dernières études confirment, finalement, que le silence en linguistique peut acquérir une signification précise qui, dans certains cas, amène à la construction d'une nouvelle voix. Dans l'étude de Gabriela Soare (Université de Genève), ce silence s'articule au niveau de la syntaxe. À partir de différents exemples de la langue française et du roumain, elle examine le rôle des éléments dits « nuls ou silencieux » à l'intérieur de quelques syntagmes nominaux et l'effet qu'ils peuvent avoir pour l'accord sujet-verbe. Il y a, en tout cas, des formes de communication où il est question de véhiculer le message linguistique entier sous forme de langue « silencieuse ». En étudiant la manière dans laquelle les gestes dits co-verbaux se combinent avec la parole dans l'expression de locuteurs ayant un trouble linguistique acquis, Valentina Bianchi (Università per Stranieri di Siena) défend que la complémentarité de la parole et du geste, qui permet de « dire autrement », forme un nouveau système là où l'aphasie a porté à une désagrégation des structures linguistiques. L'article de Sabina Fontana et d'Erika Raniolo (Università di Catania, a Ragusa) se concentre, en dernier lieu, sur un autre aspect de l'interaction entre langage verbal et éléments non-verbaux, en analysant le phénomène des labialisations dans la langue des signes italienne. Les tests réalisés par les deux chercheuses montrent que ces fragments de langue parlée articulés en combinaison avec les signes ne sont pas redondants par rapport aux gestes, mais représentent plutôt un phénomène structurel, cohérent avec la nature multimodale des langues. Entre voix et silence il existe dans ce cas-là, par conséquent, une sorte de complémentarité.

Le comité organisateur du VII^e *Dies Romanicus Turicensis* (Cristina Albizu, Julie Dekens, Mario A. Della Costanza, Cyril Dubois, Bénédicte Élie, Valeria Frei, Maria Chiara Janner, Charlotte Meisner, Gina Maria Schneider et Michael Schwarzenbach) exprime sa gratitude à tous ceux qui ont contribué à la réalisation du congrès et à la publication des actes : le *Romanisches Seminar* de l'Université de Zurich et son programme doctoral « Méthodes et perspectives », ainsi que le *Zürcher Universitätsverein* et le Prof. Dr. em. Georg Bossong. Nous remercions, en particulier, tous les auteurs sans lesquels cette publication n'aurait jamais vu le jour.

Zurich, le 20 décembre 2014

Gina Maria Schneider, Maria Chiara Janner & Bénédicte Élie

II Contributions

What does voice and silence tell us about speaker identity? An introduction to temporal speaker individualities and their use for forensic speaker comparison

VOLKER DELLWO
(Universität Zürich)

1 Introduction

Next to an intellectual interest in why speakers possess individual voices there are applications where such knowledge is vital: Recordings of a criminal's voice during the conduction of crime frequently become evidence in post-crime forensic investigations (Dellwo *et al.* 2014, Rose 2002, Nolan 1997, Künzel 1987). Typical cases range from more trivial issues like recordings of hoax calls to emergency numbers over sexual harassment via the phone to highly serious crime, like calls of kidnappers to victims or the police or phone calls or taped messages from terrorist groups. In such cases a forensic phonetic expert's task is typically to estimate the probability by which two or more speech samples are from the same or different speakers (e.g. whether voice samples from a suspect and a perpetrator are from an identical speaker or not). This process is referred to as forensic phonetic voice comparison (Nolan 1997). Forensic phonetic voice comparison is based on the one hand on a systematic analysis of auditory characteristics concerning a speaker's dialect, sociolect, emotional state, health condition, etc. (Jessen 2007) and, on the other hand, expert witnesses use acoustic measurements to obtain data that is typically not retrievable by the human ear (e.g. average frequencies of vocal tract resonances or their dynamics). In the present article we are particularly concerned with retrieving evidence from the acoustic analysis of speech.

