人類行爲訊號處理（BSP）：一個新興的跨領域研究，其未來在教育領域自動評估的應用

李祈均
國立清華大學電機工程學系助理教授

一、前言

人類行為訊號處理（Behavioral Signal Processing, BSP）（Narayanan & Georgiou, 2013）是一個新興的跨領域研究，目的是以統計計算方式來對人類行為的訊號進行量化、分析、及建模。這個研究領域產物被稱為 – behavior informatics – 先進的跨領域整合式人類行為演算法。這些演算法的研究成果可提供不同領域的專家更客觀（objective）、更有效率（efficient）、及更大規模（large scale）的方法來分析及評斷人類行為。此概念是來自於人類的行為會表現在顯性和隱性的線索（overt and covert cues），這些外顯行為的線索經過演算法萃取及模型，可以提供專家來進行分析與決策。很多的研究都集中在開發相關的數學演算法以用來對於人主觀評價（subjective evaluation）的評量過程進行系統式的自動化。

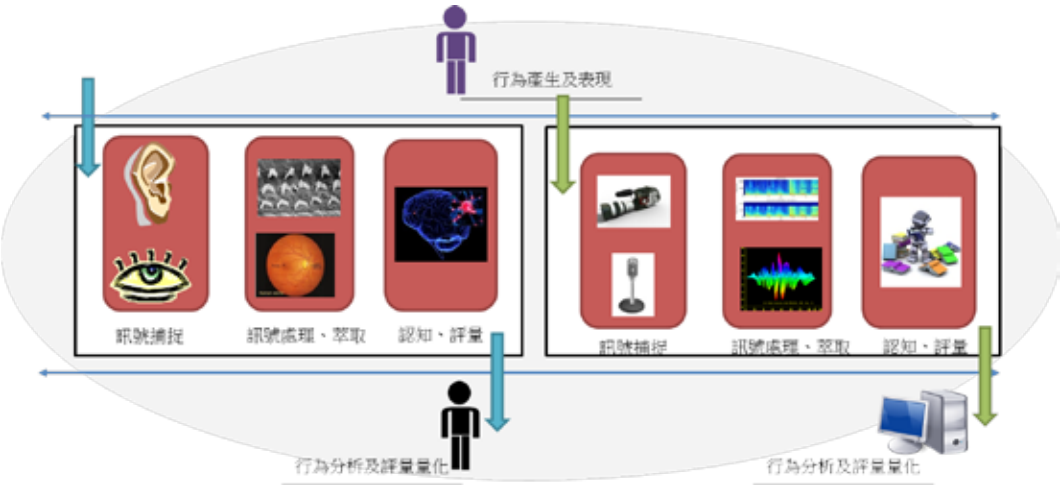

圖 1 行為分析及評量：專家處理的三個要件及電腦訊號處理的三個相對應要件

圖 1 很統括的把這種系統的可行性概念給表達出來。舉一個在精神醫療例子來說，A（病人）去 B（醫生）進行診斷及治療。因為 A 的精神狀態導至 A 行為表現即會透露出一些被醫生注意到的行為線索。醫生會注意到的行為線索通常是依據該科的診斷書（diagnostic instrument）或是專業上訓練和經驗判斷。依此，B 可以對 A 的精神狀態透過觀察 A 的行為做出一個主觀的判斷，並接續著提供相關的治療。這種 B 的主觀判斷能力可以抽象化的分成三個主要要

件：行為訊號捕捉、行為訊號分析、及行為訊號辨識。行為訊號捕捉管道可透過人的感官（sensing），譬如視覺及聽覺。這些訊號接著會被人的感官神經進行處理及分析後（processing），接著把處理過後的特徵（features）送往大腦進行辨識（recognition）。另一個簡單的生活化例子如下：A 生氣時會說話變得很大聲，這個‘大聲’的特徵值經過 B 感官，即耳朵和聽覺神經的捕捉跟處理，接著將此特徵值送往 B 大腦。而 B 的大腦早已學習過‘大聲 = 生氣’的判別，因此可以很自然的可以對 A 的情緒進行自動辨識。

