Published in Revue Tranel (Travaux neuchâtelois de linguistique) 59, 149-166, 2013
which should be used for any reference to this work

Rhythmische Variabilität bei synchronem Sprechen und ihre Bedeutung für die forensische Sprecheridentifizierung

Daniel FRIEDRICHS und Volker DELLWO
Phonetisches Laboratorium der Universität Zürich

Humans are able to speak in synchrony with each other. The present study investigated whether four temporal correlates of speech rhythm (%V, nPVI-C, nPVI-V, nPVI-CV) adapt when speaking in synchrony with a recording. The experimental setting consisted of three conditions: Eight participants read three sentences out loud *(read)*, four of them were asked to speak in synchrony *(sync)* with the same sentences recorded by the four others *(target)*. Correlation analysis was carried out between the rhythmic measurements of the sync condition and the two others (read/target). Results revealed that there are typically strong significant correlations between speech rhythm of the read and synchronous speech. The impact on speaker identification, in particular under forensic circumstances, is discussed.

1. Einführung

1.1 Forensische Sprecheridentifizierung

Die Grundannahme der forensischen Sprecheridentifizierung ist, dass die Möglichkeit besteht, einen Sprecher anhand idiosynkratischer oder biometrischer Informationen in seiner Stimme zu identifizieren (Nolan, 1997; Dellwo *et al.*, 2007). Diese Annahme wird durch alltägliche Erfahrungen unterstützt. So ist es für einen Menschen z.B. recht einfach, einen Anrufer (auch ohne visuellen Kontakt) bereits anhand eines einfachen “Hallo“ zu erkennen. Bei dem Versuch einen Sprecher zu identifizieren, nahmen über die letzten Jahre spektrale Charakteristika der Stimme die entscheidende Rolle ein (Nolan, 1991; McDougall, 2007a; Nolan *et al.*, 2009). Beispiele für diese messbaren Erkennungsmerkmale sind die Grundfrequenz (f0) oder die Formantfrequenzen, welche in den Resonanzspektren der menschlichen Stimme entstehen. Dass in deren Ausprägung sprecherspezifische Merkmale repräsentiert werden können, erscheint durchaus nachvollziehbar, wenn man bedenkt, dass sie in enger Verbindung mit der anatomischen Beschaffenheit eines Sprechers stehen dürften. So sollte gerade die Kombination von anatomisch spezifischen Eigenschaften, wie z.B. die Grösse des Kehlkopfes oder die Länge des Vokaltraktes, verständlicherweise auch zu einer höchst individuellen Stimme führen. Nun wurde jedoch bereits gezeigt, dass die spektralen Charakteristika dennoch einer hohen Variabilität unterliegen können

(Dellwo *et al.*, 2007). Das bedeutet Stimmen können nicht nur zwischen unterschiedlichen Sprechern (Inter-Sprecher-Variabilität), sondern auch für einen einzelnen Sprecher (Intra-Sprecher-Variabilität) unterschieden werden und somit differierende Ergebnisse in einer spektralen Analyse hervorbringen. Zur Veranschaulichung können einfache Beispiele dienen: Der emotionale Zustand eines Sprechers (z.B. Angst, Wut etc.) oder eine einfache physische Veränderung (z.B. beim Sprechen mit einem Stift zwischen den Lippen) kann bereits signifikante Abweichungen hervorrufen. Da Aufnahmen von zu identifizierenden Personen (z.B. Tatverdächtigen) nicht immer unter optimalen und vergleichbaren Bedingungen entstehen, werden schnell die Schwächen einer einseitigen, spektralen Sprecheridentifizierung erkennbar. In den vergangenen Jahren wurden daher Versuche unternommen, die Methoden der Sprecheridentifizierung zu verbessern. Es hat z.B. Bemühungen gegeben, Regelmässigkeiten und Zusammenhänge in der Variabilität (Nolan *et al.*, 2009) und dynamische und statische Darstellungsformen der spektralen Stimmcharakteristika zu finden (McDougall, 2007a, 2007b).

Die hier präsentierte Studie verfolgt den Ansatz, die forensische Sprecheridentifizierung um den Parameter *Zeit* zu erweitern. Hierzu sollen vokalische und konsonantische Intervalle von Äusserungen darauf untersucht werden, ob sie spezifische Informationen über einen Sprecher transportieren. Diese Überlegung ist durchaus begründet. Die Produktion von Sprache geschieht über eine komplexe motorische Steuerung einzelner Muskelbewegungen durch das Gehirn. Somit besteht eine Analogie zur Steuerung anderer Muskelbewegungen, wie z.B. der motorischen Kontrolle über Arme und Beine. Und eben diese Bewegungsabläufe von Gliedmassen können in ihrer zeitlichen Abfolge für einen Menschen spezifisch sein (Cunado *et al.*, 2003; Foster *et al.*, 2003). Von dieser Analogie ausgehend besteht Anlass zu der Vermutung, dass es sich ganz ähnlich mit der ebenfalls muskulär gesteuerten Produktion einer Stimme verhalten könnte. Weitere Hinweise darauf finden sich in den Arbeiten von McDougall (2007a, 2007b). Auch sie erkennt den Zusammenhang von muskulären Körperbewegungen und artikulatorischer Produktion von Sprache. Ferner kann er sogar zeigen, dass die muskulär initiierte und gesteuerte Artikulation einen Einfluss auf die zeitliche Aussteuerung der Formantfrequenzen haben muss. Dellwo *et al.* (2009) konnten zudem beobachten, dass beispielsweise der prozentuale Anteil einer vokalischen Äusserung (%V, nach Ramus *et al.*, 1999) selbst bei der Imitation einer fremden Stimme relativ konstant zu bleiben scheint. Somit gibt es einen konkreten Hinweis darauf, dass ein auf zeitlichen Intervallen basierendes Mass eine wichtige Rolle für die Sprecheridentifizierung spielen könnte. In dieser Studie sollen daher suprasegmentale Zusammenhänge von Sprache im Hinblick auf ihre temporale Spezifität untersucht werden.

Um eine temporale Veränderung der natürlichen Stimme herbeizuführen, bietet sich das synchrone Sprechen mit Audioaufnahmen von Zielsprechern an. Hierdurch kann effektiv Einfluss auf die zeitliche Konstruktion des Stimmsignals genommen werden.

1.2 Synchrones Sprechen

Menschen können ohne grosse Mühe synchron sprechen oder singen. Die einfachste Form synchronen Sprechens wird erzeugt, wenn zwei Sprecher einen Text gemeinsam vorlesen. In einem dynamischen Prozess, bei dem es keine eindeutige "leader-follower-relation" gibt (Cummins, 2009), passen sich hierbei beide Sprecher einander zeitlich bis zu einem hohen Grad an (Cummins, 2003; Krivokapic, 2007). Die experimentell ermittelte Asynchronität bewegt sich in derartigen Fällen in einer Spanne von lediglich 9 bis 70ms (Crystal, 1982; Cummins, 2002). Auch ohne Übung und Vertrautheit mit dem Text ist die Verzögerung nicht wesentlich grösser (Cummins, 2003). Obwohl die menschliche Stimme höchst individuell ist, scheint sie somit auch überaus anpassungsfähig zu sein. Doch wie weit reicht diese Anpassungsfähigkeit? Es stellt sich die Frage, ob sich in der menschlichen Stimme auch bei einer derart starken Angleichung noch zeitliche Merkmale finden lassen, die auf einen Sprecher zurückzuführen sind. Um dieser Frage nachzugehen, wird die Sprechersynchronisierung nicht als dynamischer Prozess, sondern als einseitiger Anpassungsversuch untersucht. Dies kann am einfachsten durch die Synchronisierung mit einer Aufnahme erfolgen. Auch wenn bereits experimentell gezeigt werden konnte, dass mit einer Aufnahme ein sehr hoher Grad an Synchronität erreicht werden kann (Cummins, 2009), muss darauf hingewiesen werden, dass diese Methode einige Schwierigkeiten für die Versuchspersonen birgt, da im Gegensatz zum dynamischen Prozess kein Entgegenkommen eines Sprechpartners zu erwarten ist (Poore & Ferguson, 2008). Um adäquates Material zu generieren, muss somit ein Einüben gestattet sein.