Forensic phonetic speaker comparison underlies the assumption that speech contains speaker idiosyncratic or biometric information (Dellwo

et al. 2014, Dellwo, Huckvale & Ashby 2007, Nolan 1991, 1997). This assumption is supported by everyday observations, which demonstrate that it is to a high degree possible to identify individuals on the basis of their voice only (e.g. a caller on the phone may be identified on the basis of a simple 'hi'). Even though human speaker identification abilities are far from perfect and subject to human perceptual error (Köster & Schiller 1997, Eriksson *et al.* 2010), it is evident that there must be some speaker specific information in the speech signal responsible for the basic speaker identification ability of listeners. Such parameters are, for example, frequency-domain or spectral parameters like fundamental frequency (f_0) and its harmonics and the resonance frequencies of the different vocal tract cavities (formant frequencies). The rationale behind these characteristics is that they are to some degree a result of the individual physical dimensions of a speaker's organs of speech. A long vocal tract and a large larynx typically lead to overall lower vocal tract resonances (i.e. lower formant frequencies) and a lower rate of vocal fold vibrations (i.e. lower f_0) respectively. The most central problem connected to these characteristics is that they are highly variable and used for a number of different types of information transfer in speech (Dellwo *et al.* 2014, Nolan 1991). Additionally speakers also have the tendency to adapt the way they realize the acoustic signal in accordance to their environment (e.g. in silent environment or with loud background noise), their mental state (e.g. in emotional situations of sadness), their state of health (e.g. when suffering a cold), physical obstruction of their organs of speech (e.g. with a cigarette or a pencil between the lips), etc. Numerous other influential factors are observable. For this reason, both the above mentioned spectral parameters, fundamental frequency and formant frequencies, are prone to high within-speaker variability. In particular f_0 can vary tremendously as a result of a speaker's emotional state, for example, which is a common problem for forensically relevant recordings as speakers are not seldom under some form of stress when recorded during a crime. Additionally formant and f_0 characteristics can be very similar between speakers which share similar vocal tract configurations and larynx size (e.g. in father and son, mother and daughter, or siblings of the same sex). For this reason the acoustic identification of speakers based on such parameters is often limited and it can typically not be excluded that two voices which sound and are measurably very similar are from different speakers (between-speaker similarity); nor can it be excluded that voice samples which sound and are

measurably very different are from the same speaker (within-speaker variability). As such, speech evidence is not as unambiguous as other widely obtained forensic evidence like fingerprints or DNA.

Given the discussion in the previous paragraph it is evidently the case that we need to extend our knowledge about speaker individual characteristics to different dimensions. In the following we will first provide a rationale for why temporal characteristics should matter for explaining speaker individuality (2) and we will then discuss which temporal characteristics of speech are most likely to reveal speaker individualities (3). Finally we will demonstrate a number of case studies which support the point that temporal characteristics enhance our understanding of speaker individuality.

2 Why should speakers reveal temporal individualities in speech?

Speech is produced by a large number of brain controlled muscle movements that steer articulators like the tongue, jaw, velum, larynx, lips, etc. in a very complex fashion (Perrier 2012, Dellwo, Huckvale & Ashby 2007). All movements must be carefully controlled by a large number of motor actions. As such human speech articulation is a very similar process to other motor controlled actions as arm or leg movements and it has been demonstrated that such movements can be idiosyncratic (McDougall 2006). A particular analogy can probably be drawn with human gait (Anders Eriksson, personal communication), which requires a careful control over a large number of muscles to balance the entire body correctly during walking. In the case of gait it has been demonstrated in great detail that the temporal information of the involved movements are highly walker idiosyncratic and that humans can be identified with high accuracy based on the observation of three discrete points along their legs (Loula *et al.* 2005, Matovski *et al.* 2012, Bouchrika & Nixon 2008). Studies further revealed that even a front recording of human gait, for example, can deliver enough detail for high identification performance (Goffredo, Carter & Nixon 2008). But also gait recognition faces problems of within-walker

variability in movement or clothing; however, such problems can increasingly be reduced (Matovski *et al.* 2012).