現代數位科技對於這三個不同要件已有長足的發展，因此也慢慢整合出一套電腦可以完成的端至端（end-to-end）‘人類行為訊號處理’系統。數位科技的進步使的訊號捕捉的方式變的普及化，譬如錄音錄影設備可以方便及快速的錄製大量的聲頻及視頻；日漸便宜的儲存媒介也讓這些資料可以做長期及大量的保存。訊號處理技術的進步也使的這些影像聲音的捕捉更加清晰、低噪音，並且也同時發展出各種不同的計算方法來對原始聲頻及視頻串流進行快速的‘特徵’計算（feature computation），這些特徵的計算是仿效人類感官的處理（perceptual processing）。最後，統計模型、人工智慧、機器人學習等數學及工程領域的發展，隨著資料量的增加、特徵 更精準的萃取、以及更快速的電腦運算，也產生了許許多多複雜且高階的演算法使機器能從資料中‘學習’並進行以數學統計為基礎的自動模式識別（pattern recognition）。這三個在不同科技領域的發展慢慢在匯流形成一個新興的研究方向，讓電腦也可以以高可靠度的方式做一些一向以來皆由人類大腦在執行的的主觀判別。

這篇文章首先要介紹這個新興領域對於跨領域人類行為的研究遠景及動機，簡略的介紹既有的精神醫療方面的研究，以及最後如何將此概念移植到教育領域。透過密合的跨領域合作，即資訊科技及教育專家，在教育領域裡的各式各樣評量或也可以慢慢的採用電腦透過人類行為的客觀模型之演算法。

人類行為訊號處理（Behavioral Signal Processing, BSP）（Narayanan & Georgiou, 2013）是一個新興的跨領域研究，目的是以統計計算方式來對人類行為的訊號進行量化、分析、及建模。這個研究領域產物被稱為 – behavior informatics – 先進的跨領域整合式人類行為演算法。這些演算法的研究成果可提供不同領域的專家更客觀（objective）、更有效率（efficient）、及更大規模（large scale）的方法來分析及評斷人類行為。此概念是來自於人類的行為會表現在顯性和隱性的線索（overt and covert cues），這些外顯行為的線索經過演算法萃取及模型，可以提供專家來進行分析與決策。很多的研究都集中在開發相關的數學演算法以用來對於人主觀評價（subjective evaluation）的評量過程進行系統式的自動化。

二、研究遠景及動機

人類行為訊號處理的技術發展，主要目的之一是在於幫忙不同領域的專家解決對於人行為評量既有方法的問題。這種依靠人（專家）來對其他人的行為進行評量的過程被統稱為人工觀測方法（manual observational approach）（Margolin et

al., 1998），幾個長期困擾的問題如下（Baucom & Iturralde, 2012）：

（一）主觀評量 (subjective evaluation)

人工觀測方法，因為是‘人工’，所有的評量皆屬於主觀（subjective）評量。所有主觀評量都會存在著因為人的因素（human factors）而產生不確性的問題（issue of reliability）。例如，疲勞，經驗不足，觀念不同，及其他相關非跟行為評量本身相關因素都會造成人工評量結果存在著不可避免也不容易控制的變數。

（二）耗時 (time-consuming)

人工觀測方法，因為是‘人工’，也相對比較耗時。再加上主觀評量有的先天問題，在各別專業領域，譬如醫療或心理，相當多的研究都著重在如何訓練人員進行有一致性（consistent）及可靠（reliable）的評量。解決主觀評量的問題包含了對個別人員進行紮實訓練，亦或是運用多位評量員來減低個別評量員可能造成的問題。這樣為了確保一致性的方式進一步的使得人工觀測方法更加費時。

（三）不容易擴展 (non-scalable)

由於前兩項的問題，利用人力來進行行為評量很難進行大規模應用。無法擴展成大規模應用的後果導致在行為研究方面僅能維持相對小量的樣本數來加以分析，或是將評量方式改以問卷或其他非行為類評量方式以便執行大規模應用。在現今數位科技的時代，行為訊號捕捉的工具已相當普遍，資料量的累積也逐漸變得快速且大量。透過行為評量可以比問卷或其他類似的測量對要測量的行為或能力進行更精準即客觀的評斷。