Da bei der vorliegenden Studie Probanden versucht haben, sich mit einer Aufnahme zu synchronisieren, handelt es sich ferner um ein Experiment, das Aspekte des *begleitenden Nachsprechens* einschliesst, welches in der Forschung als *shadowing* bezeichnet wird (für einen genaueren Überblick und eine umfangreiche Einführung siehe Marslen-Wilson, 1973). Beim *shadowing* wird der Versuch unternommen, einem auditiven Stimulus mit der eigenen Stimme als *Schatten* zu folgen. Wie schon frühe Studien zeigen, ist dies für kognitiv gesunde Probanden ohne Probleme möglich (Alekin, 1962; Porter & Lubker 1982). Um nun bei synchronem Sprechen Rückschlüsse auf zeitliche Anpassungsphänomene festzustellen, bietet sich zudem die Einbeziehung von Forschungsergebnissen auf dem Feld der Sprachrhythmusforschung an. Die hier dargebrachten akustischen Rhythmuskorrelate beruhen nämlich allesamt auf zeitlichen Intervallen.

1.3 Akustische Rhythmuskorrelate

Die in dieser Studie untersuchten Rhythmuskorrelate beruhen auf den Forschungsarbeiten von Ramus *et al*. (1999), Grabe und Low (2002) und Barry *et al*. (2003). Nachdem zu Beginn der 1990er-Jahre die wissenschaftliche Diskussion über die Einteilung von Sprachen in Rhythmusklassen mit dem vorläufigen Ergebnis endete, dass man keine Möglichkeiten mehr sah, den Rhythmus einer Sprache über das Sprechsignal zu messen und zu beschreiben sowie die allgemeine Annahme herrschte, Rhythmus müsse ein rein perzeptives Phänomen sein, dass mit bisherigen Beobachtungen nicht zu erklären sei (Auer, 1993), brachte die Berechnung auf Grundlage neuer empirischer Methoden von akustischen Rhythmuskorrelaten wieder Bewegung in die Sprachrhythmusforschung. In Anlehnung an Dauer (1987) verwarfen Ramus *et al*. (1999) das Konzept, welches Akzent und Silbendauer für den Sprachrhythmus zugrunde legte und präsentierten ein rein phonetisches Modell. Grundlage hierfür war die von Roach (1982) formulierte Annahme, dass der Eindruck von Rhythmus durch Vokalreduktion und die Varianz der Silbenstruktur in einer Sprache entstehe. So erklären sich die von Ramus *et al*. (1999) ermittelten akustischen Rhythmuskorrelate (%V, ΔC, ΔV), welche konsonantische und vokalische Intervalle ins Zentrum der Beobachtung rücken.

Grabe und Low (2002) präsentieren nur wenig später Korrelate, welche ebenfalls auf der Segmentierung konsonantischer und vokalischer Intervalle beruhen und deren Dauer berücksichtigen, indem sie einen paarweisen Index für deren Variabilität (rPVI) berechnen. Dieses “rohe” Variabilitätsmass (r für Engl. raw = roh) wird von ihnen ferner für die Sprechgeschwindigkeit normalisiert (nPVI) und kann sowohl für konsonantische (nPVI-C) als auch für vokalische Intervalle (nPVI-V) berechnet werden.

Barry *et al*. (2003) schlagen wenig später ein weiteres PVI-Korrelat vor, welches sowohl die vokalischen als auch konsonantischen Intervalle einbezieht (nPVI-CV). Dies sei nötig, um dem auditiven Effekt Rechnung zu tragen, der bei der Kombination beider Lautklassen während des Sprechaktes entstünde.

Für die hier präsentierte Studie wurden aus diesen Forschungsarbeiten vier Rhythmuskorrelate ausgewählt, welche für die Sprechgeschwindigkeit normalisiert wurden (nPVI-Masse) oder sich ihr gegenüber als relativ resistent erwiesen haben (%V). Korrelate, die nicht in diese Kategorien fallen (dies sind z.B. ΔC, ΔV, rPVI etc.), wurden nicht berücksichtigt, da ihre Ergebnisse auf keiner einheitlichen Basis beruhen würden, d.h. die vorgelesenen Sätze der Versuchspersonen in ihrer Dauer natürlich von der Dauer der Zielsätze, mit denen es sich zu synchronisieren galt, abweichen.

Folgende vier Rhythmuskorrelate wurden daher im Rahmen dieser Studie untersucht:

- %V (Ramus *et al.*, 1999), der prozentuale vokalische Anteil einer Äußerung.

- nPVI-C, nPVI-V (Grabe & Low, 2002) und nPVI-CV (Barry *et al.*, 2003), paarweiser Index für die Variabilität der Dauer von vokalischen (V) und konsonantischen Intervallen (C), welcher für die Sprechgeschwindigkeit normalisiert wurde. Dieser wird folgendermassen berechnet:

$$nPVI = 100 \times \left[\sum_{k=1}^{m-1} \left| \frac{d_k - d_{k+1}}{(d_k - d_{k+1})/2} \right| \Big/ (m-1) \right]$$

m steht hierbei für die Anzahl der Intervalle und d stellt die Dauer des k-ten Intervalls dar.

2. Daten und Methoden

2.1 Sprecher

Für das Experiment wurden zwei Gruppen von insgesamt 8 Sprechern untersucht. Die erste Gruppe (Sprecher 1-4; je zwei weibliche und männliche Probanden) im Alter von 20-30 Jahren wurde aus dem Korpus des Phonetischen Laboratoriums der Universität Zürich anhand hoher differierender %V-Werte ausgewählt. Um eine möglichst hohe Variabilität der Zielwerte zu erreichen, wurden zwei Deutsch-Muttersprachler (Sprecher 1 und 2) und zwei Italienisch-Muttersprachler (Sprecher 3 und 4) mit hohem Deutsch-L2-Niveau ausgesucht. Durch diese Vorauswahl sollte der Einfluss der Synchronisierung auf zumindest ein akustisches Rhythmusmass besser beobachtet werden können. %V wurde gewählt, da es sich in den zuvor präsentierten Studien (z.B. Dellwo *et al.*, 2009) als ein besonders resistenter Faktor gegenüber externen Einflüssen, wie beispielsweise der Sprechgeschwindigkeit, gezeigt hatte. Die zweite Gruppe von Probanden (Sprecher 5-8; eine weibliche Probandin und drei männliche Probanden) im Alter von 20-25 Jahren waren durchweg Deutsch-Muttersprachler (Sprecher 6-7 Standarddeutsch, 5 und 8 Schweizerdeutsch).