Given these promising results from gait research, it seems conceivable that similar idiosyncratic movements are performed during the control of the human articulators and that speakers should be identifiable by the way they move their jaw, for example (Mark Nixon, personal communication). A most obvious idea might thus be to copy techniques from human gait recognition and apply them to human articulator movement recognition. This would imply to define some points along the human articulators like the jaw and the lips and apply the mathematic processing from gait identification. Such research is potentially interesting from a theoretical point of view and it might enhance our understanding of the articulation process. For forensic contexts, however, it lacks applicability as the necessary articulatory movement data for carrying out a speaker comparison will typically not be available in forensic casework (and if some articulatory movements are available, e.g. jaw movements through CCTV recordings, the low resolution of the recordings and the small quantity of observed articulators movement will limit the recognition options severely). We argue, however, that we can already go one step further as the timing of the articulators is directly responsible for the temporal characteristics of the speech signal. In other words, it is in the acoustic signal where we should be able to obtain speaker individualities arising from articulatory movements. Evidence for such assumptions and the relevance for speaker identification purposes has been provided by McDougall (2004, 2006) who demonstrated that temporal features of segmental formant movements (formant dynamics) are highly speaker specific.

3 Speaker specific speech rhythmic characteristics

Next to segmental characteristics there is a higher-level temporal organization of the articulatory process over the course of an utterance. Such suprasegmental timing characteristics are generally referred to as speech rhythm. Basic actions like the onset and offset of vocal fold vibration or the on- and offset of certain vowels and consonants might as well

be influenced by idiosyncratic timing of the articulators as might be the temporal distribution of syllable peaks or the temporal development of f_0 rises, etc. Support for such a view can be gained from the field of speech rhythm research. In the following we will discuss two types of rhythmic properties, one that is based on human phonological knowledge about speech (consonantal and vocalic intervals; 3.1) and another that is based on bottom-up acoustic characteristics of the signal (voice and silences; 3.2).

3.1 Rhythmic characteristics derived from consonantal and vocalic durations

Measures based on durational characteristics of vocalic and consonantal segments have been argued to be acoustic correlates of a language's rhythm class, i.e. whether a language is stress- or syllable-timed, for example (Ramus *et al.* 1999a, 1999b; Grabe & Low 2002). Most of these measures either calculate the durational variability of consonantal and/or vocalic intervals (e.g. ΔC and ΔV , the standard deviation of consonantal and vocalic intervals respectively, Ramus *et al.* 1999b; or rate normalized version of these measures, VarcoC: Dellwo 2006; VarcoV: White & Mattys 2007) or are simple ratio measures between the proportion of time over which speech is vocalic as opposed to consonantal (%V). Other measures obtain the local variability between consecutive consonantal or vocalic intervals (Pairwise Variability Index, PVI; Grabe & Low 2002). A detailed overview over the measures is provided in Dellwo (2010). Whether such measures are strong correlates of speech rhythm and whether they capture between language rhythmic differences is a matter of heavy debate (Arvaniti 2009, Wiget *et al.* 2010, White & Mattys 2007, Dellwo 2010) as is the whole question of whether categorical rhythmic differences between languages at all exist (see discussion in Arvaniti 2009). This discussion, however, is irrelevant in the present context and for simplicity reasons the measures are referred to as 'rhythm measures' in the remainder of the paper. Of importance is the fact that there is an increasing number of findings from different datasets providing evidence that measures like %V, ΔV or the PVI vary significantly between speakers (Wiget *et al.* 2010, Yoon 2010, Dellwo *et al.* 2012a). Wiget *et al.* (2010) showed for five speakers of English that there is a significant variability of %V and VarcoV between speakers but not for the vocalic

variability measure nPVI. Yoon (2010) analyzed 10 speakers from the same language variety of Standard Northern American English (Buckeye corpus). He found significant differences in %V across the speakers. Dellwo *et al.* (2012a) investigated 8 speakers of Swiss German under different speaking style conditions and found speaker individualities for a variety of temporal characteristics (see below). Such results give rise to the assumption that temporal measures of speech are influenced by individual timing strategies that are possibly related to idiosyncratic timing behavior of the articulators.