透過人類行為訊號處理的技術發展，目的在於提供自動演算法來解決以上述的相關問題。從一些既有的專家評量中，讓電腦演算法可以從這些既有行為資料中先學習如何模擬專家觀察的標準。在確保可信度（reliability）後，這套自動化評量系統即可進一步達成一致性（consistent），可複製性（repeatable），及大量且快速的評量（scalable）。人工智慧演算法的可信度也會隨著資料量的不斷累積而變得更準確。

同時，這樣的技術發展及運用其實更可以讓專家做出明智的決策（informed decision），而並非要讓專家被電腦給取代。因為在多數的領域，行為的評量只是介於最後專家決策與一開始人類行為產生中間的一個過程（見圖 2）。譬如說，醫生依據診斷的評量書的結果來進行合適的臨床醫療行為。現今大多數人工行為評量的方式其實是在這個過程中導致了資訊瓶頸（information bottleneck）－使的多數的資訊並無法被最後的決策者給使用。透過自動化的評量，電腦可以提供各領域專家快速而且更客觀的行為評量，加上其他相關資料以便進行最後總和的決策。

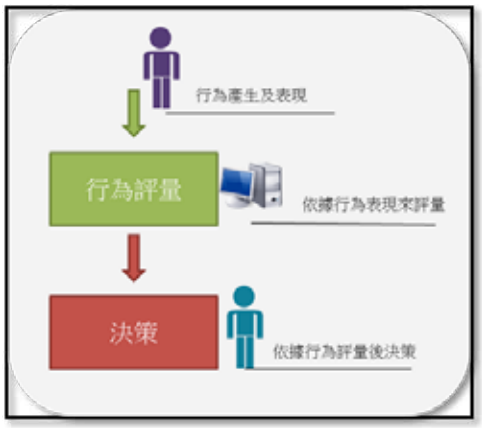

圖 2 專家決策過程：評量後依結果做決策

三、醫療上的應用

人類行為訊號處理領域在過去幾年中對精神醫療方面有深入的研究。我們在這簡略的介紹其中兩個在國外子領域的發展：婚姻諮詢（couples counseling and therapy）、泛自閉症症狀（autism spectrum disorder, ASD）。

（一）婚姻諮詢

長年有婚姻生活有問題的夫妻尋求夫妻心理治療（couple therapy）在國外已是常態。臨床心理學家已建構對這些夫妻數種不同的治療方式，譬如 Traditional Behavioral Couple Therapy（Neil S Jacobson & Margolin, 1979)、Integrative Behavioral Couple Therapy（Christensen et al., 2004; N. S. Jacobson, Christensen, Prince, Cordova, & Eldridge, 2000)。 在作為治療的一部分，夫妻都需要經歷解決問題的談話互動（problem-solving interaction）。在這十分鐘的談話互動中，夫妻必須各選一個他們婚姻中的問題提出來討論並尋求解決的方法。臨床心理學家會將這個對談互動進行錄音錄影。專家們為了研究治療的效果，夫妻互動對談的模式，及將其行為來預測未來產生離婚或家庭暴力的可能，臨床心理學家會請數名人員（通常 3 至 5 名)，經過訓練後，以觀看錄音錄影檔的方式來對夫妻的互動進行行為評量。

行為評量的準則是根據兩份已被認可的行為評量表：Couples Interaction Rating System（CIRS)（Heavey, Gill, & Christensen, 2002) 和 Social Support Interaction Rating System （SSIRS)（Jones & Christensen, 1998)。這兩份量表對夫妻各別在解決問題的互動的行為進行評分，總共評分項目有 33 項。以下舉例其中一個評分項目，內容是直接引用 SSIRS 量表：

“Blame: Blames, accuses, or criticizes the partner, uses critical sarcasm; makes character assassinations such as, “you’re a real jackass,” “all you do is eat,” or “why are you such a jerk about it?” Explicit blaming statements (e.g., “you made me do it,” or “you prevent me from doing it”), in which the spouse is the causal agent for the problem or the subject’s reactions, warrant a high score.”