2.2 Material

Alle Versuchspersonen wurden gebeten, die folgenden drei Sätze im eigenen Rhythmus und ohne jegliche Vorgaben vorzulesen:

(1) Die Frau des Apothekers weiss immer was sie will.

(2) Das Theater hat viele neue Aufführungen geplant.

(3) Er wollte sich seiner Schwächen einfach nicht bewusst werden.

Die Sprecher der zweiten Gruppe mussten sich nach einer kurzen Pause mit den Versionen jedes einzelnen Satzes der Sprecher 1-4 (i.e., Gruppe 1) synchronisieren. Durch die Berechnung der akustischen Rhythmuskorrelate erhielten wir drei unterschiedliche Konditionen für die spätere Auswertung. Die Ergebnisse der vorgelesenen Versionen von Sprecher 5-8 dienten als Ausgangswerte (read), die der gelesenen Sätze von Sprecher 1-4 als Zielwerte (target) und die Messwerte der Synchronisierungsversuche von Sprecher 5-8 als Vergleichswerte (sync).

2.3 Versuchsablauf

Die Aufnahmen wurden in einer Audiometriekabine des Phonetischen Laboratoriums der Universität Zürich durchgeführt. Die Synchronsprecher (sync) bekamen während der zweiten Phase des Experiments die Zielsprecher (target) über Kopfhörer als Stimuli zu hören. Die Probanden benutzten hierzu halboffene Kopfhörer, um die eigene Stimme während der Einspielungen noch hören zu können. Dieses Verfahren sollte eine möglichst hohe Sensibilisierung der Sprachproduktion ermöglichen. Da sowohl die eigene Stimme als auch die Zielstimme während des Experimentes akustisch wahrnehmbar waren, bestand eine grössere Chance, während des Versuches eine Asynchronität zu korrigieren.

Ferner wurde jeder Satz eines Zielsprechers (target) fünf mal eingespielt. Eingeleitet wurde er von drei 1kHz-Tönen im gleichbleibenden Abstand von 500ms. Der erste dieser fünf Stimuli diente zur Orientierung und musste noch nicht synchronisiert werden. Für die Auswertung wurde später stets der letzte erfolgreiche Synchronisierungsversuch verwendet, da durch den mehrmaligen Versuch bzw. durch die Einübung ein höherer Grad an Synchronität gegeben war. Lediglich in zwei Fällen musste die dritte, in einem Fall die zweite Aufnahme verwendet werden, da sich der Proband entweder versprochen oder den Einsatz verpasst hatte. Somit ergab sich ein Korpus von 24 gelesen Sätzen (8x3 read-Versionen) und 48 synchron gesprochenen Sätzen (4x4x3 sync-Versionen). Insgesamt dauerte das Experiment ungefähr 14 Minuten, so dass eine Ermüdung oder das Nachlassen der Konzentration der Versuchspersonen weitgehend ausgeschlossen werden konnte.

2.4 Aufbereitung der Daten

Die Daten wurden mit der Audioproduktionssoftware Pro Tools (www.avid.com) auf ein bzw. zwei Kanälen aufgenommen. Die Segmentierung der einzelnen Sätze in vokalische und konsonantische Intervalle erfolgte in Praat (www.praat.org) manuell durch den ersten Autor. Vokal- und Konsonantencluster wurden jeweils zu vokalischen bzw. konsonantischen Einheiten zusammengezogen, dessen Dauern dann gemessen werden konnten. Dargestellt wird der Segmentierungsprozess in Abbildung 1.

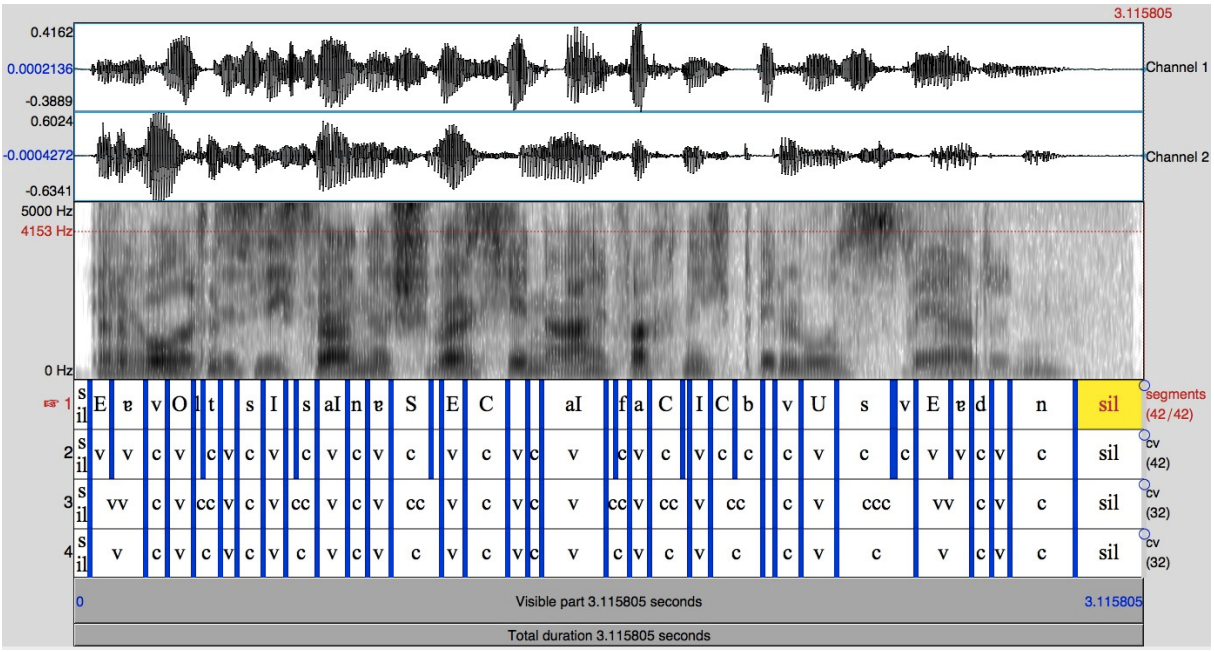

Abb. 1: Segmentierung des Satzfragments "Er wollte sich seiner Schwächen" in konsonantische und vokalische Intervalle (c-v). Für die Bearbeitung wurde das Programm Praat (www.praat.org) verwendet.

2.5 Statistische Auswertung

Auf Grundlage der Dauern der vokalischen und konsonantischen Intervalle konnten die akustischen Rhythmusmasse berechnet werden. Mit Hilfe von Korrelationsanalysen ist es daraufhin möglich, den Grad des linearen Zusammenhangs der einzelnen Konditionen darzustellen. Sollte eine starke Korrelation der Ausgangswerte (read) und Synchronisierungswerte (sync) festgestellt werden, während die Korrelation zwischen den Zielwerten (target) und den Synchronisierungswerten (sync) nur schwach ist, könnte dies bedeuten, dass jenes untersuchte Rhythmusmass an den Sprecher gebunden ist. Aus diesem Grund lohnt sich ebenfalls eine sprecherspezifische Analyse der Messwerte, denn in diesem Fall müssten die Werte für das jeweilige Rhythmuskorrelat bei der Synchronisierung eines einzelnen Satzes eine kleinere Streuung zeigen als die Zielwerte.