3.2 *Rhythmic characteristics derived from voiced and voiceless intervals*

To measure durations of consonantal and vocalic intervals, on- and off-sets of these intervals need to be labeled individually as procedures to do this process automatically have been demonstrated to be too imprecise (see discussion in Dellwo 2010). In Dellwo, Fourcin & Abberton (2007) and Dellwo & Fourcin (2013) we thus tried a different approach. Instead of basing the rhythm metrics described above on vocalic and consonantal durations we based them on voiced and voiceless sections of the signal. We used metrics from Ramus *et al.* (1999b) like the percentage over which speech is vocalic (%V) and the standard deviation of consonantal and vocalic intervals (deltaC and deltaV; we also took the rate normalized variants of this measure VarcoC and VarcoV, see Dellwo 2006) created %VO (the percentage over which speech is voiced) and, for example, deltaUV (the standard deviation of unvoiced intervals; see Dellwo, Fourcin & Abberton 2007, for the rationale). As such we turned from higher level phonetic detail (consonants and vowels) to lower level phonetic detail (voicing patterns). We provided evidence that languages show similar differences based on voiced/voiceless intervals as they show for measures based on consonantal and vocalic intervals by Ramus *et al.* (1999b). In Dellwo & Fourcin (2013) we showed that such characteristics may also vary between speakers. As such the temporal organization of voiced and unvoiced parts of speech seems to be steered by speaker individual behavior, probably related to individual articulatory processes. Below we provide further evidence for such variables.

4 Case studies on rhythmic differences between speakers

In the following we reviewed data from our own studies on rhythmic speaker individualities and provide evidence for temporal measures as described in 3 to reveal speaker specific characteristics. We tested the influence of the native language on such characteristics in bilinguals (4.1), the influence of deliberate voice disguise (4.2), we further collected evidence from larger datasets (4.3) and created our own tailored datasets for studying speaker individual rhythmic characteristics (4.4). In addition, we are studying new ways for measuring the dynamics of temporal variability in the speech signal (4.5).

4.1 *Evidence from second language (L2) and bilingual speech*

In research on the development of speech rhythm in second language acquisition (Dellwo, Gutiérrez Díez & Gavalda 2009, Gutiérrez Díez *et al.* 2008) we aimed at investigating whether acoustically measurable rhythm changes over the course of an acquisition period of a second language. In a longitudinal study we therefore recorded Spanish learners of English pre and post a one year English instruction period at Murcia University. The data situation turned out slightly problematic as only nine out of thirty speakers were available after the instruction period and of these nine speakers some took considerable instruction in English speaking countries while others remained in Spain. We thus dealt with a rather heterogeneous group. However, as a group and in most individual cases we found that there was no significant change in rhythmic measurements like %V or variability measures like deltaC, deltaV or the PVI (Dellwo, Gutiérrez Díez & Gavalda 2009). This was also true for the speakers who gained a drastic increase in English competence during the instruction period. For the second language research this might have been a disappointing result, however, for the present investigation it serves as further support for the view that rhythmic measurements are speaker specific and do not change with competence in a language.

The evidence from the L2 domain was further supported by results from the speech of bilinguals (Dellwo *et al.* 2012b). We investigated how temporal characteristics of speech can contribute to speaker individuality for the same speaker speaking in different languages by studying the

speech of bilingual Italian/German speakers. Our assumption was that if speaker idiosyncratic rhythmic characteristics exist, then they should be present across utterances from different languages produced by the same speaker. 5 Italian-German bilinguals, 5 German natives with L2 Italian and 5 Italian natives with L2 German were recorded reading 10 sentences (about 15 syllables on average) in each Italian and German. Speakers were selected to be similarly competent in their L2 across the two native language conditions. Durational characteristics of consonantal and vocalic intervals (e.g. syllable rate, %V, VarcoC and V, PVI, etc.) as well as voiced and unvoiced intervals (e.g. %Voiced, VarcoVoiced) were calculated for each sentence.