這樣的行為評量方法已行之有年，但臨床心理學家對於這樣的人工觀測方法感到非常耗時而且有時並無法達到必要的可靠度。第一個首要的問題是必須要徵求多名評量員，然後再根據量表內不同的行為定義對評量員加以訓練。所以臨床心理學家對於可以將這樣乏味且耗時的工作自動化覺得不但可以促進研究效率，更可以對這些有心裡有障礙的夫妻們提供更適時的幫忙。

在人類行為訊號處理的研究中，已有數篇文章提出相關技術來對這些專屬的量表進行自動化。這些系統有的是單純藉由語音（acoustics）（Black et al., 2013），有的是語音加用詞（acoustic + lexical）（Black, Georgiou, Katsamanis, Baucom, & Narayanan, 2011; Georgiou, Black, Lammert, Baucom, & Narayanan, 2011），有的是透過對夫妻彼此頭部的動作（head motion）（Xiao, Georgiou, Lee, Baucom, & Narayanan, 2013）。這些初步研究都已顯示自動化系統的可行性，系統準確率也都至少大約七成左右。

（二）泛自閉症症狀

美國 Centers for Disease Control and Prevention（CDC）預估在美國差不多每 88 個小孩中就有一個有泛自閉症症狀。這是個嚴重的且昂貴的社會問題。在研究及臨床診斷自閉症的過程中，常被當成是黃金標準（gold standard）的是 Autism Diagnostic Observation Schedule（ADOS）（Lord et al., 2000）標準化觀察行為量表。當父母帶小孩進醫院進行診斷，臨床心理學家必須要花 30 至 45 分鐘之間的時間來跟小孩互動，並在互動的過程中將小孩不同的行為根據 ADOS 量表進行記錄。

這樣的過程不只是耗時，即診斷時間方面，還有很多其他潛在不確定因素。譬如，臨床心理學家在這個狀態下不只是第三方觀察者，他們同時也是在跟小孩互動的其中一方。這樣的診斷設計必須要求被認可進行診斷的醫療人員經歷更長時間的訓練 – 這尤其重要因為這些診斷會影響小孩子接下來所要接受的醫療行為。自動化部分量表內容不儘可以節省時間，其實更可以提供家長一個‘早期發現’的機會。在研究方面，部分量表的行為其實相對的不容易透過人，即診斷的醫師，來進行評量。譬如說 ADOS – Module 3（內含有 28 個不同行為評量項目）的其中一項行為評量是 Speech Abnormality，在量表內的形容引用如下：

“slow and halting; inappropriately rapid; jerky and irregular in rhythm . . . odd intonation or inappropriate pitch and stress, markedly flat and toneless, . . . consistently abnormal volume”

對於語音變化的量化其實透過電腦的運算比人工判別客觀。事實上已有些初步的人類行為訊號處理的研究指出，透過電腦量化語音及用詞（Bone et al., 2013）或肢體行為（Delaherche et al., 2013）後，可以對小孩的自閉症的嚴重程度評分（autism severity score）進行自動化的評量。此外因為是透過電腦的運算，可以直接對小孩和臨床心理學家在診斷的過程建立一個共同行為的量化及模型（joint modeling）（Bone et al.）。這樣雙邊行為的同時觀測甚至可以對小孩的自閉症的嚴重程度評分的準確率更加提升。

四、教育上的應用

人類行為訊號處理技術在對於教育相關的評量在概念上其實是跟精神疾病醫療應用相當類似。因為在教育評量領域裡所遭遇的問題跟第二段所形容的人工觀測評量是相仿的。雖然在教育科技的發展目前還沒有像醫療應用的多面向且縝密的跨領域整合，但在國外教育類科技的發展中，已經有些初步技術對‘第二外語’學生的發音進行自動評量。近期也有數篇論文在研發針對老師的網路上教學的好壞評比研發自動化的評量系統。

第二外語的學生學習英文，在發音上會被母語的影響而導致部分發音的不準確。英語教學的一個重點在確保這些第二外語的學生的發音得準確度。雖然沒有像醫療用的制式量表，但英語老師，甚或一般母語為英文的人皆可以判別發音的正確與否，並對不準確的發音進行修正。一直以來這些方式都是透過人工的方法，使的外語學習普及化的速度並不快速。