Sollte eine starke Korrelation der Synchronisierungswerte (sync) und Zielwerte (target) ermittelt werden, wäre die Betrachtung ebenfalls sinnvoll. Die Streuung der Messwerte müsste dann (bei einer perfekten Anpassung) genau jener der Gruppe der Zielsprecher entsprechen.

3. Ergebnisse

Auf den ersten Blick zeigt sich eine randomisierte Verteilung der Messwerte. Für alle untersuchten Rhythmuskorrelate (%V, nPVI-C, nPVI-V, nPVI-CV) ist zunächst kein einhaltliches Muster erkennbar. Bei der Synchronisierung können die Werte konstant bleiben, sich einem Zielwert annähern oder sich sogar (scheinbar) unabhängig von Ausgangswert (read) und Zielwert (target) verändern. Die graphische Darstellung gibt hierüber einen schnellen und einfachen Überblick. In den Abbildungen 2-5 sind alle Messwerte für die drei Konditionen (read/sync/target) ablesbar. Die Relation gibt hierbei das jeweilige Sprecherpaar an, d.h. 51 steht beispielsweise für den Synchronisierungsversuch von Sprecher 5 mit Sprecher 1. Als Referenz zur Synchronisierung (sync) ist immer der Ausgangswert (read) und Zielwert (target) angegeben.

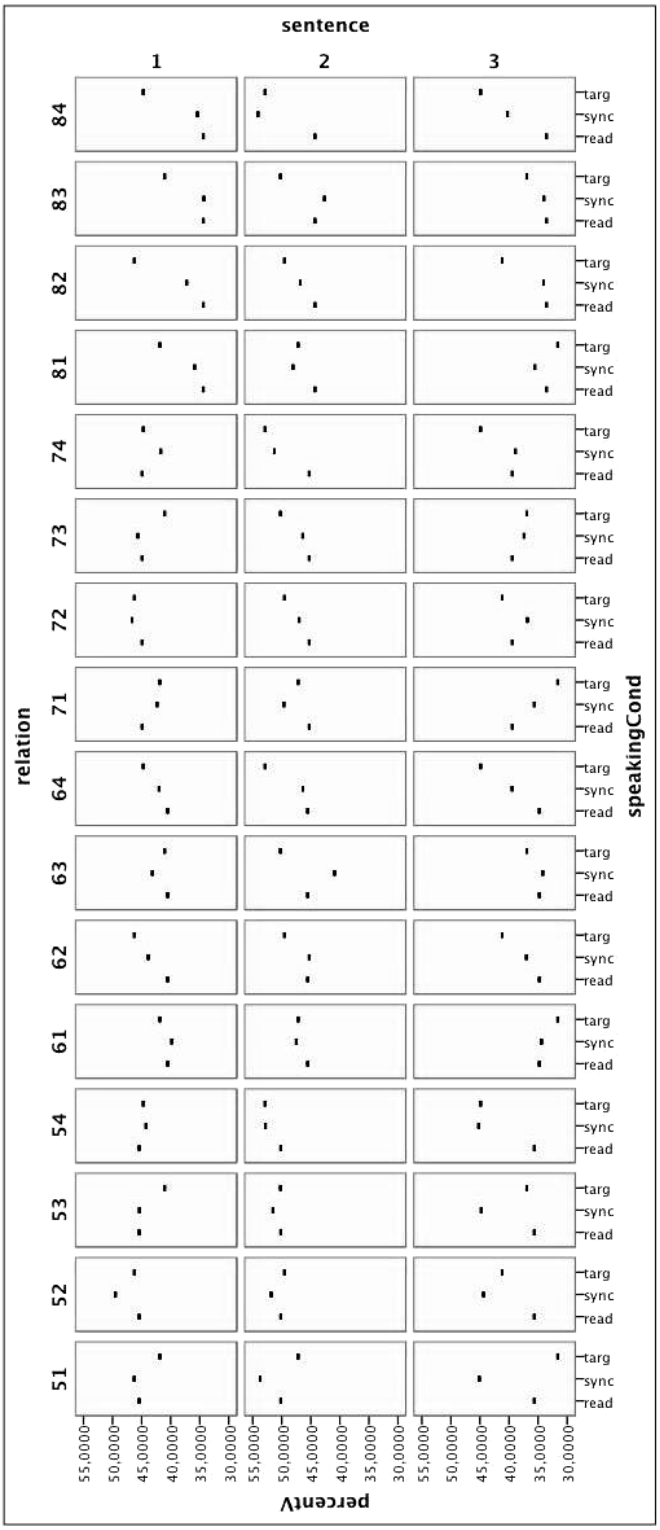

Abb. 2: Messwerte für %V

Korrelationsanalysen konnten allerdings zeigen, dass bei jedem der vier untersuchten Rhythmuskorrelate eine signifikante Korrelation sowohl zwischen Ausgangs- und Synchronisierungswert (read/sync) als auch Ziel – und Synchronisierungswert (target/sync) besteht. Die Korrelation der Ausgangswerte (read) und der Synchronisierungswerte (sync) ist hierbei

jedoch stark, während lediglich eine schwache bis mittlere Korrelation der Zielwerte (target) mit den Synchronisierungswerten (sync) berechnet werden konnte. In Tabelle 1 sind die Ergebnisse der Analyse aufgelistet.