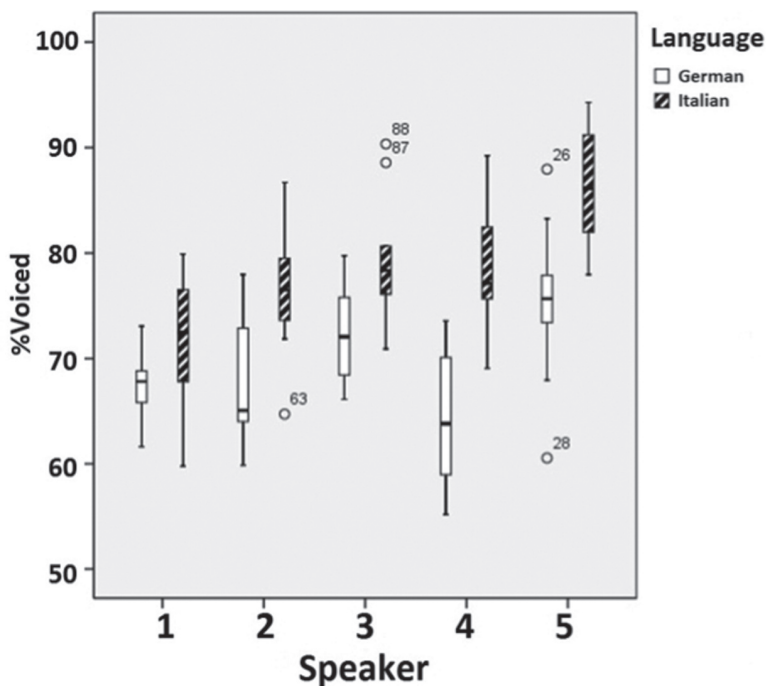


Figure 1: Box plot showing %Voiced for 5 bilinguals speaking German (white plots) and Italian (striped plots).

The most supportive result in this study was found to be %VO which is plotted in Figure 1. Apart from the German condition of speaker 5 it can be observed that %Voiced has the tendency to increase from the

leftmost to the rightmost speaker in both languages. A factorial ANOVA (language * speaker) revealed that there is no interaction between the factors ($p = 0.065$; the relatively low p -value is assumed to be caused by the German of speaker 4). However, highly significant main effects were found for both language ($p < .001$) and speaker ($p < .001$). The effects could be replicated for other metrics (e.g. VarcoV or syllable rate). These results clearly supported the view that %VO varies between languages (lower in Italian compared to German) but it also varies significantly between speakers independent of language. This is to say that speakers with a high %VO in one language also tend to have a high %VO in another language. It thus seems conceivable that %VO might be influenced by idiosyncratic articulatory processes that are carried over between languages.

4.2 The influence of voice disguise on temporal measurements of speech

Within-speaker variability in forensic data is often caused by speakers disguising their voice to increase the chance not to be identified. We therefore wanted to know whether voice disguise has an effect on temporal measures like %V and other rhythm metrics. In a single subject pilot study we asked a speaker of Standard Northern British English to read 120 English sentences once in his normal accent and then in a disguised voice. The form of disguise was the speaker's choice, and he chose to imitate a Mancunian accent. Disguised material was collected until we had a set of 30 sentences for which expert phonetic listeners disputed the speaker identity between disguised and non-disguised samples based on informal auditory listening exercises. An acoustic comparison was then performed between the 30 disguised sentences and the respective 30 non-disguised peers. Results revealed that spectral parameters, in particular fundamental frequency characteristics (mean and variability) varied significantly between normal and disguised voice (mean f_0 and normalized standard-deviation were higher in disguise).

Figure 2 contains the results for the temporal measures %V. The top box-plot in the figure reveals that there was close to no variability between the two conditions. A t -test provided support for this observation showing no significant variability between the groups. It can be seen in the figure that there is high within-condition variability. Figure 2 (bottom), however, reveals that this variability is almost exclusively a result of the influence of

sentence which reflects the strong between sentence variability found in rhythmic measurements (cf. Wiget *et al.* 2010, Dellwo 2010).