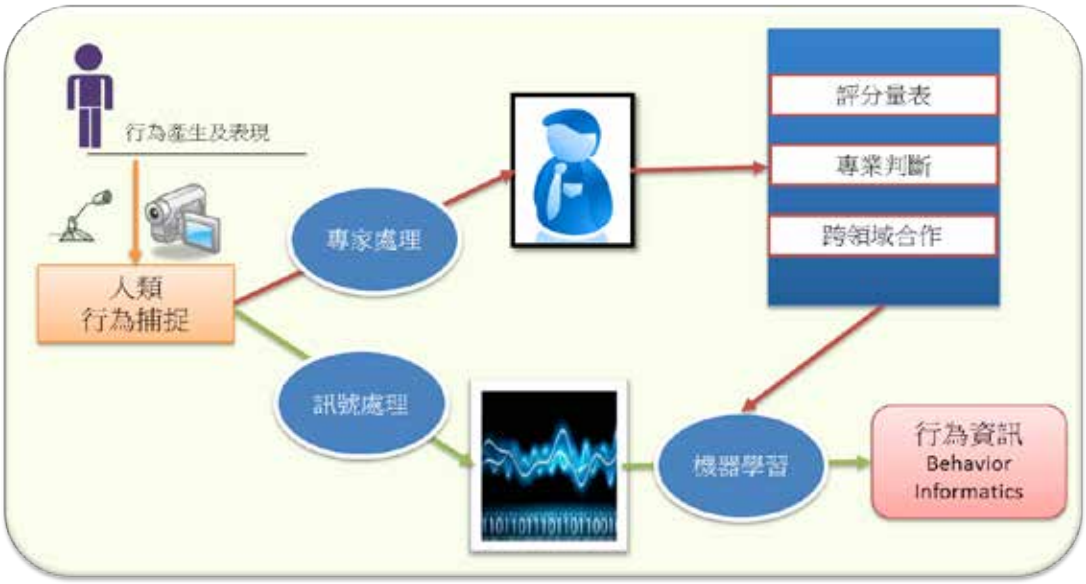

圖 3 人類行為訊號處理：研究領域的發展必須與各領域專家有縝密的結合，行為資訊的產生才會有意義

已有數篇論文著重對此問題進行自動評量（Alwan et al., 2007; Black, Kazemzadeh, Tepperman, & Narayanan, 2011; Black, Tepperman, & Narayanan, 2011）。這些自動評分的系統在一些客觀的驗證下，甚至比人跟人之間的對發音正確與否判斷的 agreement 來的更高。

這樣概念的自動評量系統可以將基礎的外語學習用電腦軟體系統先落實到一般生活中。而專家，即專業英語教師，反而可以更將教學著重在比較抽象的概念，譬如說語感、語意（connotation）、甚或是面對面溝通的技巧。

此外，最近也有一些新興的研究在對於線上教學（http://videolectures.net/）的視頻進行自動化評量（Cheng et al., 2014; Salvagnini, Salamin, Cristani, Vinciarelli, & Murino, 2012）。在現在網路發達的時代，在國外，甚至一些國內大學，已經將大量不同教學科目放置在網路上供一般民眾學習。這些教學影帶在網路上都有好壞評分（類似 youtube 的五顆星的評比）。透過影像處理跟語音處理的技術，一些初步研究已經展現出電腦可以幫忙提供自動化評分的可行性。持續發展這類技術可以在未來讓‘學習推薦系統（learning recommendation system）’變得更人性化及更加精準。

五、未來教育上的應用

雖然已經有些資訊科技使用在教育的應用，但在未來可進步的空間，及加上更縝密的跨領域整合可以讓人類行為訊號處理技術的應用更加切入重點而對教育產生實質影響力。圖 3 描述了人類行為訊號處理技術整個領域的架構。基本上人類行為訊號處理技術是必須要包含專家高密合的參與。這是其中最重要的一環，既高度跟該領域專家的互動，以便將先進的科技技術建立在該領域既有相關的知識及應用上。舉例來說，在教育界，已有長年的研究及累積的經驗針對如何培育各級學校的師資，學校校長，也甚或有一些經過縝密設計的培訓課程。這些課程中

對於該培訓的能力及內容都已經過長時間的淬煉及證實。在既有的課程設計及評量標準中如何幫忙教育專家注入更客觀、更迅速、及最重要的 – 更有用，的自動化資訊技術是未來在教育上應用的關鍵。