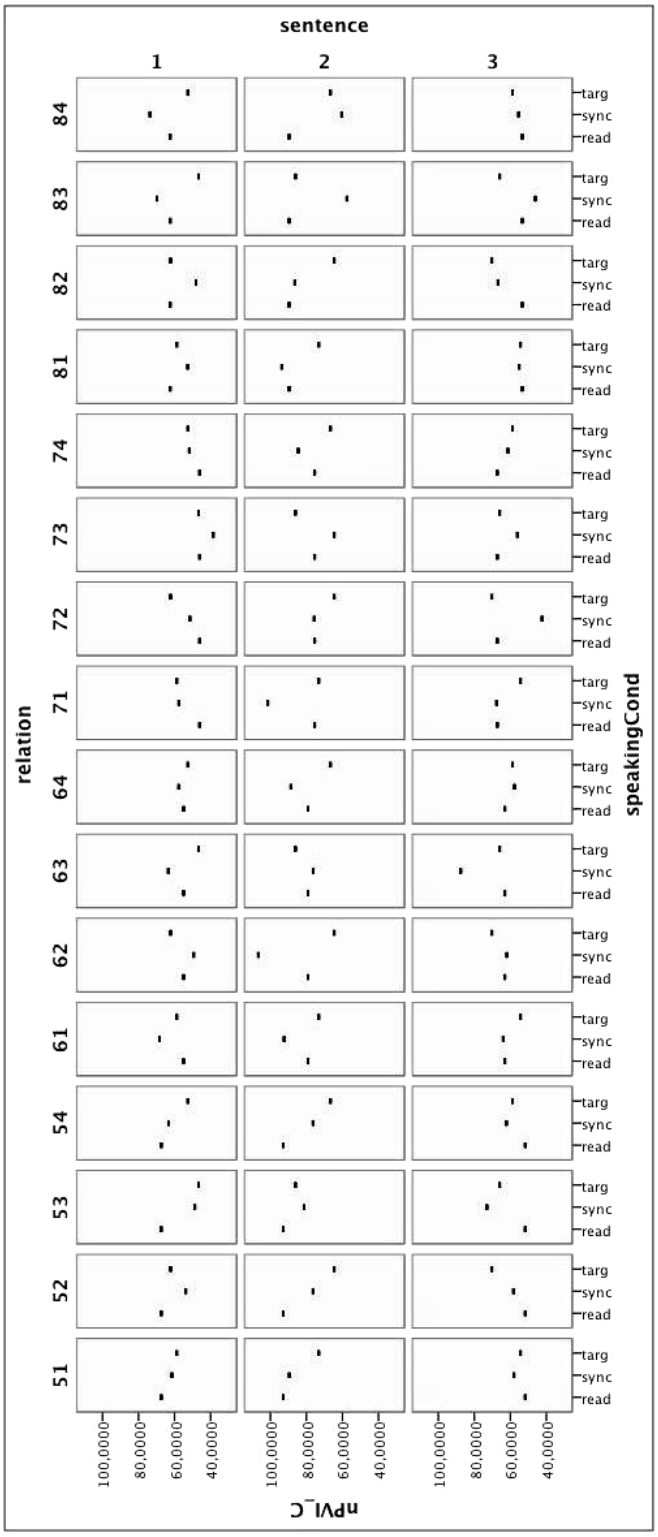

Abb. 3: Messwerte für nPVI-C

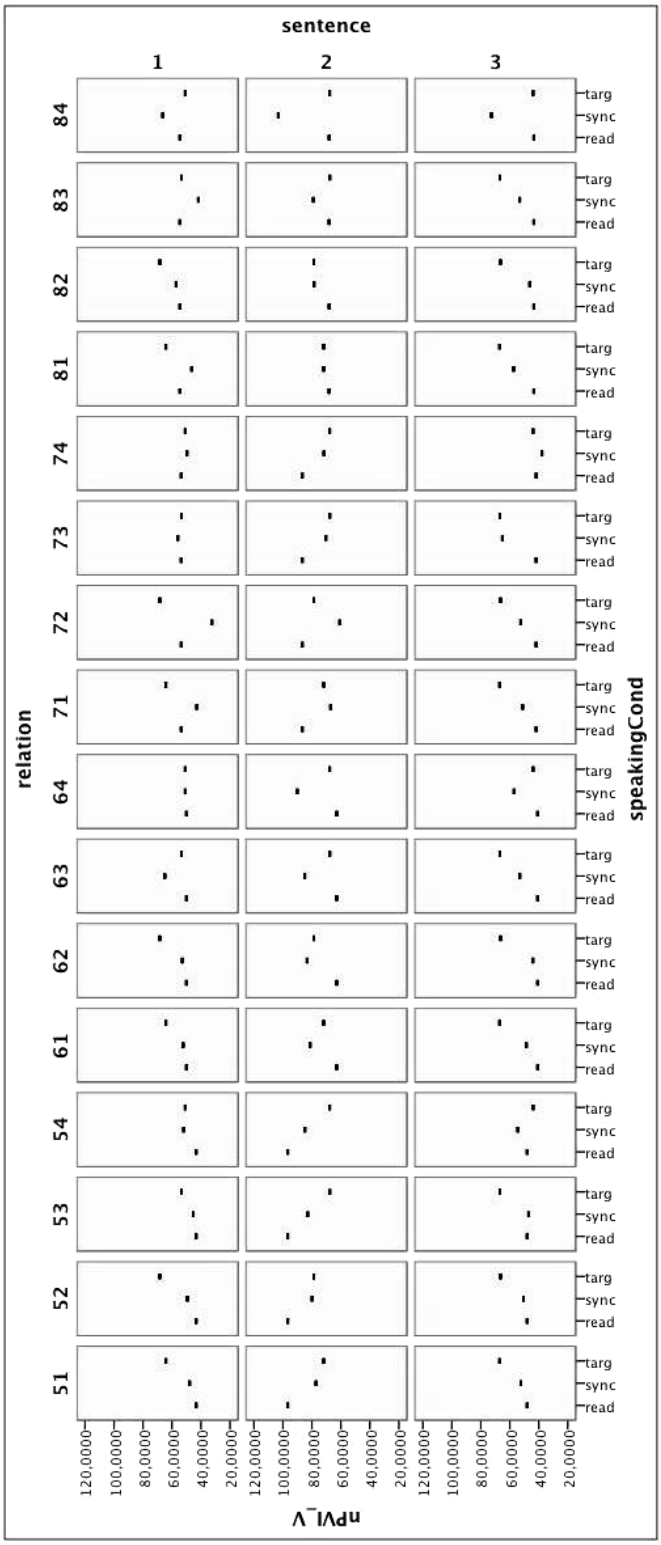

Abb. 4: Messwerte für nPVI-V

Besonders gross ist die Abweichung im Fall der Messungen des prozentualen vokalischen Anteils. Die Untersuchung der drei Konditionen für %V zeigt zwischen Ausgangswerten (read) und Synchronisierungswerten (sync) eine besonders starke Korrelation (r=0.809), während die Korrelation zwischen Zielwerten (target) und

Synchronisierungswerten (sync) verhältnismässig schwach ist (r=0.296). Das Bestimmtheitsmass spiegelt dieses Ergebnis wider und entspricht r^2=0.654 respektive r^2=0.088. Die starke Korrelation von Ausgangswerten (read) und Synchronisierungswerten (sync) und die schwache Korrelation von Synchronisierungswerten (sync) und Zielwerten (target) zeigt sich in Abbildung 6 deutlich.