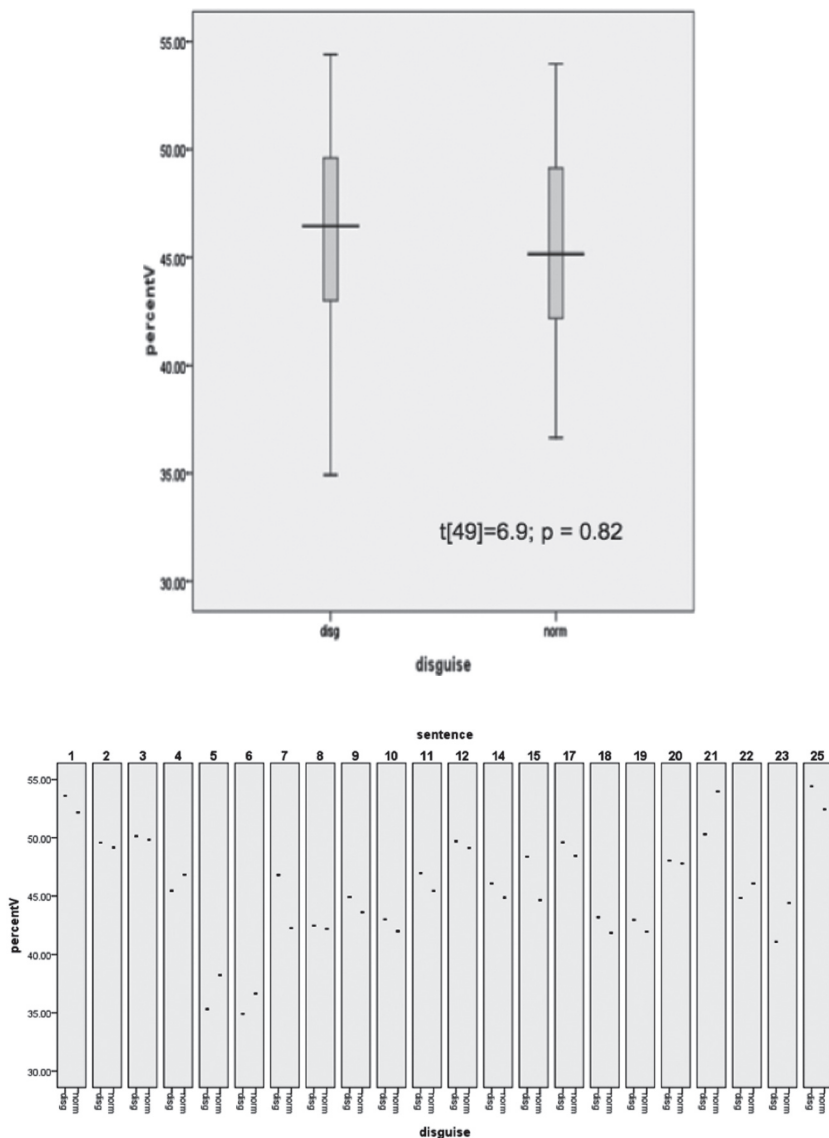


Figure 2: %V for disguised and non-disguised speech of a male English speaker (top) and the same data for each of 25 sentences (bottom).

4.3 Evidence from larger databases

In addition to the data analyzed so far we started looking into larger databases such as the TIMIT corpus (Texas Instruments, TI, and Massachusetts Institute of Technology, MIT). This corpus was designed with speech recognition investigations in mind and contains a large number of speakers (630) of eight different American dialects. Each speaker reads 10 sentences, however, only two sentences are the same across all speakers. The advantage of this corpus is that all sentences are segment annotated thus temporal measurements can be processed quickly. Results can be seen in Figure 3. The box plots in the figures show %V as a function of dialect (left) and as a function of speaker within dialect 1 (right). Results are based on the two sentences that are equal across all speakers. The graphs reveal that there is an extremely low variability of %V between dialects while there is considerable variability between speakers, for example, in dialect 1 (left graph). These results provide further support for the view that temporal measures like %V are speaker idiosyncratic. The disadvantage of this data, however, is that only two of the sentences are the same across all speakers. As there is a large variability of rhythmic measurements of the type of %V or deltaV across different sentences (Wiget *et al.* 2010, Dellwo 2010) it is vital to compare equal sentence material across the speakers at the stage we are now.

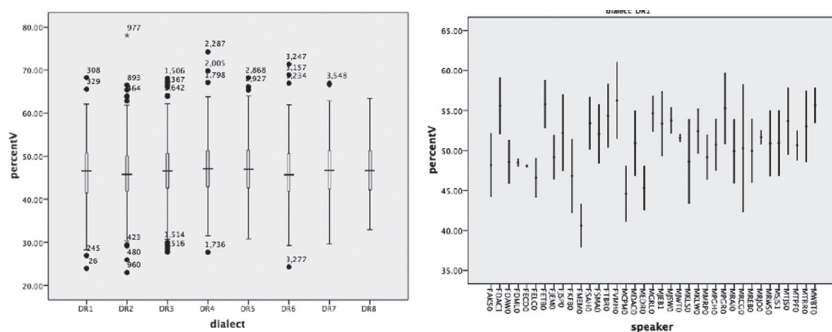


Figure 3: Left: %V across 8 different dialects in the TIMIT database. Right: %V as a function of 37 speakers (2 sentences/speaker) of dialect number 1 in the TIMIT database.