對於師資培育，可以建立一套系統來提供跟現有人工方式不一樣的行為表現評分標準，相對比較客觀也比較有可能進行大規模的應用。對於學生的學習效率，國外文獻也漸漸指出學生與老師上課中 emotional and social engagement 是一個重要因素（Blanchette & Richards, 2010; Martin & Dowson, 2009）。如何準確的評量出教學或學習效率，其實學生或老師在課堂上互動行為是非常關鍵的。在人類行為訊號處理技術慢慢的成熟後，數位科技也可以漸漸進入校園及教室。藉由這些技術根據師生行為的互動來進行數位紀錄，以及用演算法自動判別不同教學及學習的狀況 - 藉此也可以更準確的掌握教學效用。

人類行為訊號處理技術（Behavioral Signal Processing）是其中一個新興的‘以人為中心的計算（Human-centric Computing）’的研究領域。其他相關領域，譬如人類社交訊號處理（Social Signal Processing）（Pentland, 2007）、計算行為科學（Computational Behavioral Science）（Oberleitner, Abowd, & Suri, 2013）、情感計算（Affective Computing）（Picard, 2000）也都是在國外新興的研究領域。專門對於人的行為、情緒、社會互動透過客觀的行為捕捉、處理、統計建模來研究人彼此之前的行為互動。這些新興領域的發展對於將來把這些技術實際應用於醫療或教育的評量行為皆會有很大的幫助。

參考文獻

Alwan, A., Bai, Y., Black, M., Casey, L., Gerosa, M., Heritage, M., . . . Lee, S. (2007). *A system for technology based assessment of language and literacy in young children: the role of multiple information sources.* Paper presented at the Multimedia Signal Processing, 2007. MMSP 2007. IEEE 9th Workshop on.

Baucom, B. R., & Iturralde, E. (2012). *A behaviorist manifesto for the 21st century*. Paper presented at the Signal & Information Processing Association Annual Summit and Conference (APSIPA ASC), 2012 Asia-Pacific.

Black, M. P., Georgiou, P. G., Katsamanis, A., Baucom, B. R., & Narayanan, S. S. (2011). *``You made me do it": Classification of blame in married couples' interactions by fusing automatically derived speech and language information.* Paper presented at the Proceedings of Interspeech, Florence, Italy.

Black, M. P., Katsamanis, A., Baucom, B. R., Lee, C.-C., Lammert, A. C., Christensen, A., . . . Narayanan, S. S. (2013). Toward automating a human behavioral coding system for married couples' interactions using speech acoustic features. *Speech Communication, 55*(1), 1-21.

Black, M. P., Kazemzadeh, A., Tepperman, J., & Narayanan, S. S. (2011). *Automatically Assessing the ABCs: Verication of Children's Spoken Letter-Names and Letter-Sounds. ACM Transactions on Speech and Language Processing, 7:4*, 15:11 - 15:17.

Black, M. P., Tepperman, J., & Narayanan, S. S. (2011). *Automatic Prediction of Children's Reading Ability for High-Level Literacy Assessment. Ieee Transactions on Audio Speech and Language Processing, 19*(4), 1015-1028. doi: Doi 10.1109/Tasl.2010.2076389

Blanchette, I., & Richards, A. (2010). The influence of affect on higher level cognition: A review of research on interpretation, judgement, decision making and reasoning. *Cognition & Emotion, 24*(4), 561-595. doi: Pii 914096069 Doi 10.1080/02699930903132496

Bone, D., Lee, C.-C., Black, M. P., Williams, M. E., Lee, S., Levitt, P., & Narayanan, S. The Psychologist as an Interlocutor in Autism Spectrum Disorder Assessment: Insights from a Study of Spontaneous Prosody. *Journal of Speech, Language, and Hearing Research.*

Bone, D., Lee, C.-C., Chaspari, T., Black, M. P., Williams, M. E., Lee, S., . . . Narayanan, S. (2013). *Acoustic-Prosodic, Turn-taking, and Language Cues in Child-Psychologist Interactions for Varying Social Demand.* Paper presented at the Interspeech, Lyon, France.