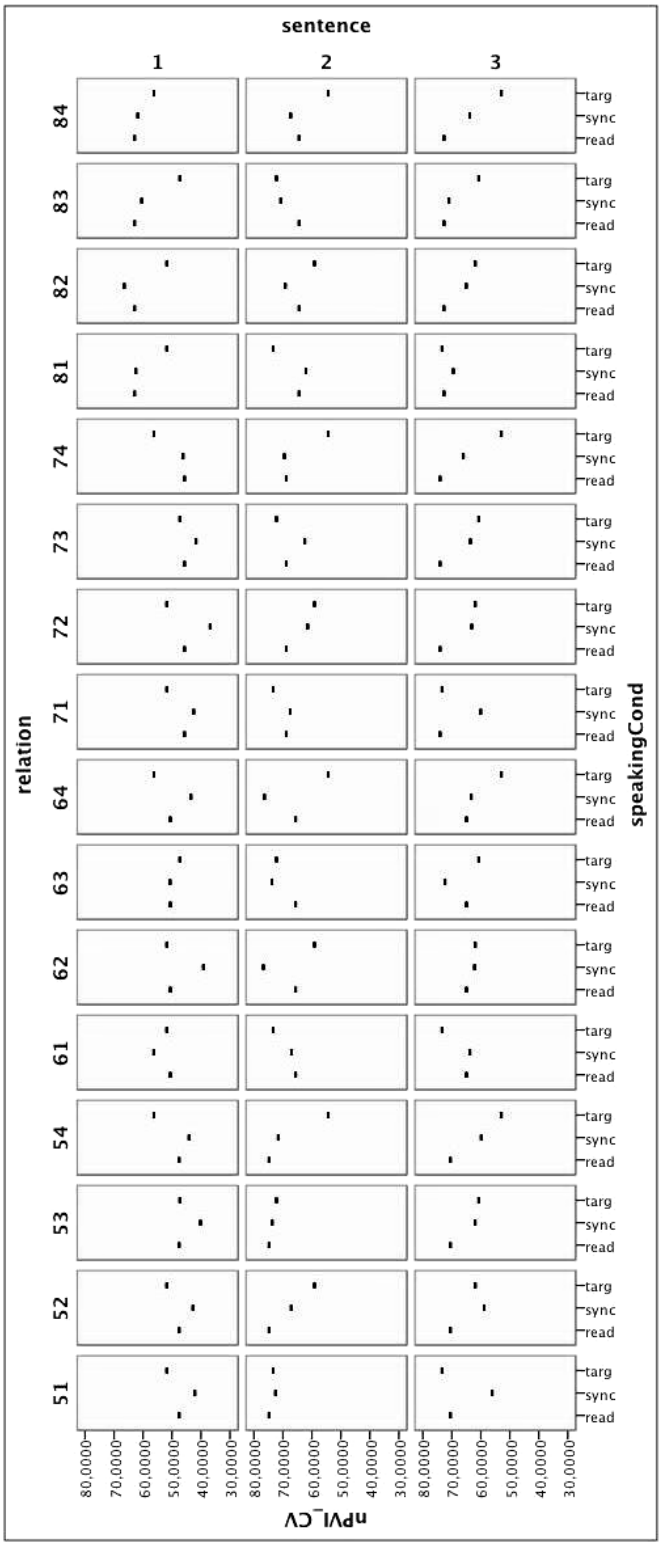

Abb. 5: Messwerte für nPVI-CV

Akustisches Rhythmuskorrelat	r (sync/read)	r² (sync/read)	Sig. (sync/read)
%V	0.809	0.654	p<0.001
nPVI-C	0.609	0.371	p<0.001
nPVI-V	0.65	0.423	p<0.001
nPVI-CV	0.828	0.686	p<0.001

Akustisches Rhythmuskorrelat	r (sync/target)	r² (sync/target)	Sig. (sync/target)
%V	0.296	0.088	p=0.041
nPVI-C	0.383	0.147	p=0.007
nPVI-V	0.367	0.135	p=0.010
nPVI-CV	0.484	0.234	p<0.001

Tab. 1: Ergebnisse der Korrelationsanalysen für die vier untersuchten Rhythmuskorrelate

Ferner ist für %V in vielen Fällen eine verhältnismässig enge Verteilung (Streuung) der Messwerte der Synchronisierungsversuche erkennbar. Dabei können diese jedoch auch deutlich vom Ausgangswert abweichen. Ein Blick auf die Interquartilabstände der synchronisierten Versionen von Sprecher 5 verdeutlich diese Beobachtung (siehe Abb. 7).

Die PVI-Rhythmusmasse nPVI-C und nPVI-V zeigen ebenfalls eine starke Korrelation von Ausgangs- und Synchronisierungswerten und eine schwache Korrelation von Ziel- und Synchronisierungswerten (siehe Tab. 1). Die graphische Darstellung vermag in diesen Fällen dieses Ergebnis zwar anzudeuten, jedoch nicht so deutlich wie im Fall von %V. Abbildung 8 verdeutlicht am Beispiel von nPVI-V diese Beobachtung.

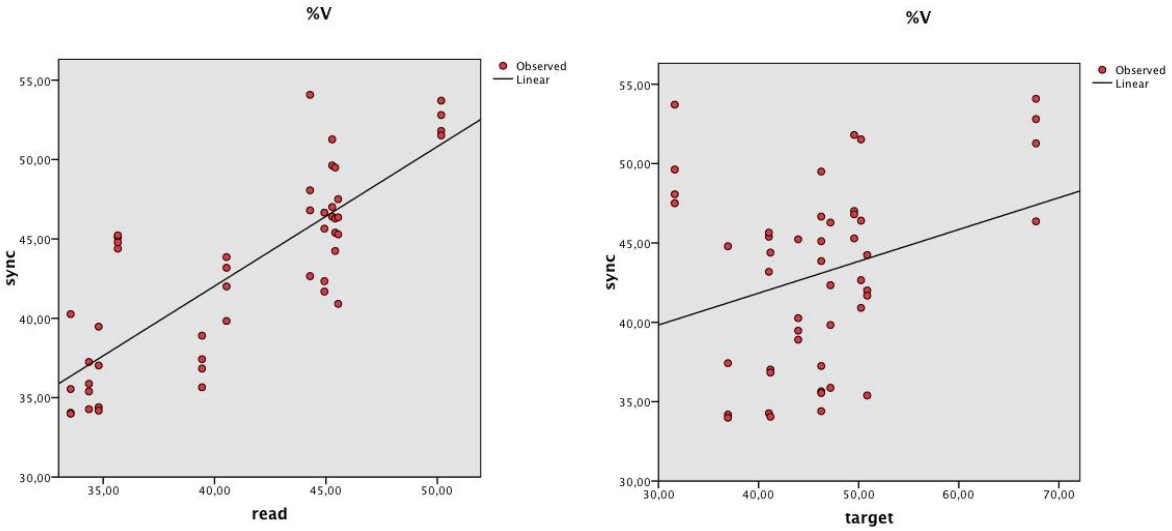

Abb. 6: Korrelation der %V-Werte für die Kombinationen read/sync (links) und target/sync.

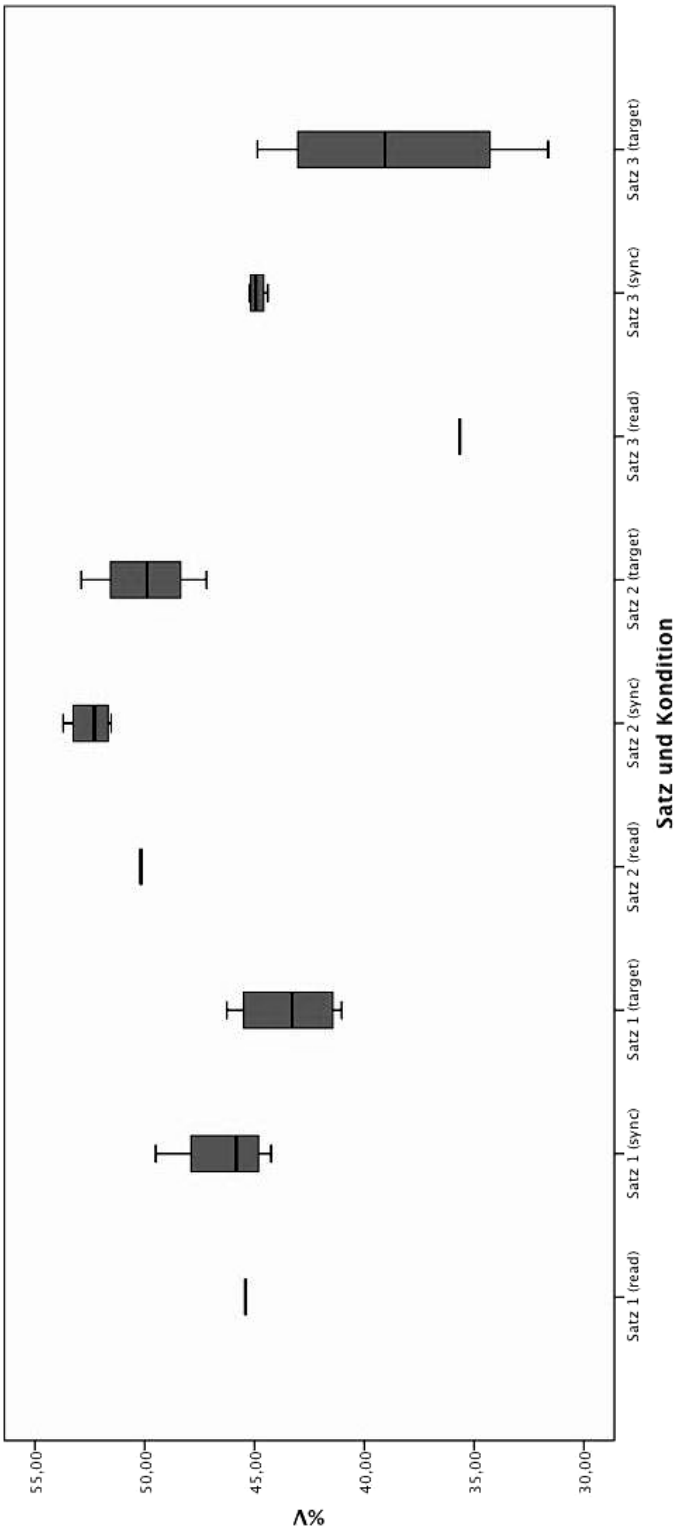

Abb. 7: Boxplots der Verteilung der Messwerte für %V bei Sprecher 5 für die Sätze 1-3 in allen drei Konditionen (read/sync/target)

Die PVI-Rhythmusmasse nPVI-C und nPVI-V zeigen ebenfalls eine starke Korrelation von Ausgangs- und Synchronisierungswerten und eine schwache Korrelation von Ziel- und Synchronisierungswerten (siehe Tab. 1). Die graphische Darstellung vermag in diesen Fällen dieses Ergebnis

zwar anzudeuten, jedoch nicht so deutlich wie im Fall von %V. Abbildung 8 verdeutlicht am Beispiel von nPVI-V diese Beobachtung.

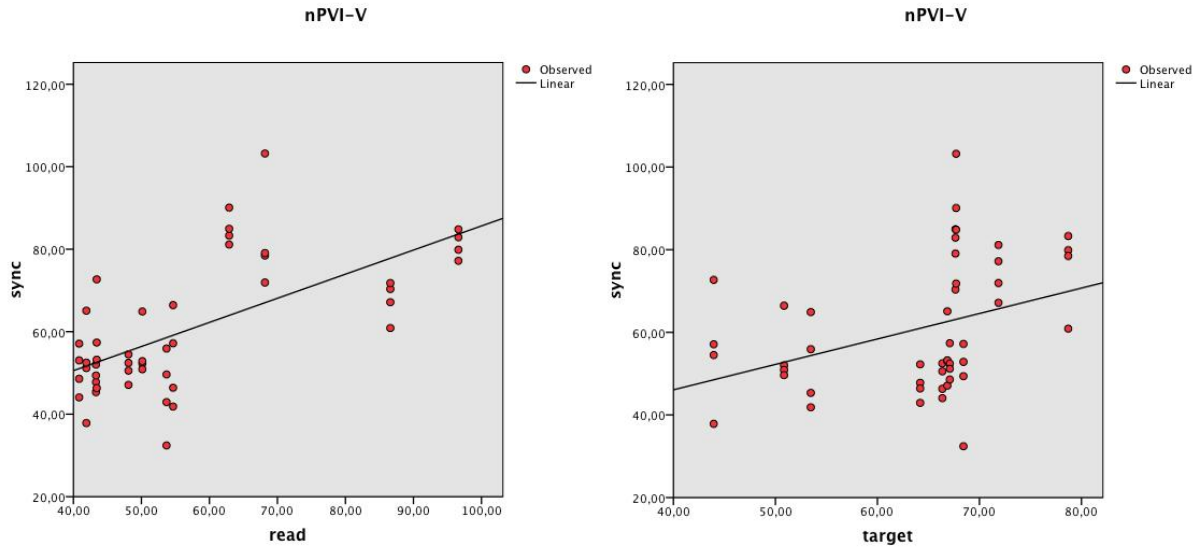

Abb. 8: Korrelation der nPVI-V-Werte für die Kombinationen read/sync (links) und target/sync

Für das Rhythmuskorrelat nPVI-CV wurde eine signifikante Korrelation mit $p<0.001$ von sowohl Ausgangs- als auch Zielwerten mit den Synchronisierungswerten berechnet. Die stärkere Korrelation von Ausgangs- und Synchronisierungswerten (read/sync) lässt sich für dieses Rhythmuskorrelat hingegen wieder mittels der graphischen Darstellung (Abb. 9) gut erkennen.

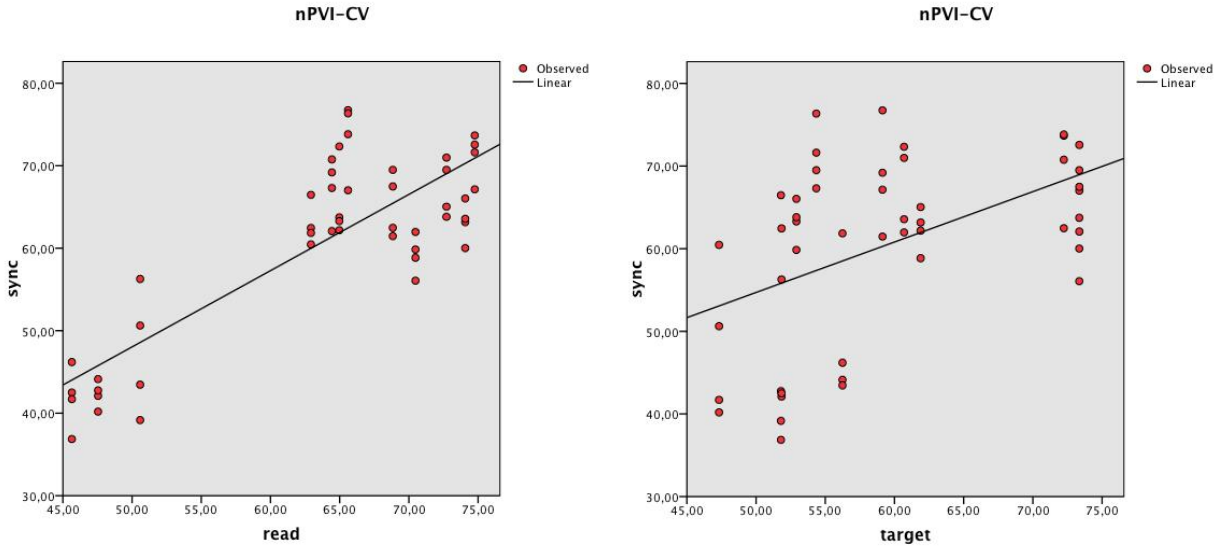

Abb. 9: Korrelation der nPVI-CV-Werte für die Kombinationen read/sync (links) und target/sync

4. Diskussion

Es konnte eine starke Korrelation von Ausgangs- und Synchronisierungswerten (read/sync) und eine schwache bis mittlere Korrelation von Ziel- und Synchronisierungswerten (target/sync) für die Rhythmuskorrelate %V, nPVI-C, nPVI-V und nPVI-CV berechnet werden. Dies ist ein Indiz dafür, dass diese vier Rhythmusmasse im Wesentlichen von der eigenen Sprache abhängig sein könnten, obgleich die zeitliche Veränderung des Sprachflusses ebenfalls einen Effekt auf die ermittelten Werte zu haben scheint, wie am Rhythmuskorrelat nPVI-CV deutlich wurde. Die Analyse einzelner Sprecherdaten lässt ferner vermuten, dass das Spektrum der erreichbaren Werte jedoch sprecherspezifisch limitiert ist. So konnte beispiesweise für %V bei den Synchronisierungsversuchen eine geringe Streuung der Messwerte (kleinere Interquartilabstände) gezeigt werden, während die Streuung der Zielwerte in fast allen Fällen grösser war (vgl. hierzu Abb. 2 und vor allem Abb. 7).