Cheng, D. S., Salamin, H., Salvagnini, P., Cristani, M., Vinciarelli, A., & Murino, V. (2014). Predicting online lecture ratings based on gesturing and vocal behavior. *Journal on Multimodal User Interfaces, 1-10.*

Christensen, A., Atkins, D. C., Berns, S., Wheeler, J., Baucom, D. H., & Simpson, L. E. (2004). Traditional versus integrative behavioral couple therapy for significantly and chronically distressed married couples. *Journal of Consulting and Clinical Psychology, 72*(2), 176-191. doi: Doi 10.1037/0022-006x.72.2.176

Delaherche, E., Chetouani, M., Bigouret, F., Xavier, J., Plaza, M., & Cohen, D. (2013). Assessment of the communicative and coordination skills of children with Autism Spectrum Disorders and typically developing children using social signal processing. *Research in Autism Spectrum Disorders, 7*(6), 741-756. doi: DOI 10.1016/j.rasd.2013.02.003

Georgiou, P. G., Black, M. P., Lammert, A. C., Baucom, B. R., & Narayanan, S. S. (2011). *``That's aggravating, very aggravating": Is it possible to classify behaviors in couple interactions using automatically derived lexical features?* Paper presented at the Affective Computing and Intelligent Interaction.

Heavey, C., Gill, D., & Christensen, A. (2002). Couples interaction rating system 2 (CIRS2): University of California, Los Angeles. Retrieved from http://christensenresearch.psych.ucla.edu/

Jacobson, N. S., Christensen, A., Prince, S. E., Cordova, J., & Eldridge, K. (2000). Integrative behavioral couple therapy: an acceptance-based, promising new treatment for couple discord. *J Consult Clin Psychol, 68(2), 351-355.*

Jacobson, N. S., & Margolin, G. (1979). *Marital therapy: Strategies based on social learning and behavior exchange principles:* Psychology Press.

Jones, J., & Christensen, A. (1998). Couples interaction study: Social support interaction rating system. University of California, Los Angeles. Retrieved from http://christensenresearch.psych.ucla.edu/

Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Jr., Leventhal, B. L., DiLavore, P. C., . . . Rutter, M. (2000). The autism diagnostic observation schedule-generic: a standard measure of social and communication deficits associated with the spectrum of autism. *J Autism Dev Disord, 30*(3), 205-223.

Margolin, G., Oliver, P. H., Gordis, E. B., O'Hearn, H. G., Medina, A. M., Ghosh, C. M., & Morland, L. (1998). The nuts and bolts of behavioral observation of marital and family interaction. *Clin Child Fam Psychol Rev, 1*(4), 195-213.

Martin, A. J., & Dowson, M. (2009). *Interpersonal Relationships, Motivation, Engagement, and Achievement: Yields for Theory, Current Issues, and Educational Practice. Review of Educational Research, 79*(1), 327-365. doi: Doi 10.3102/0034654308325583

Narayanan, S., & Georgiou, P. G. (2013). Behavioral Signal Processing: Deriving Human Behavioral Informatics From Speech and Language: Computational techniques are presented to analyze and model expressed and perceived human behavior-variedly characterized as typical, atypical, distressed, and disordered-from

speech and language cues and their applications in health, commerce, education, and beyond. *Proc IEEE Inst Electr Electron Eng, 101*(5), 1203-1233. doi: 10.1109/JPROC.2012.2236291

Oberleitner, R., Abowd, G., & Suri, J. S. (2013). Behavior Imaging®'s Assessment Technology: A Mobile Infrastructure to Transform Autism Diagnosis and Treatment *Imaging the Brain in Autism* (pp. 371-380): Springer.

Pentland, A. (2007). Social signal processing. *Ieee Signal Processing Magazine, 24*(4), 108-111. doi: Doi 10.1109/Msp.2007.4286569

Picard, R. W. (2000). Affective computing: MIT press.

Salvagnini, P., Salamin, H., Cristani, M., Vinciarelli, A., & Murino, V. (2012). *Learning how to teach from "Videolectures": automatic prediction of lecture ratings based on teacher's nonverbal behavior.* Paper presented at the Cognitive Infocommunications (CogInfoCom), 2012 IEEE 3rd International Conference on.

Xiao, B., Georgiou, P. G., Lee, C.-C., Baucom, B., & Narayanan, S. S. (2013). *Head motion synchrony and its correlation to affectivity in dyadic interactions.* Paper presented at the Multimedia and Expo (ICME), 2013 IEEE International Conference on.