Aufgrund der kleinen Datenmenge kann an dieser Stelle noch nicht mit Sicherheit behauptet werden, dass es sich um weitestgehend sprecherspezifische Rhythmusmasse handelt. Die Ergebnisse lassen allerdings weitere Experimente zur rhythmischen Variabilität eines Sprechers als sinnvoll erscheinen.

Dass die Messwerte trotz des Hinweises auf die mögliche Sprechergebundenheit (durch die hohen Korrelationswerte) nicht konstant bleiben (vgl. Abb. 2-5), könnte ein Indiz für eine Intra-Sprecher-Variabilität der zeitlichen Intervalle sein. D.h. ein Sprecher könnte selbst ohne dies anzustreben und ohne externe Einflüsse variierende konsonantische und vokalische Intervalle und damit unterschiedliche Messwerte erzeugen. Auch die bei Sprecher 5 beobachtete Veränderung der %V-Messwerte während der Synchronisierung auf ein einheitliches Niveau könnte mit dieser temporalen Intra-Sprecher-Variabilität in Kombination mit dem evtl. sprecherspezifischen Charakter von %V erklärt werden.

Zu beachten ist im Rahmen der hier präsentierten Studie ferner, dass Ausgangs- und Zielwerte für einige Rhythmuskorrelate sehr ähnlich waren. Aus diesem Blickwinkel erscheinen hohe Korrelationskoeffizienten natürlich in einem anderen Licht.

Sollten sich die hier gemachten Ergebnisse in weiteren Versuchen jedoch bestätigen, könnten die untersuchten Rhythmuskorrelate zur Identifizierung oder Unterscheidung von Personen herangezogen werden. Da hierbei die Intra-Sprecher-Variabilität scheinbar berücksichtigt werden muss, wäre es allerdings notwendig, über ausreichend grosses Vergleichsmaterial zu verfügen. Ferner müssten die Aufnahmen den identischen Inhalt haben, da natürlich unterschiedliche Lautsequenzen differierende Messwerte erzeugen würden. Wäre dies gewährleistet,

könnten mit Hilfe von Korrelationsanalysen mehrere Aufnahmen verglichen werden, um herauszufinden, ob es sich um identische oder unterschiedliche Sprecher handelt. Sollten sich die Werte eines Rhythmuskorrelates stark unterscheiden und ggf. sogar nicht signifikant korrelieren, dürfte dies dazu führen, dass eine Übereinstimmung eines Sprechers bei zwei oder mehreren Aufnahmen mit hoher Wahrscheinlichkeit ausgeschlossen werden kann. Mit dieser Methode könnten beispielweise Personen, welche auf Grundlage einer Audioaufnahme (z.B. eines Telefongesprächs) des Begehens einer Straftat verdächtig sind, entlastet werden. Denkbar wäre in diesem Zusammenhang ferner auch die Kombination mehrerer geeigneter Rhythmusmasse.

Bibliographische Angaben

Alekin, RO, Klaas, Y.A, Christovich, LA (1962): Human reaction time in the copying of aurally perceived vowels. In: Soviet physics: Acoustics 8, (1), 17ff.

Auer, P. (1993): Is a rythm-based typology possible? A study of the role of prosody in phonological typology. KontRI Working Paper, 21.

Barry, W.J., Andreeva, B., Russo, M., Dimitrova, S., Kostadinova, T. u. a. (2003): Do rhythm measures tell us anything about language type. In: Proceedings of the 15th ICPhS Barcelona, 2693-2696.

Cummins, F. (2002): On synchronous speech. In: Acoustic Research Letters Online, 3, (1), 7-11.

— (2003): Practice and performance in speech produced synchronously. In: Journal of Phonetics, 31, (2), 139-148.

— (2009): Rhythm as entrainment: The case of synchronous speech. In: Journal of Phonetics, 37, (1), 16-28.

Cunado, D., Nixon, M.S., Carter, J.N. (2003): Automatic extraction and description of human gait models for recognition purposes. In: Computer Vision and Image Understanding, 90, (1), 1-41.

Crystal, T.H. (1982): House, A.S.: Segmental durations in connected speech signals: Preliminary results. In: The journal of the acoustical society of America, 72, 705-716.

Dauer, R.M. (1987): Phonetic and phonological components of language rhythm. In: Proceedings of the XIth International Congress of Phonetic Sciences Tallinn, Bd., 5, 447-450.

Dellwo, V., Huckvale, M., Ashby, M. (2007): How is individuality expressed in voice? An introduction to speech production and description for speaker classification. In: Speaker Classification I, S. 1-20.

Dellwo, V., Ramyead, S., Dankovicova, J. (2009): The influence of voice disguise on temporal character- istics of speech. Abstract presented at the annual IAFPA meeting 2009, Cambridge/UK.

Grabe, E., Low, E.L. (2002): Durational variability in speech and the rhythm class hypothesis. In: Papers in laboratory phonology, 7, 515-546.

Foster, J.P., Nixon, M.S., Prügel-Bennett, A. (2003): Automatic gait recognition using area-based metrics. In: Pattern Recognition Letters, 24, (14), 2489-2497.

Krivokapic, J. (2007): Prosodic planning: Effects of phrasal length and complexity on pause duration. In: Journal of phonetics, 35, (2), 162-179.

Marslen-Wilson, W. (1973): Linguistic structure and speech shadowing at very short latencies. In: Nature, 244(5417), 522-523.

McDougall, K. (2007a): Dynamic features of speech and the characterization of speakers: Towards a new approach using formant frequencies. In: International Journal of Speech Language and the Law, 13, (1), 89-126.

— (2007b): Dynamic features of speech and the characterization of speakers: Towards a new approach using formant frequencies. In: International Journal of Speech Language and the Law, 13, (1), 89-126.

Nolan, F. (1991): Forensic phonetics. In: Journal of Linguistics, 27, (2), 483-493.

— (1997): Speaker recognition and forensic phonetics. In: The handbook of phonetic sciences, 744-767.

Nolan, F., McDougall, K., De Jong, G., Hudson, T. (2009): The DyViS database: style-controlled recordings of 100 homogeneous speakers for forensic phonetic research. In: International Journal of Speech Language and the Law, 16, (1), 31-57.

Poore, M.A., Ferguson, S.H. (2008): Methodological variables in choral reading. In: Clinical linguistics & phonetics, 22, (1), 13-24.

Porter, R.J., Lubker, J.F. (1980): Rapid reproduction of vowel–vowel sequences: Evidence for a fast and direct acoustic–motoric linkage in speech. In: Journal of Speech & Hearing Research, 593-602.

Ramus, F., Nespor, M., Mehler, J. (1999): Correlates of linguistic rhythm in the speech signal. In: Cognition, 73, (1), 265-292.

Roach, P. (1982): On the distinction between ‘stress-timed’ and ‘syllable-timed’ languages. In: Linguistic controversies, 73-79.

Shockley, K., Sabadini, L., Fowler, C.A. (2004): Imitation in shadowing words. In: Attention, Perception, & Psychophysics, 66, (3), 422-429.